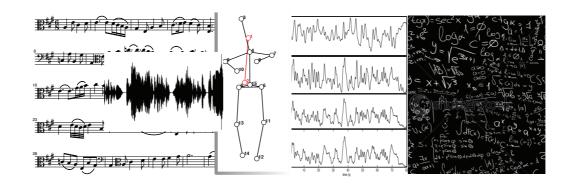# A Statistical Framework

For Embodied Music Cognition

Frank Desmet

Ghent University
Faculty of Arts and Philosophy

IPEM – Dept. of Musicology.

Ghent University

Promotor: Prof. Dr. Marc Leman

Thesis submitted to fulfill
the requirements for the degree of

Doctor in Art Sciences

Ghent University

Faculty of Arts and Philosophy

Department of Musicology

A Statistical Framework

For Embodied Music Cognition

Frank Desmet

Thesis submitted to fulfill the requirements for the degree of

Doctor in Art Sciences

Academic year 2010 – 2011

Candidate's Declaration

I certify that the thesis entitled: "A Statistical Framework for Embodied Music Cognition", submitted for the degree of doctor of art science, is the result of my own research, except where otherwise acknowledged, and that this thesis in whole or in part has not been submitted for an award, including a higher degree, to any other university or institution.

Acknowledgements

This work is the result of a remarkable combination of many things. I recall, as if it happened yesterday, the first contact with IPEM, where somebody was needed to support this small research group at that time, for the maintenance of the network, with interest and experience in statistics and of course had affinity with music. I could not have imagined that this was the start of a wonderful and amazing journey.

First of all, I became aware of the unique position of IPEM, which is nowadays a multi-disciplinary and stimulating environment and a center of excellence in the domain of systematic musicology. I am privileged to have been able to expand my horizon and to share knowledge with skilled and experienced researchers, which all have one thing in common: passion.

Therefore, I want to thank all my colleagues for their support and kindness during this work. Their enthusiasm and encouragements and the appreciation for my work make me humble.

But I want also express my gratitude to some of my colleagues in particular. First of all I want to thank profoundly Micheline. Without her support and engagement I would not have been able to finalize this work. Even when things went not as intended and I encountered problems she helped me to overcome them and stimulated me to continue my work.

Special thanks go also to Luc with whom I had the occasion to discuss and change ideas and who enabled me to re-discover my passion for philosophy and reminded me how important playing music is for myself, to Michiel for his highly appreciated contribution and constructive critical support, to Aagje for the review of my texts and to Katrien for her support and understanding.

And of course, I want to profoundly thank my promotor Marc. It is really a privilege to have had the opportunity to work with him. Not only I discovered a brilliant intellectual mind but above all a man with a great engagement and

commitment, without prejudices. His amazing knowledge in nearly all fields of the multi-disciplinary domain of systematic musicology stimulated me in my work. But I also experienced a man with a deep concern and I want to express my gratitude for that.

And last but not least I want to thank my friends for their patience and their help whenever it was needed.

Ghent University

Faculty of Arts and Philosophy

Department of Musicology

Promotor: Prof. Dr. Marc Leman

Reading committee: dr. Michiel Demey
                              Prof. dr. Giovanni De Poli
                              dr. Micheline Lesaffre
                              Prof. dr. ir. Jean-Pierre Martens
                              Prof. dr. ir. Leon van Noorden

Examination committee: Prof. dr. Hendrik Pinxten (chair)
                              dr. Michiel Demey
                              Prof. dr. Giovanni De Poli
                              dr. Micheline Lesaffre
                              Prof. dr. Francis Maes
                              Prof. dr. ir. Jean-Pierre Martens
                              Prof. dr. Maximiliaan Martens
                              Prof. dr. ir. Leon van Noorden

Thesis submitted to fulfill the requirements for the degree of

Doctor in Art Sciences

Academic year 2010 – 2011

Table of Contents

List of Figures

List of Tables

English summary

In this study, the development of a framework in order to enable a better understanding of the relation between human movements as a response to music and the concepts of intentionality and expressivity is presented. The methodology is based on the concept of embodiment (Leman, 2007). This paradigm states that humans are capable of decoding music and translate this into intended expressive movements (gestures). Understanding how people move during a performance (as a performer or as a listener), when listening using media devices or when interacting with devices (computers or sensors) are research topics within the domain of systematic musicology.

During the past decades the musical landscape has changed tremendously. Due to the availability of computers, mobile devices, sensors, ... music has not only become a part of our daily life but it is nowadays also possible to create (compose) and distribute music over networks with common available hardware or to use wireless devices to interact with music. Another important fact is that large collections of music are accessible. In short, at least in our Western culture, everybody can benefit from music. Music is no longer the exclusive domain of a relative small fraction of the population who can afford it to go to concerts and the production of music is no longer limited to the environment of a recording studio. As a consequence people interact more directly with music without limitations in time and place. The interaction between humans and music is known to have an influence on people as music induces emotions, which result in movements. In 1872 already, Charles Darwin understood that there was a link between emotions and movement. In 'The Expressions of Emotions in Man and Animals' this connection between the mental states and the neurological organization of movement was central in Darwin's concept of emotion (Darwin, 1872).

In this work a clear distinction is made between movement and gestures. In the present thesis, movements are considered as displacements in space and

time of so called rigid bodies, while gestures are defined within a general framework that considers them from the viewpoints of communication, control and metaphor.

It is not only important to understand movements. Due to the fact that music induced movements are connected with the concept of gestures, it is also important to investigate this domain as well. The concept of gestures related to music is a complex phenomenon (Godøy & Leman, 2010). Investigating movements belongs to the domain of natural sciences while the research on gestures is mostly part of the humanities. In general a gesture (having a meaning, inducing an aesthetic experience) is positioned between chaotic (completely unpredictable) and redundant (completely predictable) information.

From the point of view of the scientific researcher a systematic methodology to explore the domain of movement and gestures and their relationship is vital to understand the way music plays a role in several domains of human life e.g. music as a social phenomenon, in medical treatment, in sports, in musical performance, in education, in recreation and neuroscience. In the last decade hardware has been developed to measure human movement at high sampling rates resulting in large data sets over several dimensions. Moreover, in the case of human movement on music, the study of movements cannot be seen without taking into consideration other sources of information in a broad range of scientific domains e.g. sociology, psychology, music production, musical content. Finally there is no straightforward methodology for the analysis of human movement as a result of response to music when compared to other domains of statistical analysis as can be found in economics, medicine and psychology.

This makes the analysis of human movements in the musical context and the relation with gestures within the context of systematic musicology inherently multi-disciplinary and is a real challenge in science today.

This thesis is organized as follows. In the introduction (chapter 1) the goal of this thesis is explained and motivated. My objective is to establish an empirical methodology for embodied music cognition research, and to prove that this

framework can contribute to handle core problems regarding musical communication: I particularly want to understand two essential ingredients of musicality (expressiveness and intentionality) in their relation to body movement.

The next chapter (chapter 2) describes the basic concepts of expressiveness, intentionality and embodiment. Expressiveness and intentionality are considered essential aspects of musicality, whose articulation can be studied from the viewpoint of embodiment.

First the meaning of expressiveness is described, as it is a rather elusive concept with differences in meaning. The approach of Leman (Leman, 2007) is then explained. A major consequence of Leman's approach is that the study of expressiveness necessitates the study of corporeal gestures, and therefore, the development of a new methodology that allows the analysis of gestures.

Second, the concept of intentionality is explained. This concept is based on the simulation of movement patterns. The perception of intentionality is based on a motor system that allows the simulation of the observed movement with an internal model that allows the prediction of the movement. Intentions in music can thus best be seen as anticipations, or directions towards targets.

I am convinced that expressiveness and intentionality in music can be perceived as cues that facilitate musical engagement (listening, playing, dancing). Instead of perceiving them as deviations from the neutral form (in the case of expressiveness) or perhaps as deviations from observed movement (in the case of intentionality), both expressiveness and intentionality are accepted as the core foundation of musicality and therefore of musical communication.

Third, the concept of embodiment is described. The theory states that our interaction with music is embodied, and thus is mediated by the human body. It is through corporeal articulations that experiences can be encoded in musical sound (through music playing). It is through corporeal imitations that musical sounds make sense. The concept of a repertoire of gestures, or corporeal articulations, also called the action-oriented ontology, is an essential component of the theory.

In our study, the musical communication model is accepted, including the a-modal concept of gesture as the basis for our study on expressiveness and intentionality. I focus on sound-facilitating gestures, as these are the gestures that display the expressive and intentional features of musicality. I show that the methodological framework developed in this thesis can be applied to the study of expressiveness and intentionally in a way that is in full agreement with the musical communication model. As such, the study of embodied music cognition becomes fully embedded within a modern empirical approach.

In the last section of this chapter attention is given to the evolution from a more general towards a more local approach. It was found that this approach to expressiveness and intentionality requires a methodology that combines both a bottom-up and a top-down approach. The bottom-up approach is concerned with the observation of body movement, which is based on motion capture and sensing technologies. The top-down approach is concerned with the identification of the local targets of a particular performance, which is based on the performer's introspective analysis.

Chapter 3 deals with a general description of the framework proposed in this work. I found that up till now no framework was established to consider movements and combine movement analysis based on the methods used in the natural sciences with the results of the interpretation from the viewpoint of the humanities. During my research, it became apparent that statistical analysis can contribute to connect natural sciences and the humanities.

Chapter 4 reviews the data sets used in this work. In this chapter the data sets used in this thesis are presented as relational sets. As a consequence of the research topic each data set is multi-modal and contains time series based on human movement, captured by means of different methods. The challenge of this work was to define and develop strategies in order to investigate the relations between the different modalities. In all data sets time series of movement data are the vital clue to define the strategies and methods, which are explained in the next chapter.

Chapter 5 handles about the methodology developed throughout this work. First an overview of the evolution of data sets and the soft- and hardware is given. During the course of the thesis work, there was a substantial development of data sets, as well as of software and measurement methods. This had a great impact on the realization of the framework. Starting from relative small data sets, basic soft- and hardware and basic analytical tools the possibilities were gradually expanded throughout this work.

Second, attention is given to the importance of exploratory data analysis as a mandatory first step in statistical analysis. By a carefully inspection and exploration of the initial data errors, missing values and outliers can be avoided. In the case of nominal data attention is also given to the importance of normality and a method is given to deal with data deviating from the normal distribution.

The third section deals with the importance of variability. A general variability model for the analysis of human movement in response to music is proposed. The model relies on a mapping of the sources of variability as a result of task constraints, human variability, aggregation, sonic constraints and dynamic re-parameterization.

The next section contains selected examples of methods used in this work. For each method an example from the data sets used is added.

The fifth section deals with the statistical paths used in this work. This approach enables to select appropriate methods in order to investigate the research questions (hypotheses). The paths for the analysis of all data sets used in this work are presented and described.

Finally, in the last section of this chapter, the establishment of the framework is presented. First the concepts of the top-down approach are described and in the next section the methodology to link the bottom-up and top-down approaches is given.

In chapters 6, 7, 8 and 9 the texts of four papers realized in this work are added.

In the first paper, sharing musical expression through embodied listening is presented. The analysis reveals that the listeners' movement velocity patterns tend to correlate with each other, and with the movement velocity patterns of the player's shoulders. The findings support the hypothesis that listeners and player share, to a certain degree, a sensitivity for musical expression and its associated corporeal intentionality.

The second paper examines the effect of music on human movement in response to music with respect to group interactions. A method was developed to quantify group similarity and the results revealed that synchronicity increases significantly in a social context. Furthermore, it was shown that the type of music is the predominant factor on how people move to musical stimuli.

The third paper gives the results of a study of a musical performer. Using a combination of principal components analysis and analysis of variance it was possible to uncover measures that reflect coordinated action.

The fourth paper describes the establishment of the framework. By combining a top-down analysis based on annotations made by the performer and a bottom-up method based on segmentation of kinematic data it was shown that the mental focus on musical targets is related to bodily expression.

Finally in chapter 10 a discussion is given followed by the general conclusions.

Nederlandstalige samenvatting

In deze studie wordt de ontwikkeling van een onderzoekskader voorgesteld met het doel een beter inzicht te verwerven in de relatie tussen de menselijke bewegingen als respons op muziek enerzijds en de concepten van intentionaliteit en expressiviteit anderzijds. De methode is gebaseerd op het theoretisch model van lijfelijkheid (Leman, 2007). Dit paradigma stelt dat de mens in staat is om muziek te decoderen en om te zetten in betekenisvolle expressieve bewegingen (gebaren). Onderzoek over hoe mensen bewegen tijdens een uitvoering (als uitvoerder of als luisteraar), bij het luisteren wanneer ze media apparatuur gebruiken of in interactieve toepassingen (computers, sensoren) zijn onderzoekthema's binnen het domein van de systematische musicologie.

Gedurende de laatste decennia, is het muzieklandschap enorm veranderd. Dankzij de beschikbaarheid van computers, mobiele toestellen, sensoren en dergelijke is muziek niet alleen een deel van ons dagelijks leven geworden maar biedt dit ook de mogelijkheid om muziek te creëren, te verspreiden over netwerken door middel van toegankelijke hardware en om interactief te reageren op muziek door middel van draadloze apparatuur. Een ander belangrijke vaststelling heeft betrekking op de toegankelijkheid tot grote verzamelingen muziek. Vandaag is muziek voor bijna iedereen bereikbaar, zeker in onze Westerse cultuur. Muziek is niet langer het exclusieve privilege van een relatief kleine bevolkingsgroep die concerten kan bijwonen en registratie van muziek is niet meer beperkt tot de omgeving van een opnamestudio. Als een gevolg daarvan gaan mensen muziek meer interactief gebruiken, zonder beperking in tijd en ruimte. Van de wisselwerking tussen mens en muziek is bekend dat deze een invloed heeft op de mens, omdat muziek emoties teweeg brengt en dit in beweging resulteert. Reeds in 1872 begreep Charles Darwin dat er een verband was tussen emoties en beweging. In het boek "The Expression of Emotions in Man and Animals" stond de connectie tussen de mentale toestand en de

neurologische structuur van de beweging centraal in het concept van Darwin over emoties (Darwin, 1872).

In mijn thesis wordt er een duidelijk onderscheid gemaakt tussen beweging en gebaar. Bewegingen worden beschouwd als verplaatsingen in tijd en ruimte van starre lichamen, terwijl gebaren gedefinieerd worden binnen een algemene structuur, gebaseerd op de begrippen communicatie, controle en metafoor.

Als gevolg van dit onderscheid is het dus niet enkel belangrijk om bewegingen te begrijpen. Doordat de bewegingen die door muziek veroorzaakt worden door intentionaliteit en expressiviteit verbonden zijn met het concept van gebaren, is het noodzakelijk om dit domein te eveneens te onderzoeken. Dit moet dan de mogelijkheid bieden om het verband tussen bewegingen en gebaren te onderzoeken. Het concept van gebaren gekoppeld aan muziek is een complex fenomeen (Godøy & Leman, 2010). Onderzoek naar bewegingen valt binnen het domein van de natuurwetenschappen terwijl onderzoek naar gebaren deel uitmaakt van de humane wetenschappen. Algemeen wordt een gebaar dat een betekenis heeft en/of een esthetische expressie uitdrukt gepositioneerd tussen chaotische (compleet onvoorspelbaar) en overtollige (compleet voorspelbaar) informatie.

Vanuit het standpunt van de wetenschappelijke onderzoeker is het van belang om, gebaseerd op de empirische wetenschappelijke onderzoeksmethode, het domein van bewegingen en gebaren en hun onderlinge samenhang te onderzoeken. Daardoor kan inzicht bekomen worden over hoe muziek een rol speelt in de verschillende domeinen van het menselijk leven, bijvoorbeeld als sociaal fenomeen, in de geneeskunde, in de sport, in de productie van muziek, in het onderwijs, in de vrije tijd en in de neurowetenschappen. Gedurende het laatste decennium is er hardware ontwikkeld om menselijke beweging met hoge frequenties te meten. Dit resulteert in grote datasets over verschillende dimensies. Bovendien kan men in het geval van menselijke beweging op muziek de studie van de bewegingen niet zien zonder rekening te houden met andere

informatiebronnen in een brede wetenschappelijke context, zoals sociologie, psychologie, muziekproductie en muziekinhoud. Het ontbreekt momenteel aan een eenduidige methodologie voor de analyse van de menselijke beweging als gevolg van een respons op muziek, wanneer men dit vergelijkt met andere domeinen van statistische analyses zoals in de economie, de geneeskunde, biologie en de psychologie.

Dit maakt de analyse van menselijke bewegingen in de muzikale context en de relatie met gebaren binnen de context van de systematische musicologie inherent multidisciplinair en vorm daardoor een echte uitdaging voor de wetenschap van vandaag.

Deze thesis is als volgt ingedeeld. In de introductie (hoofdstuk 1) wordt het doel van deze thesis uitgelegd en gemotiveerd. Mijn doel is om een empirische methodologie te ontwikkelen om inzicht te verwerven over lijfelijke muziek cognitie en om aan te tonen dat dit onderzoekskader kan bijdragen tot de oplossing van cruciale problemen in de muzikale communicatie. In het bijzonder wil ik twee essentiële ingrediënten van muzikaliteit namelijk expressiviteit en intentionaliteit begrijpen in hun relatie tot het bewegen van het menselijk lichaam.

Het volgend hoofdstuk (hoofdstuk 2) beschrijft de basisbegrippen van expressiviteit, intentionaliteit en lijfelijkheid. Expressiviteit en intentionaliteit worden beschouwd als essentiële aspecten van muzikaliteit, waarvan de articulatie bestudeerd kan worden vanuit het oogpunt van lijfelijkheid.

Eerst wordt de betekenis van expressiviteit beschreven als een eerder moeilijk te definiëren begrip met verschillende betekenissen. Vervolgens wordt de benadering van Leman (Leman, 2007) toegelicht. Een belangrijk gevolg van het concept van Leman is dat de studie van expressiviteit de studie van de lichamelijke gebaren vereist en bijgevolg de ontwikkeling van een nieuwe methodologie die de analyse van gebaren toelaat.

Ten tweede wordt het begrip intentionaliteit toegelicht. Dit begrip is gebaseerd op de simulatie van bewegingsmodellen. De perceptie van intentionaliteit is gebaseerd op een motorisch systeem dat de simulatie van de geobserveerde beweging toelaat, met een intern model dat de voorspelling van de beweging mogelijk maakt. Intenties in muziek kunnen dus het best beschouwd worden als anticipaties of bewegingen naar doelen.

Ik ben ervan overtuigd dat expressiviteit en intentionaliteit in muziek als signalen kunnen worden waargenomen die het muzikale engagement (luisteren, spelen, dansen) vergemakkelijken. In plaats van deze waar te nemen als afwijkingen van de neutrale vorm (in het geval van expressiviteit) of misschien als afwijkingen van geobserveerde momenten (in het geval van intentionaliteit), worden zowel expressiviteit als intentionaliteit als fundamenten van muzikaliteit en als gevolg van de muzikale communicatie beschouwd.

Ten derde wordt het begrip lijfelijkheid omschreven. De theorie stelt dat onze interactie met muziek lijfelijk is en dus doorgegeven wordt door het menselijk lichaam. Het is via lichamelijke bewegingen dat ervaringen in muzikaal geluid gecodeerd kunnen worden (door muziek te spelen). Het is door lichamelijke imitaties dat muzikale geluiden zinvol worden. Het begrip van een repertoire aan gebaren, of lichamelijke bewegingen, ook de actiegerichte ontologie genoemd, is een essentieel onderdeel van de theorie.

In deze studie nemen we het muzikale communicatiemodel aan, inclusief het amodale begrip van gebaar als basis voor onze studie over expressiviteit en intentionaliteit. Ik focus me op gebaren die door geluid (muziek) worden opgeroepen, omdat het deze de gebaren zijn, die de expressieve en intentionele functies van muzikaliteit weergeven. We tonen aan dat het methodologisch model uitgewerkt in deze thesis toegepast kan worden op de studie van expressiviteit en intentionaliteit, op een manier die volledig overeenstemt met het muzikale communicatiemodel. Zodoende wordt de studie van lijfelijke muziekwaarneming volledig ingebed in een moderne, empirische aanpak.

In het laatste deel van dit hoofdstuk wordt er aandacht besteed aan de evolutie van een meer algemene aanpak naar een meer locale aanpak. Er werd vastgesteld dat deze aanpak van expressiviteit en intentionaliteit een methodologie vereist die zowel een "bottom-up" als een "top-down" aanpak combineert. De "bottom-up" aanpak houdt zich bezig met de waarneming van de lichaamsbeweging, die gebaseerd is op de technologie van het vastleggen van bewegingen. De "top-down" aanpak houdt zich bezig met de identificatie van locale doelen van een bepaalde prestatie, die gebaseerd is op de introspectieve analyse van de uitvoerder.

Hoofdstuk 3 handelt over een algemene beschrijving van het model dat voorgesteld wordt in dit werk. Ik stelde vast dat er geen bestaand model bestond om bewegingen in aanmerking te nemen en wetenschappelijke bewegingsanalyse te combineren met subjectieve data vanuit het oogpunt van de humane wetenschappen. Tijdens mijn onderzoek werd het duidelijk dat de statistische analyse gebruikt kon worden om de natuurwetenschappen met de humane wetenschappen te verbinden.

Hoofdstuk 4 beschrijft de datasets die in dit werk gebruikt zijn. In dit hoofdstuk werden de datasets gebruikt in deze thesis voorgesteld als relationele sets. Als een gevolg van het onderzoeksonderwerp is elke dataset multimodaal en bevat het tijdreeksen gebaseerd op menselijke beweging, vastgelegd door middel van verschillende methodes. De uitdaging van dit werk was om strategieën te definiëren en te ontwikkelen om de verhoudingen tussen de verschillende modaliteiten te onderzoeken. Het is duidelijk dat analyse tijdsreeksen van bewegingsdata hierbij centraal staat in dit werk.

Hoofdstuk 5 handelt over de methodologie ontwikkeld tijdens dit werk. Eerst volgt een overzicht van de datasets, sofware en meetmethodes. Gedurende de loop van mijn thesis was er een beduidende ontwikkeling op het vlak van datasets, software en meetmethodes. Dit had een grote impact op de realisatie van het onderzoekskader. Startend van relatief kleine datasets, eenvoudige software en meetapparatuur werden de mogelijkheden geleidelijk uitgebreid.

Ten tweede wordt er aandacht besteed aan het belang van verkennende data analyse als een verplichte eerste stap in statistische analyse. Door een voorzichtige inspectie en exploratie van de initiële fouten in de data kunnen ontbrekende waarden en uitschieters vermeden worden. In het geval van nominale data wordt ook het belang van de normaliteit onderlijnd.

Het derde deel gaat over het belang van variabiliteit. Een algemeen variabiliteitmodel voor de analyse van menselijke beweging in reactie op muziek wordt voorgesteld. Het model is gebaseerd op het in kaart brengen van de bronnen van variabiliteit als gevolg van de beperkingen van de taak, menselijke variabiliteit, aggregatie, sonische beperkingen en dynamische re-parametrisering.

Het volgend deel bevat geselecteerde voorbeelden van gebruikte methodes in dit werk. Voor elke methode is een voorbeeld van de datasets toegevoegd.

Het vijfde deel gaat over de statistische paden gebruikt in dit werk. Het gebruik van deze stroomdiagrammen maakt het mogelijk om geschikte methodes te selecteren om onderzoeksvragen (hypotheses) te testen. De paden voor de analyse van alle datasets gebruikt in dit werk worden voorgesteld en omschreven.

Tenslotte wordt in het laatste deel van dit hoofdstuk de ontwikkeling van het onderzoekskader weergegeven. Eerst worden de concepten van de "top-down" aanpak omschreven en dan wordt in het volgende deel de methodologie weergegeven om de "bottom-up" en de "top-down" aanpakken te linken.

In de hoofdstukken 6, 7, 8 en 9 worden de teksten van vier papers die verwezenlijkt werden toegevoegd. In de eerste paper wordt het delen van muzikale expressie door middel van lijfelijk luisteren voorgesteld. De analyse leert dat snelheid van de bewegingspatronen van de luisteraar correleert met elkaar, en met de beweging de snelheid van de bewegingspatronen van de schouder van de luisteraar. De bevindingen ondersteunen de hypothese dat de luisteraars en de speler, tot op een zekere hoogte, een gevoeligheid voor muzikale expressie en zijn gerelateerde lijfelijke intensiviteit delen.

De tweede paper onderzoekt het effect van muziek op menselijke beweging als reactie op muziek rekening houdend met de groepinteracties. Er werd een methode ontwikkeld om de gelijkenissen van de groep te kwantificeren en de resultaten tonen aan dat de synchronisatie beduidend stijgt in een sociale context. Bovendien werd ook aangetoond dat het type van muziek een overheersende factor in welke mate mensen synchroon bewegen op muzikale prikkels.

De derde paper toont de resultaten van een case studie van een muziekuitvoerder. Door een combinatie van principale componenten analyse en variantieanalyse te gebruiken was het mogelijk om gecoördineerde acties van een muziekuitvoerder in kaart te brengen.

De vierde paper beschrijft de ontwikkeling van het onderzoekskader. Door een "top-down" analyse gebaseerd op annotaties uitgevoerd door de uitvoerder en een "bottom-up" methode gebaseerd op segmentatie van kinematische data te combineren, werd aangetoond dat de mentale focus op muzikale doelen gerelateerd zijn met lichamelijke expressie.

Tenslotte wordt er in hoofdstuk 10 een discussie gegeven, gevolgd door algemene conclusies.

1 Introduction

Music is a phenomenon that manifests itself in different domains of our reality. Some people may define music as something that they experience as a value, both intellectually and emotionally, or as means to relax, to get excited, to divert attention, to change mood, to come into contact with other people (such as during dancing), and so on. They may argue that music has a lot of functions related to our subjective experiences. Other people may define music as something that they have on their shelf, or on their iPod: it is a kind of material substrate that encodes physical energy that leads to sound, and consequently to a musical experience. For some people, it may refer to structures, to brain activations, or perhaps to the magic of an audience in a concert hall. The point is that subjective experience and the objective substrate are typically linked to each other, and that music can be seen as a phenomenon that reflects human interaction and experience by physically encoding sounds in an environment.

However, in adopting this broad viewpoint, the question remains how music is actually brought from its physical substrate into our subjective experience, or from our subjective experience into the physical substrate. Stated otherwise, this question is about how people can access the physical environment in which music manifests itself, so that they can make it part of their subjective experience, so it becomes the object of meaning formation, emotional engagement, and social interaction. As a listener, you would typically absorb the vibrations of musical sounds and transform them into experiences (perception). As a musician, the conversion of vibrations into experiences is linked with the production of these vibrations (action). Our research question is very much related to a better understanding of these processes of action and perception.

The major research question of the present thesis is whether it is possible to capture the essential aspects of the corporeal articulations that reveal and likely

also facilitate the nature of the mediation between subjective experience and physical environment. More specifically, our scientific research question focuses on corporal aspects of expressiveness and intentionality, and on the development of a methodology that makes the scientific study of these aspects possible.

The goal is to investigate whether expressiveness and intentionality can be explained and perhaps even predicted by means of body movements. Along with Leman (2007), we believe that body movements play an important role in musical communication in the sense that they form the basis for the encoding and decoding of sonic patterns that make up music (the so-called "moving sonic forms"). Accordingly, we believe that expressiveness and intentionality have a corporeal basis that becomes articulated in these sonic patterns. By focusing explicitly on the corporeal articulations of expressiveness and intentionality, we hope to be able to contribute to a better understanding of the way in which body movement relates to musicality.

In traditional musicology, both expression and intention form core concepts of our understanding of music (Broeckx, 1981). More recently, they have been reconsidered within an embodied approach to musical meaning formation (Leman, 2007). This embodied approach to musical meaning formation provides a powerful framework for a better understanding of musical communication. However, as mentioned in the final chapter of the book (Leman, 2007), the challenge is to extend the empirical methodology of disembodied music cognition research to an empirical methodology of *embodied* music. The latter requires a methodological framework that reaches further than the currently available methodologies.

Therefore, our goal is to contribute to an empirical methodology for embodied music cognition research, and to prove that this framework can contribute to solving core problems regarding musical communication: we particularly want to understand two essential ingredients of musicality (expressiveness and intentionality) in their relation to body movement. In the

long run, we believe that the analysis of body movement can be combined with audio analysis approaches. This link with audio analysis has not been made in the present thesis. However, as musical communication is a multi-modal phenomenon in which different modalities of our human faculties are combined, we believe that our focus on body movement makes our research innovative. Once a good methodology for studying human movement is found, it will be quite natural to extend the resulting methodology to a combined body-audio framework for musical analysis.

To summarize, our work can be conceived along two parallel research lines. The first research line concerns a better understanding of corporeal expressiveness and intentionality as essential components of musicality. The second research line concerns the development of an empirical framework that allows the scientific study of body movement in relation to musicality.

This thesis is organized as follows. In chapter 2, the concepts of expressiveness, intentionality and embodiment are explained. These concepts form the basic underlying concepts of this work. In chapter 3, the empirical framework, which combines subjective and objective aspects related to expressiveness and intentionality from the viewpoint of body movements, is described. In chapter 4, an overview of the data sets used in this work is given. In chapter 5, the methodologies used in this work are explained. In this chapter, the general workflow is presented first. An overview of the development of data sets, the available software and measurement methods is added as these changed a lot during this research. Before explaining a selection of analytical methods, the importance of data exploration and variability are presented. For each data set, a statistical path is then discussed. Finally, the realization of the framework is then presented. Chapters 6, 7, 8 and 9 contain the texts of four papers. In chapter 10, the discussion is followed by the conclusion.

2 Expressiveness, Intentionality and Embodiment

In what follows, we describe into more detail three basic concepts that form the core of our empirical study, namely: expressiveness, intentionality and embodiment. Expressiveness and intentionality are considered essential aspects of musicality, whose articulation can be studied from the viewpoint of embodiment.

2.1 Expressiveness

Empirical studies of musical expressiveness often evolve around a particular expressive performance style that is associated with an entire (often short) musical piece. Typically, musicians are asked to perform a musical piece according to a number of targeted expressive sensitivities, such as 'light', 'heavy', 'expressive', 'immobile', and, based on recordings, the goal is then to extract objective features from audio and body movement, and correlate them with the expressed sensitivity. Based on this correlation analysis, it is possible to set up an explanatory model. In some cases, it is even possible to set up a control model. For example, Canazza et al., (1997) found that audio features of clarinet performances played in a "light", "heavy" or other style could be connected to a control space that would allow the reconstruction of the characteristic expressive style from a neutral example (Friberg, Sundberg & Frydén, 2000).

Yet, the concept of expressiveness is rather elusive and depending on the source and there may be differences in meaning. In the past, several studies have defined musical expression as the deviation from a mechanical playing of the score (De Poli, 2004). This definition can be linked with Nattiez's semiotic theory of music in which three levels are distinguished: (i) the neutral level, which is conceived as a trace of (ii) the poietic (creation) process and which can be picked up in (iii) the aesthetic (reception) process to reconstruct the message (Nattiez & Dale, 1985). When the notion of expressiveness as deviation is linked

to this scheme, then it is tempting to say that a traditional score contains the trace, or the core, of the musical message without its expressiveness. The score is then formed at the neutral level. The main advantage of this approach is that the expressiveness of a piece of music can be quantified with respect to a neutral reference performance. Audio parameters in the temporal and the spectral domain (e.g. timing, articulation, loudness, sound color) are simply based on the difference between a given (expressive) performance and the reference performance. Using this approach, expressiveness can be straightforwardly captured in rules, and used to control the synthesis of musical scores (Friberg et al., 2000).

Yet, the advantage of working with a score as neutral reference is at the same time also a disadvantage, and problematic, e.g. when scores are not available, as is the case with oral music traditions. Even in cultural traditions based on written traces, e.g. in the Chinese guqin music, there is no written neutral score that reduces the expressiveness to symbols. Instead, there are technical descriptions of how to perform particular gestures, and there are metaphors that suggest the expressive character of the gesture (Henbing & Leman, 2007). Moreover, as pointed out by Leman (2007, p. 143), this theory of expressiveness assumes somehow that the listener accesses expressiveness on the basis of a neutral reference scheme. However, it is not clear how such a neutral reference scheme would form itself in the mind.

Instead, Leman proposes a radical different account, based on the idea that the listener, rather than having a neutral reference scheme, has an expressive reference scheme, based on a repertoire of expressive corporeal gestures that somehow resonate with the gestures contained in the music (moving sonic forms). It is assumed that expressiveness in music is obtained by matching audio patterns (moving sonic forms) with corporeal patterns (expressive gestures). Consequently, in this approach, there is no need for a neutral schema. The

proposed schema is radically expressive and embodied, rather than neutral and disembodied.

A major consequence of Leman's approach is that the study of expressiveness necessitates the study of corporeal gestures, and therefore, the development of a new methodology that allows the analysis of gestures. If musical expressiveness is embodied, then it cannot be studied merely from the viewpoint of audio analysis. It, at least, requires the study of expressive action, or corporeal patterns in relation to components of expressiveness, as well as the study of expressive listening, or mappings of audio patterns into corporeal patterns. The main focus of our research is on expressive action rather than on expressive listening, although aspects of expressive listening have been considered as well.

2.2 Intentionality

Intentionality is concerned with the purposive or goal-directed nature of human movements. Apparently, humans are very sensitive to intentionality, in the sense that our observation of how other people move is often associated with an understanding of this movement in terms of a goal. This is often related to the concept of action. Thus, when somebody performs an action, it is somehow assumed that the observed movements, on which this action is based, are goal-directed or intended. Action can thus be defined as goal-directed movements.

In the past, intentionality has been considered in relation to a theory-of-mind theory. This theory states that the goal of an observed movement can be understood by adopting the mental perspective or theory of the other (Carruthers & Smith, 1996). Interestingly, several studies have pointed out that the connection between movement and intentionality is stronger when the movement patterns have a human biological origin (Pollick, Paterson, Bruderlin & Stanford, 2001). This suggests that the teory of mind cannot be considered indepently from bodily behavior. Therefore, an alternative theory has been formulated, based on

the simulation of movement patterns (Gallese & Goldman, 1998). This theory states that the perception of intentionality is based on a motor system that allows the simulation of the observed movement with an internal model that allows the prediction of the movement.

Leman (2007) has pointed out that the principle of understanding movements of other humans in terms of their intentions also applies to music. A particular musical phrase can thus be considered as a sonic movement, of which the perception can be linked to a goal-directed movement. However, it is possible that the perception of sonic movement relates to anticipations at more local levels, thus pointing to the directionality of a goal, rather than pointing to a clearly and explicitly defined goal as such. Intentions in music can thus best be seen as anticipations, or directions towards targets and it is likely that our repertoire of gestures, which is built up through interaction with our environment, may help us to predict the direction in which musical gestures will evolve. Cultural knowledge and acquaintance with musical styles may certainly help to understand musical goals more clearly.

In the musical domain, simulation would thus imply that a listener extracts patterns from audio, whose motion is simulated using a motor model or scheme, called the intentionality engine in Leman (2007). Although this bias towards simulation may have a biological explanation, it is likely to assume that the understanding of intentions in music is influenced by cultural habits, and therefore, we also assume that intentionality may rely on a repertoire of learned gestural patterns. This repertoire would thus facilitate the linkage of sonic movements to corporeal movements, for example, through the use of a shared neural code for gestures (a so-called mirror neuron system).

Seen from this perspective, intentionality is closely related with expressiveness. The processes that underlay these ingredients of musicality both tap into the repertoire of gestures that humans acquire during their development from child to adult.

We strongly believe that expressiveness and intentionality in music can be perceived as cues that facilitate musical engagement (listening, playing, dancing). Instead of perceiving them as deviations from the neutral form (in the case of expressiveness) or perhaps as deviations from observed movement (in the case of intentionality), we accept both expressiveness and intentionality as the core foundation of musicality and therefore of musical communication. Music stripped from expressiveness, of which the movement patterns have no gestural component (e.g. such as aleatoric music), is likely to be more difficult to understand. Such music doesn't correspond to our repertoire of gestures that is naturally expressive and intentional. Interestingly, the repertoire of gestures can be assumed to be rather common and stable among people with a similar cultural background. In addition, it is likely that more universal expressive patterns can be discerned, as all people walk, grasp objects, behave emotionally and so on.

2.3 Embodiment

The viewpoint on expressiveness and intentionality mentioned above is deeply inspired by the theory of embodied music cognition, as explained in Leman (2007). The theory states that our interaction with music is embodied, and thus is mediated by the human body. It is through corporeal articulations that experiences can be encoded in musical sound (through music playing). It is through corporeal imitations that musical sounds make sense. The concept of a repertoire of gestures, or corporeal articulations, also called the action-oriented ontology, is an essential component of the theory. As suggested in the previous paragraph, the notions of musical expression and intentionality can be linked with this gesture repertoire. Gestures are assumed to facilitate the interaction with music.

Recent work on the concept of musical gestures shows that it is a promising and useful concept for embodied music cognition (Godøy & Leman, 2010). First of all, a musical gesture can be conceived from the viewpoint of

sound as well as from the viewpoint of body movement. In other words, the concept of gesture is a-modal: it does not belong to one single modality, but to at least two different modalities, namely, sound and body movement. The a-modal nature of a gesture corresponds well with Leman's (2007) viewpoint on musical communication, which states that music can be regarded from the viewpoint of an encoding and decoding of gesture.

The model consists of a player and a listener. Thus, when playing music, the player uses body movements to encode sound patterns. This encoding will be realized in such a way that the gestural nature of the body movements, based on the action repertoire of the player, will leave traces in music. These traces can be characterized by means of an analysis of sonic temporal and spectral characteristics. The sonic result of the players' movements can be recorded and sent to the listener. According to this musical communication model, the listener will decode the music, using his/her own action repertoire. Interestingly, the action repertoire of the player and the action repertoire of the listener need not be the same. The assumption is that the corporeality of humans provides the commonality that is needed to make this non-verbal communication meaningful. The theory states that a successful decoding may establish an intentional communication layer which functions on top of a physical communication layer. The intentional layer reveals the goals of the movement, while the physical layer reveals the movement patterns.

Jensenius et al. (2010) have pointed out that the encoding of musical gestures in sound may be based on different types of body movements. A useful distinction is made between sound-producing gestures, communicative gestures and sound-facilitating gestures:

- Sound-producing gestures are those that effectively produce sound. They can be further subdivided into gestures of excitation and modification.
- Communicative gestures are mainly used among musicians or among musicians and audience to communicate.

- Sound-facilitating gestures support the sound-producing gestures in various ways. They can support phrasing or follow the contour of sonic elements.

In our study, we adopt the musical communication model described above, including the a-modal concept of gesture as the basis for our study on expressiveness and intentionality. We focus on sound-facilitating gestures, as these are the gestures that display the expressive and intentional features of musicality. We will show that the methodological framework developed in this thesis can be applied to the study of expressiveness and intentionally in a way that is in full agreement with the musical communication model. As such, the study of embodied music cognition becomes fully embedded within a modern empirical approach.

2.4 Towards an empirical framework based on the study of expressiveness and intentionality

When studying expressiveness and intentionality, we considered different aspects related to embodied listening and performance. In line with the idea that listeners simulate sonic patterns in music through movements, we studied how subjects move in response to music. In our study on dance, we found that listeners move in a synchronized way with music, both individually and in a group. However, the synchronization is stronger in the group condition, moreover, movements are more intense. In our study on the guqin music, we examined whether subjects can express the music they hear through swaying, and whether they do that in a consistent way (within subject) and in the same way as other subjects. These studies revealed that expressiveness can indeed be studied on the basis of an empirical framework.

Studies on expressive listening were followed by studies on expressive performance. We first focused on the expressive performance of a guqin player.

In that paper, we introduced the idea that expressiveness is based on the coordination of movements of body parts (head, left arm, and right arm) and we then looked at the concept of intentionality in relation to head movements and the related coordination of other body parts.

Our explorations in expressiveness and intentionality have then been redefined in a large-scale study on clarinet playing. This study is based on an entirely new idea according to which expressiveness and intentionality are studied in relation to local musical targets. We believe that this study contributes in an important way to solving the problem of expressiveness and intentionality.

Until recently, most studies focused on expressed sensitivities in terms of general stylistic targets. Less attention is thereby given to the way in which particular local sensitivities, or local musical targets, are communicated.

Several studies have addressed musical playing in relation to intended emotions or expressions, such as playing in a sad mood or gay mood, or playing 'heavy' or 'light'. For example, Canazza et al., 2004; studied audio patterns of different intended expressive styles and compared them with a neutral playing style. Wanderley et al. (2005), studied body movements during standard clarinet performances (as played in concert) opposed to exaggerated expressive and immobile performances. In similar studies, the expression of a piano performance is discussed (Castellano et al., 2008; Clarke & Davidson, 1998). All these studies have in common that expressiveness and intentionality are linked to global targets.

In line with the general idea that musicians seem to move in correlation with local sensitivities in moving sonic forms, Davidson (2005; 2007) observed that the performer uses particular movement shapes that refer to specific and identifiable expressive locations within the context of a whole performance. When studying pianists, these shapes are accounted for by the notion of an "expressive centre of moment", which is the physical centre within the body

through which the musically expressive information is produced by more extended body parts. The study suggests ways of addressing intended expressive sensitivities at a local structural level, rather than at a global level that is associated with a performance style.

However, many questions regarding the precise relationship between expressive sensitivities, musical intentions, and body movement need to be clarified and investigated further, especially also in view of a refinement of the musical communication model.

In the course of this thesis, we evolved from more a global approach on expressiveness, to a more local approach on expressiveness. The latter implies the study of how expressiveness and intentionality are related to local musical targets that is, the performer's focal points within the expressive discourse of music playing. For example, these focal points may relate to different parameters of the musical structure, such as long notes, high notes, notes related to the tonal structure, or rhythmic structure and so on. We, thereby, assume that, based on these locally intended targets, the performer's gestural skills subsume a grouping of movements into larger movement patterns or gestures (Godøy & Leman, 2010). As such, the locally intended targets can be understood as reference frames for musical motility, which reduces the cognitive load during the musical performance. Without such targets and grouping, one would focus on all notes equally as much, and therefore it is likely that there would be more cognitive load, which would make a musical performance more difficult to understand and certainly less musical (as it is with novice players). Therefore, a focus on targets and grouping in relation to musical information and gesture is a first step towards a better understanding of how possible reference frames for musical motility (the gesture-based repertoire) may facilitate music playing.

However, this approach to expressiveness and intentionality requires a methodology that combines both a bottom-up and a top-down approach. The bottom-up approach is concerned with the observation of body movement, which

is based on motion capture technologies. The top-down approach is concerned with the identification of the local targets of a particular performance, which is based on the performer's introspective analysis. Moreover, one should keep in mind that local targets are present in any expressive performance style, no matter whether this style is 'light', 'heavy', or 'exaggerated expressive'. This approach, thus, requires the collaboration of a performer who assesses his of hers own performance in terms of local intentions, and chunked gestures. This approach supports the general idea that musical targets can be understood as intended goals that result in sound through the mediation of expressive gestures.

The main purpose of this thesis is not to come to a unified theoretical or methodological approach, but to set out a range of different analytical tools and statistical methods, which can be used for systematic musicological research, focusing particularly on those methods that offer potential for further development and that propose a framework. This enables us to combine objective and subjective data in music research. By doing so, a contribution is made to the cognitive revolution in the humanities and the sciences (Honing, 2006). The transformation from music as an art form to music as a process involving human actions in a broad spectrum of domains is especially important in the field of music research (Huovinen, 2006). It presents us with the challenge to develop empirical methodologies that enables us to better understand not only the musical phenomena but also the relation between humans and music.

To summarize, our empirical study resulted in a new approach to study expressiveness and intentionality in music, based on the notion of local targets. Its analysis requires a top-down analysis of local targets, based on the performer's introspective analysis.

3 Empirical Framework

In this thesis, we develop an empirical framework that can include both subjective top-down data, as well as objective bottom-up data. This combination of subjective and objective aspects relates to the nature of the topic, namely the study of expressiveness and intentionality from the viewpoint of body movements. Such a framework should also be understood within the broader scope of the discipline of systematic musicology, its relationship with the humanities and the natural sciences, and the different research centers that are active in this research domain.

Systematic musicology is an interdisciplinary science, comprising a broad spectrum of sub-disciplines from humanities and sciences (Parncutt, 2007). However, due to the high degree of specialization, and the fundamental differences between humanities and the sciences, it is often felt that there is a serious gap between humanities and sciences. This makes real interdisciplinary research difficult, and often makes it the subject of prejudices. The discussion is known (Snow, 1993), and the debate is still going on today (Van Bendegem, 2009).

Until recently, musicology lacked an empirical framework that could cover multi-modal data. Subjective assessments were mostly combined with audio analysis (music information retrieval). However, no framework was established to consider movements and combine scientific movement analysis with subjective data from the viewpoint of the humanities. During our research, it became apparent that statistical analysis could be used to connect natural sciences and the humanities. Although in the last decades, the use of statistical analysis methods has become more common and accepted in several sub-disciplines of systematic musicology, it was found that a general empirical framework did not exist in this domain. This is especially the case regarding research on human movement in response to music, when compared with, for instance, movement

analysis in sports and medicine (Stergiou, 2004). Combining top-down (humanities) and bottom-up (natural sciences) is vital in order to establish a two-way communication, in order to cope with current and future challenges presented in the domain of systematic musicology. Scientists should be able to use the less quantifiable and more subjective analysis and findings from humanities, while researchers from the humanities should be able to better understand the benefits of the quantitative approach. A framework that links both the quantitative and qualitative needs would represent an important step forward in multi-disciplinary research.

The present thesis aims at the realization of an empirical framework to (1) make a better understanding of the relationship between movement, gestures and intentionality possible and to (2) provide researchers, regardless of their background, with a tool to analyze data from all kind of experiments. The latter implies that, keeping Ockham's razor blade in mind (Baker, 2010), it was important to focus on the goal rather than on the evaluation and development of analytical methods as such.

'Closing the gap' in the domain of systematic musicology was a big challenge of the research presented in this thesis. Four consecutive steps were taken to fulfill these objectives:

- As a first step, a statistical methodology was developed to enable researchers with little background in mathematics and statistics to analyze the data sets obtained from their experiments. Starting from the hypothetical-deductive model used in empirical science, several basic statistical concepts were combined in statistical paths as a tool for analysis.

- In a second step, the analysis of multivariate time series was investigated to cope with problems such as the absence of a clear repetitive profile and with lag and delay problems.

- In a third step, the analysis of movement and the extraction of features from movement data (*objective bottom-up approach*) were considered. Based on raw motion capture data, kinematic properties are derived from which features could be extracted based on a combination of dimension reduction and the occurrence of minima in acceleration time series.

- In a fourth step, the comparison of extracted features (*objective bottom up approach*) with gestures and annotations (*subjective top down approach*) provided by the performer was established. A framework for further analysis was obtained by combining extracted features, audio, video and animations with annotations by the performer in a multimedia editor and by developing a quantitative method to compare the subjective and objective data.

In order to develop strategies and methods to find solutions for these objectives a detailed inspection of the data sets is mandatory. In the next chapter the data sets are presented and described.

4 Data sets

The work on the empirical framework is based on five data sets that have been generated at IPEM. In chronological order, these data sets are called: (1) Initial data, (2) Movement velocity response data, (3) Group interactions data, (4) Guqin player data and (5) Clarinet player data. Throughout the text the names of the data sets will be used as a reference where applicable. The following abbreviations are used in the description of the data sets (Table 1):

Abbreviation	Description
AN	Annotations
AU	Audio data
BM	Body markers
CO	Experimental Conditions
EM	Experimental measurement
EX	Experiment
GE	Gestures
GR	Grouping
MS	Musical score
PA	Participant(s)
PL	Player(s)
QU	Post experiment questionnaire
SU	Survey (personal background)
VI	Video data

Table 1. Abbreviations used in the description of the data sets

In what follows a description of each data set is provided. For each item in a data set, the number of observations or levels is represented by n. In case of time series, n is multiplied by the time in seconds (t) multiplied by the sampling frequency (fs).

4.1 Initial data (Data 1)

The first data set resulted from an experiment that was set up in 2005 to gather knowledge on embodied listening, using data of how people move along (by swaying their arm) with the music they hear (Styns, 2008). The data set is

represented in Figure 1. It consists of (i) the swaying response of 30 participants to musical stimuli, and (ii) the movements of a musician who played 3 pieces on a guqin (a Chinese Zither). The movements of the subjects were recorded by means of a joystick. The movements of the guqin player were recorded using infrared motion capture cameras. Besides, a pre- and post-questionnaire were filled in by the subjects. Due to the fact that the complete data set was mixed (containing survey data and movement data from the subjects and the player) it was possible to use this set for a preliminary study that focused on the development of statistical paths and tested the validity of different statistical methods.
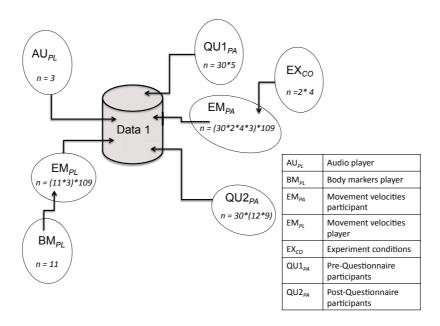
Figure 1. Schematic representation of the *Initial data* (Data 1).

The *Guqin player data* (EM_{PL}) consisted of time series of the magnitude of the velocity of 11 markers (BM_{PL}) for 3 different pieces (AU_{PL}) at a 4 Hz sampling rate for a duration of 27,15 seconds. The pre experiment questionnaire ($QU1_{PA}$) consisted of 5 variables related to personal information (age, gender, number of hours music listening/week, instrument playing and absolute hearing). The post experiment questionnaire ($QU2_{PA}$) contained 9 Likert scale responses

divided in 3 groups that are Knowledge (2 variables: music and instrument), Judgment (5 variables: degree of difficulty, constraint, discrepancy, move/music fit and general judgment) and Comparison (2 variables: Degree of improvement and Effect of information). Each subject responded in 4 subsequent sessions (EX_{CO}) on different stimuli based on the 3 pieces of the player (EM_{PA}). The four consecutive experiments were each time repeated once and designed with increasing addition of information and/or stimuli. The resulting mixed data set (containing nominal, scaled and ordinal data) was very suitable for an initial analysis that explored and evaluated different statistical techniques, which made it possible to establish basic statistical paths.

From the original data (a collection of several Excel files and text files) an MS Access Relational Database was created where all the original data files were combined in one data source ensuring data integrity and the possibility to use queries to extract subsets for more specific analysis.

4.2 Movement velocity response data (Data 2)

The second data set was a subset of the *Initial data* (Data 1) and contained the joystick movement response of the subjects (EM_{PA}) for the repeated four different conditions (EX_{CO}) and the movement data of the guqin player (EM_{PL}) for three musical Pieces (AU_{PL}) and eleven body markers (BM_{PL}). Movement velocity patterns were extracted from both the musician and the listeners in order to test correlations between the listeners and the player and also between listeners. The data set was obtained by means of a query on the preliminary data set, using Open Database Connectivity (ODBC) to import the data directly in SPSS for analysis (Figure 2).
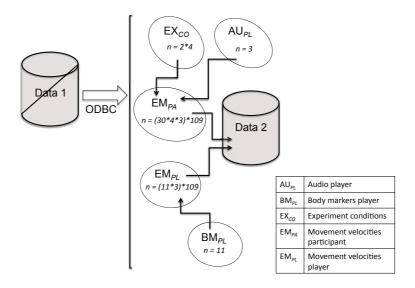
Figure 2. Representation of *Movement velocity response data* (Data 2) extracted from the *Initial data* (Data 1) by means of Open Database Connectivity (ODBC).

4.3 Group interactions data (Data 3)

A third data set (see Figure 3) was obtained from a study that investigated the impact of social interaction on movements made by groups of subjects listening and moving to music (De Bruyn et al., 2008). The data set consists of recordings of the movement of 64 participants (EM$_{PA}$), divided in 16 groups of 4 participants (GR$_{PA}$) in the ecological setting of an exhibition where audiences could watch the performances of these groups.

Each group had to perform the task in two conditions, namely, an individual condition, where the participants were blindfolded, and a social condition, where the participants could see each other (EX$_{CO}$). Each group had to move in response to 6 pieces of music (AU$_{CO}$). Each musical piece lasted about 30 seconds. Registrations of the individual movements were obtained using wireless accelerometers at a sampling rate of 100 Hz. The original data consisted of separate text files by group and condition. The complete experimental design

and the resulting acceleration time series were imported and combined in a relational MS Access database.
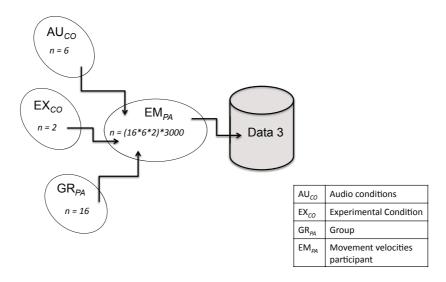
Figure 3. Schematic representation of the *Group interactions data* (Data 3).

4.4 Guqin player data (Data 4)

The fourth data set was also a subset of the *Initial data* (Data 1). The aim was to focus on the movements of the performer. The data set consisted of the magnitudes of the movement velocities of the Guqin player (EM_{PL}) of 11 different joints (BM_{PL}) for 3 different songs (AU_{PL}), see Figure 4. The data set was obtained by means of a query on the *Initial data* (Data 1), using Open Database Connectivity syntax to import the data directly in SPSS for further analysis.
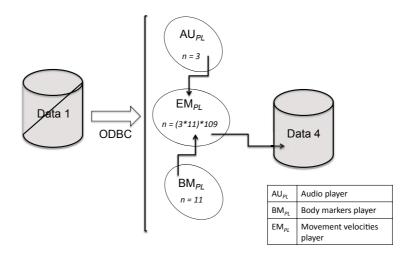
Figure 4. Schematic representation of *Guqin data* (Data 4) extracted from the *Initial data* (Data 1) by means of Open Database Connectivity (ODBC).

4.5 Clarinet player data (Data 5)

The fifth data set originates from an experiment in which a musical performance by a skilled clarinet and viol player was recorded on a motion capture, audio and video system. It was a subset of a larger experiment (Figure 5) involving 4 different musical pieces (MS_{CO}), 3 experimental conditions (PL_{CO}: clarinet solo, viol solo, viol-clarinet duo) and 3 takes (EX_{CO}). For the clarinet solo thirty-three infrared reflective markers covered the musician's full body and clarinet, for the gamba solo twenty four markers covered the upper body and the instrument and for the viol-clarinet duo 57 markers were used (BM_{PL}), recorded at a sampling rate of 100 Hz (*fs*). The length of the time series (*t*) varied from 57 to 196 seconds. In addition to the recorded positions (EM_{PL}) a video recording (VI_{PL}) and an audio recording (AU_{PL}) was made. The complete set (Data V all data) consisted of a collection of video, audio and c3d files.
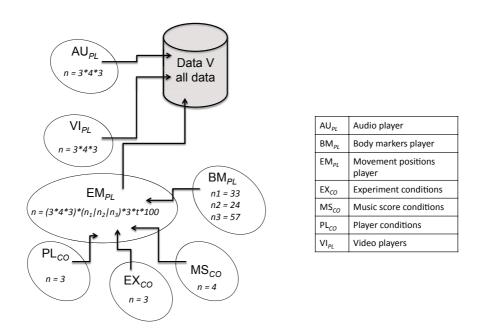
Figure 5. Schematic representation of *music performers' data* (Data V all data).

From this data, one clarinet solo of one musical piece and one take were selected for analysis (Figure 6). The main objective of the analysis based on this data set was to establish a framework for combining objective movement data (using segmentation of kinematic time series) with subjective annotations (based on score (MS_{AN}), audio (AU_{AN}) and video (GE_{AN})). The main reason for selecting only one take and one condition was a matter of time as manual annotations are labor intensive and time-consuming.
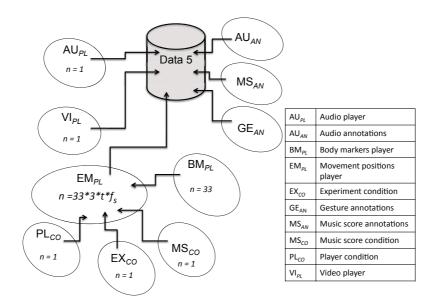
Figure 6. Selected data set for *Clarinet player data* (Data 5) as a subset from *Data V all data*.

4.6 Summary

In this chapter the data sets used in this thesis were presented as relational sets. To summarize it is clear that:

- as a consequence of the research topic each dataset is multimodal.
- each dataset contains time series based on human movement captured by means of different measurement methods (joystick, accelerometer and infrared motion capture camera's).

The challenge of this work was to define and develop strategies in order to investigate the relations between the different modalities. It is clear that time series analysis of the movement data is the vital clue to define the strategies and methods, which are explained in the next chapter.

5 Methodologies

Data sets are the result of one or more underlying hypotheses, the design of the experiment and the actual experiment. This thesis did not focus on formulating hypotheses, design of the experiment & experiments but starts from raw data, which are the outcome of an experiment for which hypotheses were already formulated by researchers at IPEM. We are aware of the fact that defining proper hypotheses, design of experiments and realizing experiments are vital parts of the scientific method, but it was found that the major problems arise when a researcher with little background in statistics and mathematics, and who suffers from what is known as 'statistical anxiety' (Pan & Tang, 2004), has to choose when it comes to which method to use and how to draw conclusions from the obtained results.

5.1 General work-flow
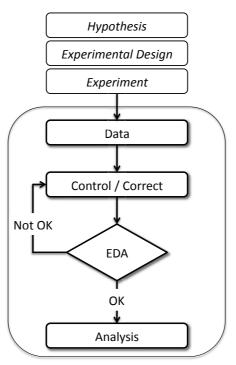
Figure 7. General workflow.

The general workflow of this thesis is shown in Figure 7. The research domain of this PhD is indicated by the rectangle around 'data' to 'analysis'. Starting from a data set from an experiment conducted by researchers working at IPEM, the first step consists of a careful inspection of the data (Control / Correct), the next step is a data exploration by means of Exploratory Data Analysis mainly based on the methodology developed by Tukey (1977). Based on the results from Exploratory Data Analysis, data are corrected and derived (calculated) data can be added. Once the data set is completed an analysis path can be defined in order to test the stated hypotheses.

However, during the course of the thesis work, there was a substantial development of data sets, as well as of software and measurement methods. This had a great impact on the realization of the framework. Starting from relative small data sets, basic software and measurement tools the possibilities were gradually expanded. A brief overview of the main development stages is first described before several aspects of the analytical methodology to establish the framework are presented.

5.2 Development of data sets, software and measurement techniques

The evolution of several levels of the research of the presented work involves multiple measurement methods that aimed at the extraction of various features. An overview of methods applied and the extracted features is given in Table 2.

In what follows, the evolution of the data sets, measurement methods and the software are briefly described.

Data set	Measurements	Features	Methods
Initial data (Data 1)	Pre- and post-questionnaire	Personal background Judgments	Non parametric Discriminant Function
Movement velocity response (Data 2)	Adapted joystick Video	Velocity	Correlation Dynamic Time Warping General Linear Model for Repeated Measures
Group interactions (Data 3)	Wireless Accelerometers Audio Video	Acceleration Intensity Synchronization	Analysis of Variance Dynamic Time Warping Fast Fourier Transform
Guqin player (Data 4)	IR motion capture Video	Velocity	Principal Components Dynamic Time Warping
Clarinet player (Data 5)	IR motion capture audio video Annotations	Kinematics Segments Gestures	Derivation Segmentation Principal Components Classification Score and video analysis

Table 2. Overview of evolution of the data sets, measurements, features and analysis methods.

5.2.1 Data sets

At the beginning of this research data sets were mainly and relatively small tables mostly available in several Excel tables or ASCII files. It was clear that the management of multiple copies of data in several spreadsheets in combination with a lot of labor-intensive management of the data could be a source of errors or even data loss. To avoid this, an improvement was made by using MS Access to combine multiple Excel tables in a Relational Database, enhancing data integrity and make it possible to select subsets of the data by means of queries. Based on the experience with the MAMI database (Lesaffre, 2005), this proved to be a good method for accessing data for statistical analysis. Due to new measurement techniques introduced in the domain of systematic musicology, the raw movement data that are mostly available as text files needed to be combined with other data sources such as audio and video. This resulted in data sets that are more difficult to combine in a single database. A method was applied where

source data are kept in a collection of different sets, which could then be accessed using appropriate software tools. It should also be noticed that, due to the increase of data storage capacities, large data (in the order of Terra bytes) sets can be managed without any problem, which was sometimes the case at the beginning of this work.

5.2.2 Measurement methods

Also the measurement methods available at IPEM changed substantially during the course of time. At the beginning, basic devices (e.g. an adapted joystick), often constructed 'in house', were used in combination with wired data capture using applications such as Pure Data (Styns, 2008). Gradually more sophisticated wireless devices became available. While writing this work, a large amount of hardware was available (e.g. IR motion capture, different types of motion sensors and bio sensors) together with specific software (e.g. Max/MSP, Java). By combining an SQL server with measurement devices it is now possible to capture data online enabling On Line Analytical Processing as a future promising development (Moens, Leman & Moelants, 2010).

5.2.3 Software tools

The software used changed gradually during this work. Starting from Microsoft Excel and SPSS (Statistical Package for the Social Sciences) in a Windows environment. Excel is good for data entry and many people are already familiar with it. The disadvantages of Excel are that it is very easy to enter data without a proper structure, the graphs are clumsy, the statistical facilities are limited and certain numerical routines are unreliable (McCullough & Heiser, 2008). In a Windows environment the latter can be avoided using visual basic for applications.

SPSS is among the most widely used programs for statistical analysis in social science. A great advantage of SPSS is the possibility to paste a selected

procedure directly in a syntax file and store the procedure for later analysis. SPSS has the disadvantage that some modern methods are not included (e.g. Procrustes, Dynamic Time Warping), there is a lot of irrelevant output, it is very difficult to produce customized analysis, the graphics are poor in quality and it is difficult to add customized functionalities. But perhaps the greatest 'danger' of using SPSS is that the menus make it very easy to get 'output' making the possibility of an inappropriate analysis very easy. S-Plus was disregarded because the license of the package at the University was halted and the package was no longer available. The open source package R was not implemented due to the steep learning curve and the not straightforward language and functions for researchers with little background in programming and working with a command line interface.

Although Excel and SPSS are still used for basic analysis and for educational purposes, the working environment for data analysis is nowadays mostly Matlab. The advantage of Matlab is that it enables to write specific functions, develop new mathematical algorithms and combine them in a toolbox, which can be made available for other researchers and is capable of automating complex calculations. Some specific toolboxes in the field of systematic musicology are already available, such as the IPEM-toolbox (Leman, Lesaffe & Tanghe, 2001), the MIR-toolbox (Lartillot & Toiviainen, 2007) and the MOCAP-toolbox (Toiviainen & Burger, 2010). These toolboxes provide a working environment that is currently used by most of the PhD researchers at IPEM. A disadvantage of Matlab is that it is an interpreted language, it consumes a lot of memory and poor programming practices can make it unacceptably slow.

Finally, in order to realize the framework presented in this thesis, a solution was needed to enable multi-modal data editing. The open source package ELAN (Wittenburg, Brugman, Russel, Klassman & Sloetjes, 2006) was found to be able to handle video, audio, tiers containing annotations and generated time

series and proved to be a very valuable tool for gesture analysis. It was found to be an excellent package to combine objective and subjective data.

5.3 Exploratory Data Analysis

When this research started, it was found that in many publications on statistical analysis in the domain of systematic musicology, little attention was given to the importance of data exploration prior to analysis. Even in the rare textbooks on statistics in musicology (e.g. Beran, 2004; Clarke & Cook, 2004), this topic is often neglected or only briefly discussed. Strictly speaking, exploring the data is not an 'analytical' (confirmatory) method and Data Mining is nowadays often used instead of the original Exploratory Data Analysis. Data Mining can be considered as an expansion of Exploratory Data Analysis, which is the first step in this process (Luan, 2002) prior to additional classification methods. The use of the terminology of Exploratory Data Analysis is still common practice in statistical analysis but it is important that it should not be confused with simply summarizing the data.

The general workflow of Exploratory Data Analysis is shown in Figure 8 and is mandatory for each statistical analysis process. Preparing and exploring data sets in this way results in reliable data and avoids the so-called 'trash in – roses out' syndrome. It is obvious that before an experiment can be defined by means of the design of experiments method, the objectives of the experiment in combination with one or more hypothesis have to be clearly described.
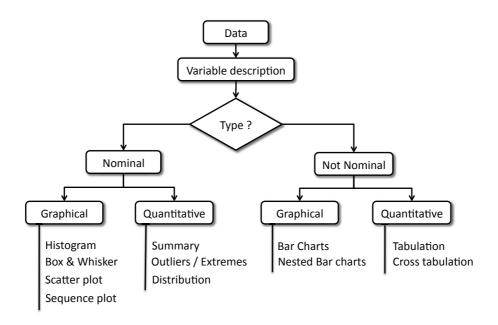
Figure 8. General overview of Exploratory Data Analysis.

In general, the Exploratory Data Analysis approach focuses on a critical evaluation of the data set; detection of outliers and anomalies; getting insight into the data set; variable extraction; uncover underlying structures; modifying variables or adding derived variables from the original data. Exploratory Data Analysis is an approach prior to data analysis and relies mostly on graphical qualitative exploration of the data. The result of Exploratory Data Analysis enables us to control the integrity of the data set, elimination of errors or outliers, develop parsimonious models (Baker, 2010), optimal factor settings and selection of appropriate statistical techniques for further analysis.

The first step in Exploratory Data Analysis deals with a detailed description of the variables. The variable's name, definition, type, limits of the values are summarized in combination with the method used to obtain the data (hardware, experimental setup). A distinction has to be made between the nominal and non-nominal data types (scaled, categorical, ordinal, bipolar). Specific methods for both main data types are divided in graphical and quantitative approaches, which are described below.

5.3.1 Nominal data

For nominal data, the four graphical methods of interest are histogram analysis, Box and Whisker plot, scatter plot and sequence plot. With regard to histogram analysis, it was found however, that the number of bins depends on the application used. In Excel and SPSS, the default number of bins is based on Sturges' rule (Sturges, 1926). However, this may cause some problems especially when the sample size is small (Hyndmann, 1995). Matlab uses a fixed default value of 10 bins but an arbitrary number of bins can be added by the user. In order to obtain consistent histograms (independent from the application) the method of Freedman and Diaconis to make an estimation of the number of bins (Freedman & Diaconis, 1981) was chosen and used in this work (Birgé & Rozenholc, 2006). For Excel, a Visual Basic for Applications program was written. In SPSS the graph editor was used to change the number of bins manually. In Matlab, an m file was written to automatically determine the number of bins and generate the corresponding histogram (the set of m files is available at http://www.ipemcms.ugent.be/?q=PHDFDesmet).

A second graphical method for nominal data is the Box and Whisker plot, which is used to detect extreme values and outliers (Tukey, 1977). The standard method of definition of the whiskers (based on 1.5 below the 25 quartile and 1.5 above the 75 quartile) was used throughout this work.

Thirdly, in case of comparison of variables (if the researcher expects a dependency between variables), a scatter plot is used to look for indications of a relationship.

A last method that was used was a sequence plot (in case of time series data). This graph enables us to look for possible repetitive patterns, possible trends as a function of time or detect anomalies (missing values, outliers, baseline drift, …).

These methods are, however, qualitative and drawing conclusions that are determined by the subjective interpretation of the researcher. The most common method to gather quantitative properties is to give an overview of the data by calculating the mean, median, mode, variance, skewness, kurtosis, inter-quartile range, minimum and maximum value (range) of the variable. Next, the data needed to be reviewed ensuring that there were no outliers and extremes. Another important issue was to evaluate the distributional properties of the data, as deviations from the normal distribution in biomechanical data often occur. Most of the parametric statistical methods have as a pre-requisite that the data come from a normal distribution (N) or that they are very sensitive to deviations from normality, we found it important to propose a method that inspects distributional properties in combination with a quantitative approach to obtain normal distributed data. This method for testing the nature of the distribution in combination with estimation of proper data transformations was added to the Exploratory Data analysis. The methods used, were the Kolmogorov-Smirnov test in case of sampling size > 30 (Massey, 1951) or the Lilliefors test for testing small data sets ($\leq$ 30) when mean and variance are unknown (Lilliefors, 1969) were added to the exploratory data analysis. When deviations from normality occurred appropriate transformations such as simple mathematical calculations (square root, *log*, *ln*) or more complex methods such as the Box Cox transformation (Box & Cox, 1964) and Weibull distribution parameter estimation (Weibull, 1951). The Box-Cox transformation alleviates heteroscedasticity when the distribution of the dependent variable is not known. Using the value of the Weibull shape parameter β enables us to choose an appropriate transformation of the data to obtain data for which the normal distribution assumption can be accepted (see Table 2).

Weibull shape parameter β	approximate distribution
1	Exponential
2	Raileigh
1.25 – 2.75	Lognormal
3 – 4	Normal

Table 3. Weibull shape parameter and corresponding approximate distributions.

An example of the effect of data transformation is shown in Figure 9.
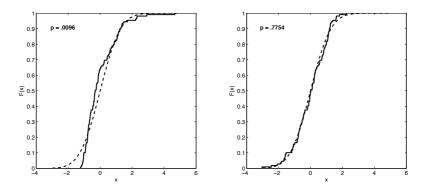
Figure 9. Example of the effect of data transformation on the distribution of data. The example comes from the Guqin player data (Data 4); namely the movement velocities of the head during the performance of a piece of music.

In the graph on the left the cumulative density distribution of the original data (solid line) is compared to the theoretical cumulative density function of the *N*ormal Distribution (dashed line). The p value ($< .05$) reveals that the assumption of *N*ormality is violated. On the right, we can see the result given after a *log* transformation of the data. In this case, the high p value (.775) leads to the conclusion that normality can be accepted after transformation of the data. The choice of transformation is based on the Weibull shape parameter, obtained

by solving the Maximum Likelihood Function of this distribution (Weibull fit). In the example, a value of 1.4 of the Weibull shape parameter β corresponds with an approximation of the Log-Normal distribution, hence the choice of taking the logarithm of the values to obtain data, which can be considered as normally distributed data. Another advantage of using data were the normality assumption can be accepted, is that the generalized extreme studentized deviate test for outliers can then be used as a quantitative method to exclude outliers (Rosner, 1983).

Another important issue is that most of the parametric tests such as the t-test and analysis of variance test (F-test) require homogeneity of variance. To assess variance homogeneity, Levene's test was used in this work (Levene, 1960) in case normality could be accepted. If not a modified Levene test was chosen (Brown & Forsythe, 1974).

5.3.2 Ordinal, Categorical and Scaled data

For ordinal, categorical and scaled data types of data tabulation, cross-tabulation, contingency tables and bar plots are mainly used to explore the data. Depending on the nature of the data scaled variables sometimes need to be recoded in order to obtain more stable, reliable data sets. This was the case in the *Initial data* (Data 1) were recoding and addition of derived (calculated) variables from the post experiment questionnaire was needed to obtain suitable data for analysis and hypothesis testing.
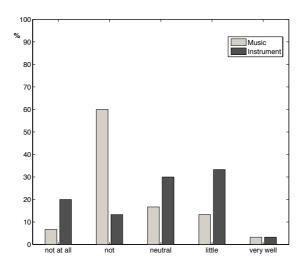
Figure 10. Example of clustered bar chart analysis from the survey data of the *Initial data set* (Data 1).

Quantification of scaled or ordinal variables (association) is possible by means of cross-tabulation and corresponding non-parametric tests of which the 2test is the best known and perhaps the most used. The degree of association can be determined using the φ, the contingency (C) or Cramer's V coefficients. Depending on the type of data and the dimensions of a contingency table other test are available such as the McNemar's test for 2×2 contingency tables with dichotomic matched pairs of subjects.

Comparative quantification (correlation) is possible using Spearman's coefficient calculation. The results of the Spearman method ($\alpha = 0.243$, $p = 0.196$) in the example (Figure 10) proved that there is no significant relationship between knowing the music and knowing the instrument. In this example the 5 point Likert scale responses (in percentage) of the participants to the questions ("Do you know the music?" (music) and "Do you know the instrument?" (instrument)) were compared.

5.4 The Variability Model

The analysis of human movement in the field of music research is a complex phenomenon due to the fact that the structure of variability, a natural feature of the behavior of living organisms, pertains to an interdisciplinary domain (Newell & Slifkin, 1998; Latash et al., 2002). Even in the case of a simple motor task (one person, who has to move within the context of a specific limited task), the potential sources of variability, which will influence the movement response pattern, are complex.

Not only is it important to describe the degree of variability (e.g. in terms of contributions of explained and unexplained variability, due to deterministic processes), but the global structure of variability should also be taken into account. Rather than focusing on a deterministic methodology by describing separate sources of variability, one should keep in mind that there is a broader context containing other sources of variability, which will interact with each other. The theory of embodied music cognition, as explained in Leman (2007), provides a solid base to define a variability framework aiming at a better understanding of the nature of human movements in a musical context.

Traditionally, four domains contribute to variability in human movement response patterns: constraints due to the task, human variability, aggregation and dynamic re-parameterization during the execution of the task. Evidently, in the domain of systematic musicology, the addition of variability due to sonic constraints is inevitable (Figure 11).
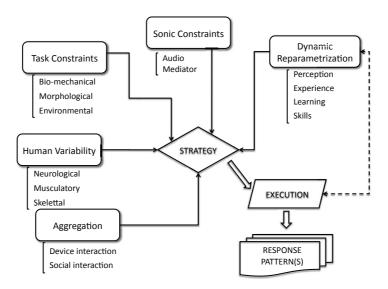
Figure 11. Overview of sources of variability for the development of the variability framework for human movements in systematic musicology. The whole variability framework is embedded in the general musical context.

In general, the total (observed) variability V_T of a system is the sum of the variability due to nonlinear dynamical processes V_n (including possible interactions) and the variability due to error V_e (eq. 5.1):

$$V_T = V_n + V_e \qquad (5.1)$$

Each of the five main domains of variability can contribute to both variability components and are described below, followed by a description of the application of this model to the different data sets.

5.4.1 Task constraints

The constraints due to the task can be subdivided into biomechanical, morphological and environmental conditions and their interactions. For example, the act of playing a violin, will be determined by either bio-mechanical constraints due to the way the instrument is played or morphological constraints if one is asked to sit down or stand up (and is able to move more freely) to play,

and environmental constraints if for instance the task is performed in a recording studio or live on a stage during a festival.

5.4.2 Human variability

Adolphe Quetelet, was one of the first scientists (mathematician) to apply statistical methods to social science (Quetelet, 1835). However, the approach of Quetelet was mainly based on the principle of the 'average man', which of course does not exist. Even up till now, we can find nearly every day in the newspapers examples of this mean approach. When I wrote these lines a newspaper revealed that "The Belgian has 66.000 € savings on his bank account", which is of course misleading and even nonsense without mentioning the variability of bank savings or taking into consideration that there is a great chance that there are outliers and the distribution is not *N*ormal.

Especially when looking at the human body, one has to take into consideration that there is a great variability between humans and that most of the information can be found in the variability rather than a value of the central tendency. Human variability is a combination of the human system (the body) itself and consists of neurological, skeletal and muscular limitations of the persons and their interactions. For example, some pianists have the possibility to reach larger scales on the keyboard due to the fact that they have longer and more flexible fingers (skeletal variability). Other examples are: the difference in movements between a healthy trained young adult and a older untrained adult, who have to move on the same piece of music (muscular variability) and the movements of people who suffered from a stroke (neurological) in response to music.

5.4.3 Aggregation

Aggregation deals with variability, which results from interactions with objects or subjects and is divided into human-device interaction and human-human

(social) interaction (or a combination of both). The device can be an interactive medium such as a motion sensor, a computer or a mediator (musical instrument) producing or influencing music, or even responding to music. It is well known for example that musical performers consider their instrument as an extension of their body. The role of the instrument for musicians is that it is the most natural mediator between subjective experience and physical reality (Nijs, Lesaffre & Leman, 2009). But music is also considered as a social phenomenon and the way people move in response to music is influenced by the presence of other people. Interactions between performers, performers and audience and between listeners are defined as inter-subject relations. The resulting movement variability will be influenced by those factors and may result in aggregation (inter subject-subject, inter subject-device). It means that the movement of a group of individuals will not be a simple addition to the movements of the individuals when moving alone (intra-subject variability).

5.4.4 Dynamic re-parameterization

This is the field of neurological, medical and behavioral science and deals mainly with the way the human brain acts and contributes to the variability of the movement. Aspects of perception, experience, learning and skills are the main issues that contribute to the variability of human movement in this domain. The main difference between this source of variability and the previous three is that there is a continuous feedback while executing a motor task, which results in non-linear dynamic behavior of human movement. If, for example, listening and moving to music arouse a human subject, the brain will respond and the neuro-muscular system will change during the execution. While the factors of the motor task the human variability and the aggregation can be controlled in an experimental design, this is rarely the case for the contribution of the dynamic re-parameterization. Dynamic re-parameterization can be either an unconscious or a conscious process. For example, a subject can deliberately 'play around' (e.g.

improving or even deliberately trying to mislead the researcher) while performing a task or is not aware of the fact that her or his movements change due to anticipation or delay.

5.4.5 Sonic Constraints

Variability due to sonic constraints can be divided into two domains. First of all several features of the sonic form can contribute to the variability in the response patterns. For example tempo, energy (loudness), pitch, tonality, and timbre characteristics of the audio may contribute to differences in movement responses. The second domain is that of the mediator, which can be a musical instrument or any type of device that produces the sound. It should be noted here that variability due to the mediator is not limited to movements of the performers but that the mediator also contributes to the variability of movements of listeners (participants in an experiment, or an audience). Also, the mediator itself can be a source of variance. An example can be found in interactive sound sculptures that allow people to move freely throughout the installation and to create a variety of sounds (Maes, 2009).

5.4.6 Applying the variability framework

Using the concept of variability as a guide, it is possible to refine the general formula of variability (equation 5.1). Depending on the design of the experiment and the hypotheses, the components of variability can be further divided into sub components and appropriate analytical methods can then be chosen to quantify them. To illustrate the benefit of this approach, we give an example based on the *Group interaction data* (Data 3).

A group of four participants was asked to move along with the beat of the music using a wireless accelerometer sensor on a stage in two conditions (blindfolded and while seeing each other) while listening to 6 different musical excerpts of 30 s each. Based on the experimental design of this data set, the

sources of variability can be divided up into variability components as represented in Figure 12. The task constraints can be split up into two variability components: an environmental controllable part $(V_{Ta(Mu)})$, the muscular constraints of the task) and a residual part $(V_{Ta(\varepsilon)})$, unknown biomechanical and morphological constraints and the effect of the presence of public and background noise of the fair). The contribution of the human variability is unknown (residual) as no data of the subjects, such as weight, length, and body mass index, were collected in this experiment $(V_{Hum(\varepsilon)})$. Aggregation can be split up in a component concerning the condition of the experiment $(V_{Agg(cond)}$, individual or social) and a residual component $(V_{Agg(\varepsilon)})$. Dynamic-parameterization can be split up in a component with respect to anticipation and delay $(V_{Dyn(LD)})$ and a residual one due to unknown contributions of skills, learning, … $(V_{Dyn(\varepsilon)})$. There is also a part of the variability caused by the sonic constraints considering that there were 6 different musical excerpts with different tempi $(V_{Son(Mus)})$ and a residual part $(V_{Son(\varepsilon)})$, for instance the effect of short failings of the mediator, which was in this case a wireless headphone. Finally, as it is always the case in any experiment, there is a source of variability, which is completely unknown (V_{Res}).

The response patterns consisted of a set of the magnitude of acceleration time series for each subject - group - condition - musical excerpt combination.
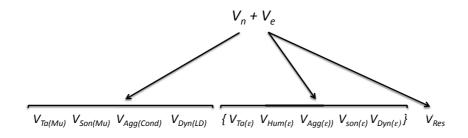
Figure 12. Variability components for the *Group interaction data* (Data 3).

Based on this model of variability, it becomes then possible to find analytical methods to investigate the stated hypothesis. In this case, the hypothesis was that taking into consideration that humans are capable to decode music through embodied cognition there would be a difference in movements due to the condition (social - individual) and due to the type of music. The solution was in this case to first define a proper measure for the movement behavior of each group (which was called group coherence) and cope with the effect of delay and/or anticipation in the acceleration time series (using Dynamic Time Warping). Another benefit of this approach is that the researcher is 'forced' to think carefully about the sources of variability, which cannot be explained and find methods to eliminate possible unknown variability. In this particular case, the time series were for instance tested for stability and drift (deviations from stationarity) in order to remove a fraction of the unknown variability caused by the measurement devices.

The logical structure of the analysis is then applied into so-called statistical pathways, which are described in detail in section 5.6.

5.4.7 Data sets in relation to the variability model

The sources of variability as shown in the variability model (Figure 11) can be separated according to two different approaches: the bottom-up and the top-down approach. Task constraints, sonic constraints, human variability and aggregation belong to the bottom-up domain while dynamic re-parameterization is part of the top-down domain.

The *Initial data* (Data 1) was mainly used to explore and test different statistical methods and can be considered as a try-out for the analysis of the other data sets. In the field of analytical methods for subjective data (dynamic re-parameterization), the methodology was already available (Lesaffre, 2005). Concerning dynamic re-parameterization, some additional methods were investigated. The *Movement velocity response data* (Data 2) belonged to the

variability domain of aggregation and sonic constraints. The *Group interaction data* (Data 3) was situated in the variability domain of aggregation, task and sonic constraints. The *Guqin player data* (Data 4) reflected sonic constraints and human variability. In this way, all different variability domains belonging to the bottom-up approach were investigated to establish this part of the framework. Finally, the *Clarinet player data* (Data 5) was used to combine and expand all the domains used in the bottom-up approach, test the top-down approach and reunite the results in the general framework.

5.5 Selected analytical methods

In this section, an overview is given of the most important statistical and analytical methods used in this thesis. Rather than summing them all up, we choose to describe those methods that were essential for our goal and that were essential in order to deal with specific problems that had to be solved before the framework could be realized. Each method is provided with an example from one of the data sets. The statistical methods that were selected in this work were chosen in such a way that they could act as a bridge between researchers from specialized research domains and researchers from the humanities (musicology).

5.5.1 Non-parametric methods

In the field of non-parametric methods, two methods were used to explore responses from the survey and questionnaire responses from the *Initial data* (Data 1). First of all, a representation of consecutive Likert scale responses was established to obtain a graphical overview, enabling us to interpret complex cross tabulation tables more straightforwardly and to define an index for the complexity of the sequential answers.

Another applied method was the use of Sequential Discriminant Function Analysis as a way to predict ordinal membership of nominal data. There are two advantages to using sequential discriminant function analysis. First, it allows us

to exclude predictors that do not contribute any more than predictors already entered into the equation. Second, as discriminant function analysis is a covariance analysis, it allows us to evaluate the contribution of a predictor variable while diminishing the influence of other predictors.

Most of the methodology in the field of non-parametric analysis was already examined in a dissertation at IPEM (Lesaffre, 2005), which formed a solid framework for the analysis of questionnaire data. In this work, two additional methods were added that are (1) graphical representation of complex cross-tabulated data and (2) sequential discriminant function analysis for group membership prediction.

5.5.1.1 *Graphical representation of complex cross-tabulated data*

Cross tabulation is a well-known method to summarize data. However, when the number of dimensions in the resulting tables increases, it is quite difficult to inspect the result and draw conclusions. From questionnaire responses in the *Initial data* (Data 1), a sequence of two cross tabs from the answers to 3 questions was selected for analysis. The sequence is represented in Figure 13.
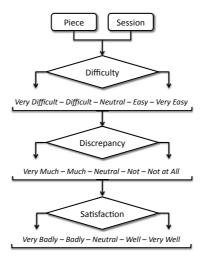
Figure 13. Example of sequential Likert Scale responses from the Questionnaire of the *Initial Data* (Data 1). For each Piece / Session combination 3 questions were answered sequentially.

The first question was "Did you find the task difficult?". The second question was "Was there discrepancy between how you intended to move and how you really moved?" and the third question was "Were you satisfied with the result of your movement?"

When using the traditional way to represent the data (in tables), two separate (20 x 15) tables needed to be inspected without the possibility to link the tables. Creating response pattern plots solved this. A graphical method was used to visualize complex tabulations and make it possible to represent sequential cross tables. In the example shown in Figure 14, one can see that in case of piece 1 and session 2, most of the participants found the task easy, did not find that there was a discrepancy between how they intended to move and how they actually moved and that they were satisfied with the obtained result. The size of the circles corresponds with the percentage of answers in the Likert scales, which are included in the lines between them; the gray circles show the highest scores in the answers. By calculating the number of actual connections and divide this by the maximum number of possible links, it is possible to quantify a complexity index for the sequential cross tabulation data. This index varies between 0 and 1, the higher this value, the more random the answers are.
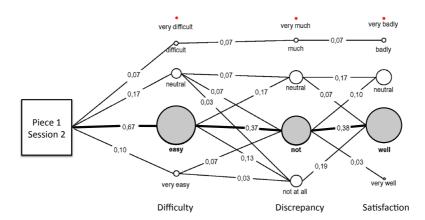
Figure 14. Example of graphical representation of combined sequential cross tabulation data. The graph shows the results for the questionnaire responses corresponding Figure 13 for Piece 1, Session 2 from the *Initial data* (Data 1).

For example, the index of piece 1 session 2 is 0.08, while for piece 3 and session 2 the index has a value of 0.20.

5.5.1.2 *Sequential Discriminant Function Analysis*

In general, discriminant function analysis is basically a classification method and can be considered as the inverse of analysis of variance (Sapatinas, 2005). As the name implies, discriminant function analysis is used in research that wants to predict group membership (dependent variable) from several independent variables. To use a one-way discriminant function analysis, the goal must be to predict group membership (dependent variable) from several independent variables. When priorities are assigned to the independent variables, it is possible to add the independents in a certain order and perform a sequence of discriminant function analysis (stepwise method). The main advantage of this method is that it enables us to exclude unimportant dependents from a data set.
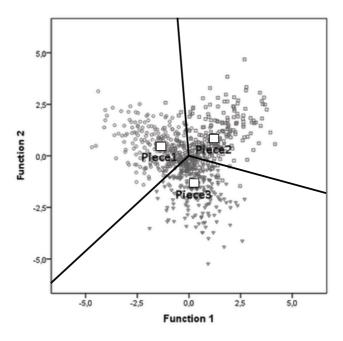
Figure 15. Example of Discriminant Function Analysis from the *Movement velocity data* (Data 2). The figure shows that the movement responses of the participants were different for the 3 pieces of music.

Test of Function(s)	Wilk's λ	χ^2	Df	P
1 through 2	0.241	942.469	218	< 0.01
2	0.522	431.138	108	< 0.01

Table 4. Wilk's lambda test results for the discriminant function analysis represented in Figure 15. The result shows that the movements of the participants from the movement velocity response data (data 2) discriminate significantly for the 3 pieces ($p < .01$).

This method was used for the analysis of *Movement velocity response data* (Data 2). The main objective was to verify that the movement responses of the participants (moving a joystick) was different over the musical stimuli. The results showed that in this case, piece 1 and 2 were discriminated by function 1, while piece 3 was discriminated by contributions of the 2 functions (Figure 15 and Table 4). This result was in good agreement with the differences of the selected pieces in this experiment. Of the three musical fragments, piece 1 and piece 2 have a rather fluent melodic line, which is clearly structured. In contrast, piece 3 has a more narrative character with a less fluent melodic line.

5.5.2 Analysis of Variance

The use of Analysis of Variance and General Linear Modeling for Repeated Measures were found to be valuable tools to estimate effects of factors on experimental data. The use of analysis of variance enabled to look at the effect of musical stimuli with differences in beat and styles in different experimental conditions in the *Group interaction data* (Data 3). In case of the *Movement velocity response data* (Data 2), General Linear Modeling for Repeated Measures was used to estimate effect sizes over time and to quantify learning effects.

5.5.2.1 General Linear Model

Comparisons of mean squares, along with F-tests (or F-like tests) allow the testing of effects of factors. Analysis of Variance is the special case of the general linear model when the effect of one factor is investigated on one variable.

The method is widely used but it was found that in many cases some important issues are either not reported or not investigated. The assumption of normally distributed data, homogeneity of variance and inspection whether the data are balanced over the factors are often not reported and are limitations of the method. Another important issue is the relation between power and sample size. For example: to achieve a power of .80 and a large effect size ($f^2 = .40$) for an *F*-test, a sample size of 76 is required to detect a significant model (Kraemer & Thieman, 1987).

An example of the use of general linear model for univariate measures is given for the *Group interaction data* (Data 3) where the effect of the factors '*Song*' and '*Condition*' on synchronization of the movements with the tempo of the music was tested. In this case homogeneity of variances (Modified Levene test, $\alpha = 0.05$) and normality (Kolmogorov-Smirnov, $\alpha = 0.05$) could be accepted.
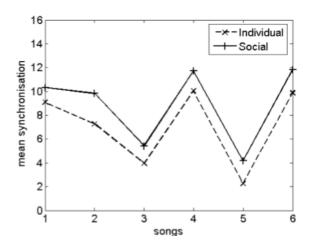
Figure 16. Example of Univariate General Linear Model from the *Group interaction data* (Data 3). In the social condition there is a significant higher synchronization then in the individual condition. The synchronization varies also by the song.

The results show that the synchronization of the participants are significantly better in the social condition compared to the individual condition (*F*-test, $\alpha = 0.05$). This could also be influenced by the kind of song they have to synchronize to (Figure 16). The latter could be illustrated by means of a multiple

comparison Tukey analysis (Tukey, 1951) that leads to the result that participants synchronize significantly lower for songs 3 and 5 than for songs 1, 4 and 6, while the results of song 2 are somewhere in between. This can be explained by the rhythmical complexity of the songs: songs 1, 4 and 6 are pop songs with a very clear beat, songs 3 and 5 can be interpreted either binary or ternary, whereas song 2 can only be interpreted binary but has an unclear beat.

5.5.2.2 *General Linear Model for Repeated Measures*

The key that distinguishes a repeated measures design from the general linear model is that measurement is on the same observational units over time (Goldstein, 2010). Hence, the same subject is measured over time and/or across conditions. In order to investigate the effect of, for example, a learning process, the general linear model for repeated measures is applicable when normality and homogeneity of variance can be assumed. Another reason to use this method is to increase statistical power, especially when the sample size is small. A repeated measures analysis divides the independent variables into '*between-subjects independent variables*' and '*within-subject independent variables*'. This method is especially useful in the field of systematic musicology where learning processes are an important research topic. For example, this method can be used to evaluate the progress of movement performance control of a musician during musical education.

A disadvantage to use a repeated measure design is that it may not be possible for each subject to participate in all conditions of the experiment (time constraints, location of experiment, or a participant who stops before the end of the experiment). Also residual (unexplained) variability can occur and have an influence on the internal validity of the design, namely a regression threat (subjects will tend to a mean in their responses over time), a maturation threat (the subjects themselves may change during the time span of the experiment) and

an outer array threat (events outside the experiment that may change the response of subjects).

An example of the use of general linear model for repeated measures is based on the *Movement velocity response data* (Data 2) (see Figure 17 and Table 5). The same participants were asked to move along with 3 different pieces of music by means of a joystick over 4 sequential sessions. Each session was different as more additional information about the music and the instrument was provided over the sessions. The results showed that there was a difference in the amount of significant inter-subject correlations over the consecutive sessions and between the 3 musical pieces in combination with an interaction of piece*session.

(A) Test for Differences Among Pieces

Source	Type IV Sum of Squares	df	Mean Square	F	p	Partial Eta Squared
Intercept	67294.68	1	67294.68	386.32	.000	.82
Piece	3522.34	2	1761.17	10.11	.000	.19
Error	15154.98	87	174.20			

(B) Test for Differences Among Sessions

Source	Type IV Sum of Squares	df	Mean Square	F	p	Partial Eta Squared
Session	1.50	1	1.50	8.14	.005	.09
Session* Piece	0.33	2	0.17	0.91	.41	.02
Error (session)	16.06	87	0.19			

Table 5. Results of the general linear model for repeated measures analysis of the *Movement velocity response data* (Data 2), revealing significant effect of piece, session and their interaction.
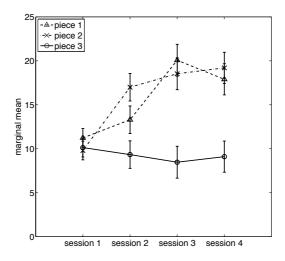
Figure 17. Example of General Linear Model for Repeated Measures from the *Movement velocity response data* (Data 2). The figure shows the effect of the number of significant inter-subject correlations by musical piece over 4 consecutive sessions.

5.5.3 Time series analysis

An important issue was the non-linear dynamic local shifts (due to anticipation and delay), which were problematic when comparison of time series is needed and Auto Correlation Function or Cross Correlation Function methods based on Euclidean distances fail. Applying Dynamic Time Warping solved this problem. This method enabled us also to quantify anticipation and delay, which were found to be very important in the case of human movement responses to music, for example in the *Group interactions data* (Data 3).

In the case of the *Clarinet player data* (Data 5), it was necessary to solve two major problems in the derivation of kinematic data from raw motion capture data (x, y, z positions of markers). First, a method was needed to solve the issue of subjectivity when applying smoothing of time series. Second, a method for segmentation (feature extraction) had to be developed.

5.5.3.1 Dynamic Time Warping

A very important topic in the analysis of human movement is the time dependency of the data. In all data sets of this work, the resulting time series were complex and did not have a clear repetitive pattern as the subjects had a great freedom to move according to there own perception and feeling as defined by the task. This resulted in time series, which were not stable over time due to small dynamic local time shifts. As a consequence, the series were not directly comparable by means of simple Euclidean based correlations, Auto-Correlation Functions, Cross Correlation Functions or Fourier Transform based analysis. Dynamic time warping was found to be a valid method to overcome the problems of small dynamic shifts (due to delay or anticipation) over time. Although the method, which originates from the domain of speech recognition, has been replaced nowadays with the Hidden Markov Modeling approach, we choose dynamic time warping in this study because we focused on the measurement of similarity between time series rather than acoustic modeling or pattern recognition.
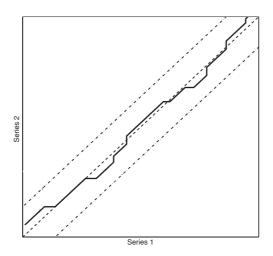
Figure 18. Dynamic time warping within the restrictions of a Sakoe-Shuba window.

In general, it is a computational method that allows us to find an optimal match between two time series with certain restrictions, i.e. the sequences are "warped" non-linearly to match each other (Berndt and Clifford, 1994).

Unconstrained dynamic time warping was found to be too flexible, resulting in an overcompensation of the observed shifts. In order to shift only neighboring peaks in between the series a Sakoe-Shuba window was used (Figure 18). By doing so, band constraints make it possible to restrict the number of vertical or horizontal transitions. The method was, furthermore, expanded with correlation optimization and enabled us to investigate multivariate series (Tomasi, van den Berg & Andersson, 2004). Dynamic time warping was successfully applied in the case of the *Group interactions data* (Data 3), see Figure 19, in order to propose a method for the measurement of social interactions between subjects (social coherence).
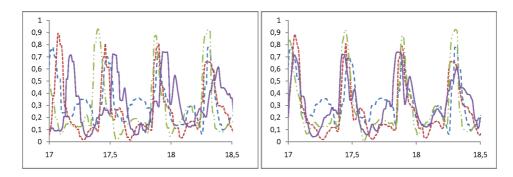
Figure 19. Example of Dynamic Time Warping based on Correlation Optimization of the *Group interactions data* (Data 3). The left plot shows the original normalized magnitude of the accelerations revealing irregular local shifts due to delay and anticipation. The plot on the right shows the result of dynamic time warping.

From the resulting warped time series, it was then possible to use a cross correlation function to define an estimate of the group coherence based on the correlation pairs ($Corr_{ij}$). The similarity between subject movements was defined as:

$$S_{ij} = 1 - \left| Corr_{ij} \right| \qquad \left(0 \le S_{ij} \le 1 \right) \tag{5.2}$$

Low values indicating a high (closer related) inter subject synchronicity. The results of this method enabled us to create graphical representations of the group coherence. An example is given in Figure 19. The length of the lines is proportional to the similarity index.

Figure 20. Representation of group coherence based on the within-group similarity of movements in the individual and social conditions on the same song from the *Group interaction data* (Data 3).

Figure 20 shows that in the case of the individual condition, participant 2 moved differently from the other 3 participants in the group. In the social condition, all participants moved more synchronously with each other, resulting in greater group coherence. The lengths of the lines between the participants are proportional to the similarity of the movements. The position of participant 2 in the individual condition is based on the average similarity with the other participants.

5.5.3.2 *Smoothing kinematic movement time series*

A major improvement for the measurement of human movement was the acquisition of the Infrared Motion Capture equipment at IPEM in 2009. This system enables us to do wireless registration of complete body movements in space at a sampling rate of 100 Hz. Due to the evolution in calculation power, capacity of data storage of computers and the use of Matlab, it was possible to develop a method for the analysis of the kinematics of movement data.

An important issue was the need to develop a smoothing methodology based on a quantitative approach rather than the subjective approach used in most cases found in literature. It is well known that due to derivation of time series with a high sampling rate the series need to be smoothed in order to be able to extract features from them as the noise increases due to derivation. Although a lot of methods are available to perform smoothing, most methods are subjective due to the presence of one or more threshold parameters, which have to be chosen by the researcher. The choice of threshold parameters is, however, subjective and can lead to under- or over-smoothing.

Several smoothing techniques (for an overview see Bowman and Azzalini, 1997) were implemented and tested and it was found that Locally-Weighted Scatter plot Smoothing (LOWESS) (Cleveland and Devlin, 1988) gave the best result. The main reasons for choosing the selected method are that a global function to fit a model to the data is not needed and that it is suitable for modeling complex processes. This technique is based on a locally weighted smoothing with a polynomial order of 1 or 2.

The method is represented in Figure 21. First, a reference dataset was recorded while the player maintained the T-pose (standing upright with arms stretched horizontal outward (a)). Second, the distribution of this recorded data was calculated for the 3D positional data of all markers on the body of the player after subtracting the mean and trend removal. The result is a Gaussian distribution around zero resembling the noise of the system with a variance related to the calibration process (b). Third, a similar distribution from the movement data during the music performance was obtained by applying a series of smoothing thresholds and subtracting the smoothed data from the corresponding data before smoothing (c). When applying too much or too little smoothing to the motion data the resulting distribution after subtraction with the original data (d) will be respectively too wide or too narrow. As a criterion for selecting the smoothing value, a F-test for equality of variances was done on the

distributions obtained from the T-pose and that from the movement data (e). The F-ratio of the reference and motion data noise with a value closest to 1 is then used to select the corresponding threshold value for the optimal smoothing.
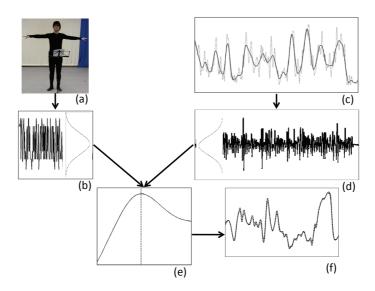
Figure 21. Determination of the optimal smoothing parameter for LOWESS smoothing based on the comparison of residuals of the calibration noise (b) and the noise of the data (c) after smoothing (d). The F-ratio (e) is then used to select the optimum smoothing parameter which is then chosen to obtain the optimize smoothed data (f). The example originates from the *Clarinet player data* (Data 5).

5.5.3.3 *Segmentation and filtering of kinematic time series*

In order to extract meaningful segments from the kinematic time series a method was developed based on the occurrence of minima in the series. A segment was defined as a fraction of the time series between minima occurring in the magnitude of the acceleration enabling us to take the variance of the segment into account, rather than to look at maxima (point estimates). In short, in order to obtain applicable segments, a filter method was applied based on the rejection of gestures, which had a low range of acceleration (elimination of gestures with no pronounced acceleration maximum). In addition, small perturbances in a larger segment, which were observed as "shoulders" or double peaks in the series, were removed. By changing two threshold values in a single function, an optimum of

the variance was then used to select the optimal parameters for the segmentation process.
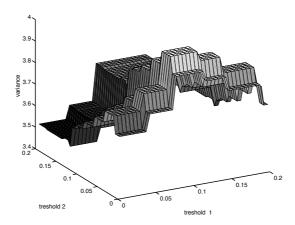
Figure 22. Determination of optimal threshold values for rejecting low range segments (threshold 1) and for small perturbances in segments (threshold 2) based on the optimum of the variance. The example is based on the segmentation process of the *Clarinet player data* (Data 5).

5.5.4 Clustering and classification

All data were multivariate sets with high order dimensionality. Principal Component Analysis was used to reduce the dimensions while preserving the information. This technique proved to be very promising for some of the data sets at hand and was used for instance for the dimension reduction of the *Guqin player data* (Data 4) and the *Clarinet player data* (Data 5).

Several classification methods were evaluated to represent the data in a hierarchical or clustered way. In this context, Procrustes Analysis in combination with Multi Dimensional Scaling was investigated.

Hierarchical Clustering, k-Means Clustering and Multi Dimensional Scaling based on a Procrustes Distance Matrix were tested. Although classification is a very important issue, it was found that classification of the type

of segments/gestures extracted from data based on a great degree of freedom (ecological setting, freedom to move) was difficult, reaching the limit of what is possible with these methods. New methods based on Machine Learning techniques are currently investigated at IPEM.

5.5.4.1 Dimension reduction: Principal Components Analysis

Another important issue was reducing the dimension of the data sets. In this thesis, the Principal Components Analysis was used to extract components to make a reduction of the dimensions of the data sets possible. Principal components analysis involves a mathematical procedure that transforms a number of (possibly) correlated variables into a (smaller) number of uncorrelated variables called *principal components*. The first principal component accounts for as much of the variability in the data as possible, and each succeeding component accounts for as much of the remaining variability as possible. It is obvious that the prerequisites of this method (sampling size, normality, sampling adequacy) were always tested before applying the method.
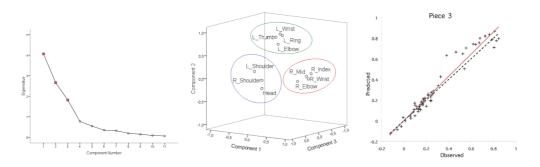
Figure 23. Example of principal components extraction of the player movements of piece 3 from the *Guqin player data* (Data 4). The original dimension of 11 can be reduced to 3 components while maintaining 77,61 % of the variance explained. The figure on the left is the scree plot and shows the eigenvalues of the 11 components. The components plot in the center shows the contribution of the 11 body markers of the player to the 3 extracted components and the right figure shows the observed versus the predicted correlations.

Principle components analysis was, for example, applied in this thesis to reduce the dimensions of the player data from the *Guqin player data* (Data 4) from 11 to 3 (see Figure 23).

The three components correspond with the playing technique of the guqin instrument: a corporal component (head and shoulders), a gliding component (right arm) and a plucking component (left arm).

5.5.4.2   *Procrustes Analysis and Multidimensional Scaling*

Due to the fact that the data sets and the derived properties in this work originate from a wide range of sources of variability, traditional methods to extract similarity measures, mostly based on correlation-like approaches, fail. An attempt was made to consider a set of properties as landmarks of a shape defined in a multidimensional space and to find similarities between these shapes by means of Procrustes Analysis (Goodall, 1991). Although the method is mostly used for real shape analysis found in disciplines as geology and in evolutionary research, a shape can also be based on three-dimensional data matrices containing Likert-type scale responses from a questionnaire (Grice & Assad, 2009) or a set of so called natural classifications used in marketing research (Steenkamp, Trijp & Verhallen, 1989). In this work, a similarity matrix of a set of properties of the segments, which were considered as a set of landmarks of a corresponding shape was used.
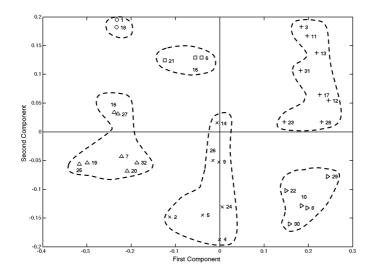
Figure 24. Example of Procrustes analysis. The example shows the result of classification of the Procrustes similarities for a set of properties of 32 segments extracted from the clarinet bell from the *Clarinet player data* (data 5)

Although the Procrustes method is very promising some limitations were encountered in our research. First of all, there is a problem with the selection of properties for an extracted segment from human movement data. A whole set of possible properties can be derived based on statics (e.g. stiffness), position (e.g. displacement, direction), kinematics (e.g. velocity, acceleration and jerk), mechanics (e.g. inertia), dynamics (e.g. force, torque), energy - work and power, shape (e.g. curvature) and complexity (e.g. entropy) and it is unclear which properties to take. Second, it was not clear how to determine the number of clusters and to relate these clusters with the perception of movement. As shown in Figure 24 for a given set of properties, six clusters could be extracted based on two components, which could not be related to the perception of (musical) gestures. A proposal to determine the best estimate for the number of clusters for shape recognition was found in literature (Daliri & Torre, 2009) but it was beyond the scope of this work to investigate this method in depth.

Instead, we adopted the approach proposed in (Naveda, 2011), namely to use Procrustes Analysis in combination with Multi Dimensional Scaling. Non-

metric Multi Dimensional Scaling with a stress minimization based on Kruskal's normalization was chosen.
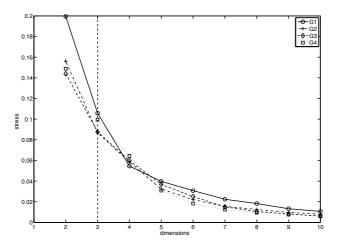
Figure 25. Scree plot of multidimensional scaling of four movement component groups. The graph represents the stress factors as a function of dimension.

The Multi Dimensional Scaling was used in order to project the gestures from the subjective approach onto the multi dimensional segmentation of the objective approach. The resulting scree plot (Figure 25) seems to suggest that group 1 (G1) has a higher stress value than the other groups, indicating more complexity (due to more sources of variance) in the properties of this group. However, the stress values for the four selected groups of movements in three dimensions are between 0.08 and 0.11, which is considered to be a poor to fair fit. We had to conclude that the classification approach did not enable a straightforward grouping of the gestures. It seems that in the current state of our research, we are reaching the limit of the bottom-up approach. To better understand how gestures are related to intentions, it is therefore necessary to rely on a comparison of the objective bottom-up approach with the subjective top-down approach. The top-down approach is discussed in section 5.7.

5.6 Statistical paths used in this work

In this section the statistical paths (flow charts) for each data set are presented. Statistical paths are solutions for the analysis of the stated hypotheses or research questions, which are described briefly. The legend of the items in the charts is given in Figure 26.
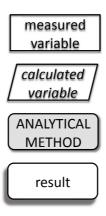
Figure 26. Legend of statistical path items

The use of statistical paths enables the researcher to find solutions for testing hypotheses. The statistical paths are also useful as guides for the non-expert researchers as a support for their analysis.

5.6.1 Initial data (Data 1)

The *Initial data* (Data 1) was divided into three paths. For clarity, only one path is shown, as the two other paths resulted in the *Movement velocity response data* (Data 2) and the *Guqin player data* (Data 4). These are discussed separately in resp. section 5.6.2 and section 5.6.4. The focus on the paths presented in this section is based on the analysis of the pre- and post-questionnaire data of the *Initial data* (Data 1). Two variables from the post-questionnaire were added to the pre-questionnaire (knowing the guqin instrument and knowing the guqin music).

The analysis of the pre-questionnaire data was limited to a descriptive analysis with summaries and charts. The relationship between knowing the guqin instrument and the guqin music was analyzed by means of a Spearman's correlation test, which is the non-parametric alternative for the Pearson correlation coefficient for ranked variables.
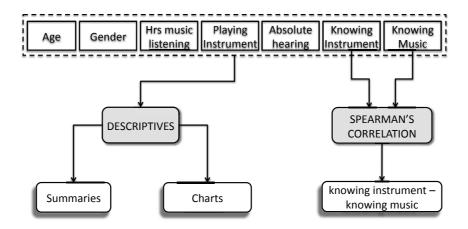
Figure 27. Statistical path for pre-questionnaire data.

For the questionnaire data, the following research questions were formulated:

1. Is there a difference in the questionnaire answers between the music/session conditions?
2. Are there correlations between different questionnaire answers?
3. Is there a possible sequential pattern between questionnaire answers?

The corresponding statistical path is shown in Figure 27.
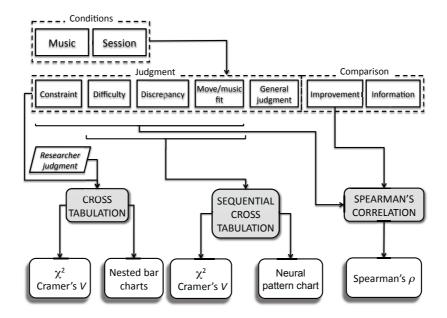
Figure 28. Path for the analysis of the post-questionnaire from the *Initial data* (Data1).

The variables from the post-questionnaire (Figure 28) were divided into two groups (judgment and comparison variables). Furthermore, an additional variable was added by the researcher (researcher judgment), based on a 3 level score after a qualitative inspection of the video takes. All variables were then tested for possible differences between the musical pieces over the sessions by means of different sets of nested cross tabulations. A χ^2 test for establishing differences between proportions (test for contingency) in combination with Cramer's *V* (test for strength of association) enables us to quantify the results. To test whether there is a correlation between the judgment variables and the comparison variable 'improvement', the Spearman's ρ test was used. Finally, sequential cross tabulation in combination with a neural chart representation made it possible to visualize and quantify the judgment variables 'difficulty' – 'discrepancy' – 'movement/music fit'.

5.6.2   Movement velocity response data (Data 2)

The *Movement velocity response data* (Data 2) were used as evidence for the following hypotheses:

1. Listeners share corporeal articulations (and thus the perception of expression) during listening.
2. There is a link between what listeners perceive as expression, and what the player produced as expression.
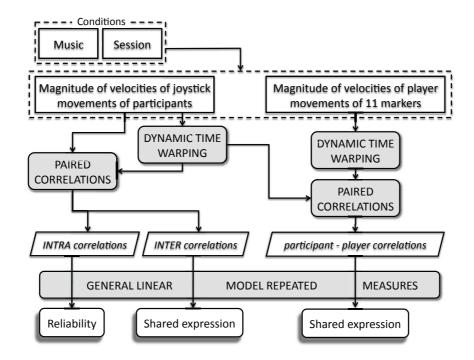
Figure 29. Statistical path for the analysis of the *Movement velocity response data* (Data 2).

In order to test the hypotheses, three variables were derived based on the number of significant paired correlations. One variable was used to examine the intra subject (within subject) correlations, a second variable was defined to inspect the inter subject (between subjects) correlations and a third variable was based on the participant - player correlations. The results of the analysis were published in the paper " Sharing musical expression through embodied listening - a case study using Chinese guqin music" (see chapter 6)

5.6.3 Group interaction data (Data 3)

For the *Group interaction data* (Data 3) the following hypotheses were formulated:

1. People synchronize better with each other in a social context (coherence).
2. The individual and social correlations of the movements are related.
3. Music and the condition are factors that influence the individual movement
4. The intensity of the movement depends on the music and condition
5. The synchronicity of the movement with the music depends on condition.
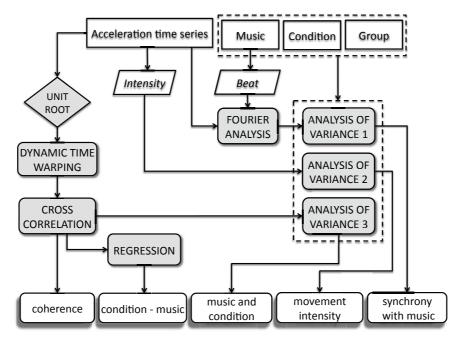
Figure 30. Statistical path for the *group interaction data* (Data 3).

The hypotheses were tested as shown in Figure 30.

The first hypothesis was tested by means of cross correlation function after the acceleration time series were first tested for stationarity (Unit Root test). In case the series were not stationary, trend removal was applied to obtain stable series. The next step was to align the time series for each group, using Dynamic Time Warping.

Figure 31. Screenshots from a video of the *Group interaction data* (Data 3). The picture on the left is the individual condition, on the right the social condition.

When calculating the cross correlations between the original time series of a group of participants, it was found that these were low and that there were no significant differences between the conditions of the experiment. This didn't correspond with what was observed on the video recordings, where differences between the movement responses were clearly observed in the two conditions of the experiment (Figure 31). It was found that small local shifts, which were not constant over time between the series were the main cause why the differences by means of cross correlations of the original series fail to be detected. To solve the problem dynamic time warping with rigid slope constraints and correlation optimized warping were applied. The cross correlations were then calculated from the warped series. This enabled us to quantify differences in correlations, which were not detected without applying warping to the series. The results presented us with the opportunity to define group coherence by establishing the similarity between the movements of participants within the same group.

The second hypothesis was investigated by means of a regression analysis of the individual and social correlations of the groups for each musical piece.

For the third hypothesis, the individual movements were investigated by means of an Analysis of Variance of the cross correlations with music and condition as factors.

In order to test the fourth hypothesis, the intensity of movement was calculated as a new variable. The intensity was based on the jerk (first derivative of the acceleration) of the movements. Analysis of Variance was then used to quantify the effects of music and condition.

Finally, calculating the degree of synchronicity with the music using Fourier Transformation tested the fifth hypothesis. The human-music interaction was studied based on the amount of seconds the participants synchronized correctly with the nominal tempo of the music. This measure of synchronicity was then further analyzed by means of Analysis of Variance.

The results of this analysis were published in the publication "Statistical analysis of human body movement and group interactions in response to music" (see chapter 7).

5.6.4   Guqin player data (Data 4)

The aim of this research is to define appropriate statistical strategies as suitable methods to process movement data related to a music performer's gestures. The following hypothesis was tested: the movements of the joints of a musician's body are driven by a small number of actions. Using a combination of principal components analysis and analysis of variance enabled us to provide evidence for the research question (Figure 32).

In this case, discriminant function analysis was used to inspect if the movement data of the player were different for the piece. A second step in the analysis was to reduce the dimensions of the data by means of principal components analysis.
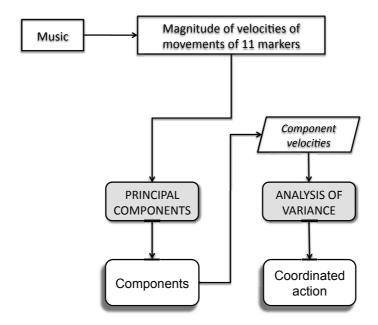
Figure 32. Statistical path for the analysis of the *Guqin player data* (Data 4).

It was found that three components made it possible to divide the movements into three groups of movements according to the body parts of the player. The results were incorporated in a case study (Gesture as coordinated action in Guqin playing, see chapter 8).

5.6.5 Clarinet player data (Data 5)

This data set was used to develop the framework. The first part of the statistical and analysis path discussed in this section focuses on the objective bottom-up strategy. The general path of the bottom up methodology is shown in Figure 34.
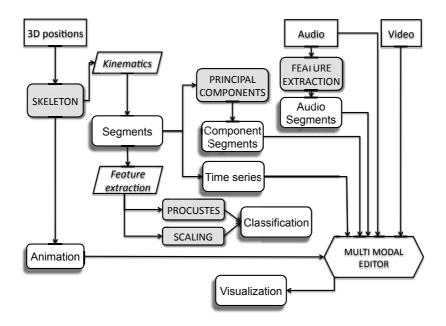
Figure 33. Analytical path of the bottom-up method for the *Clarinet player data* (Data 5).

In what follows, the path shown in Figure 33 is explained in detail. All calculations were done in Matlab R2010a by means of a set of functions written for this analysis. The objective bottom-up approach starts from the raw motion capture data and the corresponding audio file.

1. The 3D positions of 33 markers (full body + the clarinet), captured by means of the IR motion capture system at IPEM at a sampling rate of 100 Hz, were imported into Matlab. Prior to the analysis, the corresponding audio and position data were first synchronized.

2. From the resulting synchronized position time series, the 33 markers were reduced to a set of positions, enabling us to create an approximate skeleton of the player. Using the Mocap toolbox (Toiviainen and Burger, 2010), an animation movie was then created.

3. Based on the 3D positions of the skeleton data, the kinematic data (velocity, acceleration and jerk) were derived and the corresponding norms were then calculated. In order to obtain smoothed time series and

to remove noise due to sequential differentiation the method described in section 5.5.3, subsection 'smoothing kinematic time series' was applied. This resulted in a set of data from which it was possible to extract segments.

4. In order to extract segments, it was necessary to define a method to identify transition points from the kinematic data. These transitions points can then be used to break the sequences of continuous data into shorter gestural segments. The method used to obtain segments is described in section 5.5.3, subsection 'segmentation and filtering of kinematic time series'.

5. Principal components analysis was used to reduce the dimensions of the data set based on two components, which explained 74.3 % of the variance. Component 1 can be related to movements of body parts that follow the movements of the instrument, while Component 2 can be related to the magnitude of the acceleration. Although principal components analysis revealed that there are three clusters of joints, we decided that for further analysis, it is more convenient to consider four groups. We, particularly, make a further distinction within group three between the torso and the legs. By doing so, the results can be more easily compared with the annotations of the player for the movement of the legs (stepping and knee bending). For each cluster of joints, a method was then applied to combine the clusters into a single time series. The method was based on the Gaussian distributions for each transition point using the start and end time of a segment as the mean and an estimate. The resulting separate skeleton time series of the group are combined using a weight based on the acceleration range and a curve fit of the best Gauss curve is then performed to obtain a single time series for each cluster of joints containing the start and end points of gestures found from the movement data as normal distributions.

6. In addition to extracted segments of the motion capture, the audio data were also segmented using the MIR toolbox (Lartillot & Toiviainen, 2007), using the Novelty segmentation method (Foote, 2000).

7. Using the multi-modal editor ELAN, it was possible to combine all resulting data into a single graphical user interface. Video, audio and the animated skeleton were directly imported and synchronized. From the time series of the clusters obtained in step 5, it was then possible to generate tiers. Both the time series and the tiers were then imported into ELAN. A screenshot of the result is shown in Figure 34.
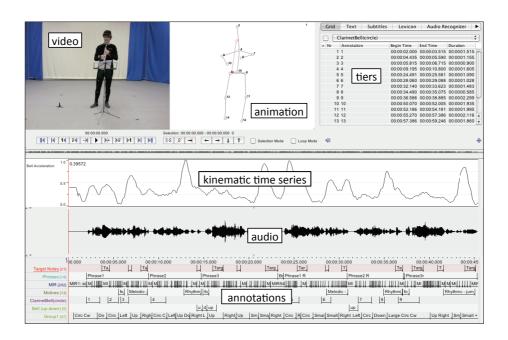
Figure 34. Example of combined data in the multi-modal editor ELAN.

The annotations in the editor can be manual annotations (e.g. from the player), the segments of the audio, separate joint annotations, clustered annotations depending on the goal of the analysis. The editor makes it also possible to select a specific segment and play it in slow motion, which decreases the labor-intensive time of a manual annotation.

8. In order to enable a detailed inspection of an annotated segment (or

gesture), a function was written in Matlab requiring the name of the matlab data variable, the start and end frame of the segment and the skeleton joint, which has the focus. The outcome is a visualization of the segment (Figure 35). In addition to the 3D rotatable skeleton plot, detailed information of the segment is added. First, a set of properties is calculated and four additional graphs are added to the figure: a cumulative distance plot for each joint, the translational kinetic energy distribution, the norm of the acceleration of the joint of interest and a distance-velocity phase plot. The skeleton is positioned at the maximum acceleration of the selected gesture. To interpret it more easily, a triangular marker is used to indicate the start position of the segment in the skeleton figure.
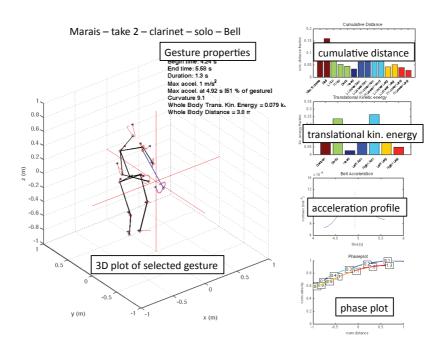
Figure 35. Example of 3D visualization of a selected segment.

9. Based on the annotated segments or gestures, a set of properties can then be calculated. Several properties were calculated, using obvious basic parameters (e.g. duration, displacement, average and variance of velocity, average and variance of acceleration,...) and more complex characteristics

(e.g. curvature, Frenet-Serret vectors, Mean Absolute Relative Phase, …). The main objective of the obtained set of properties was to use the results as the input for classification purposes, as explained in the next section.

10. Based on the set of properties for the annotated segments (gestures) an attempt was made to establish a classification. The method used, combined elements of Procrustes analysis and multi-dimensional scaling as explained in section 5.5.4 (subsection Procrustes analysis and Multi Dimensional Scaling).

The second part of the analysis was based on the top-down approach and the analysis path is shown in Figure 36.
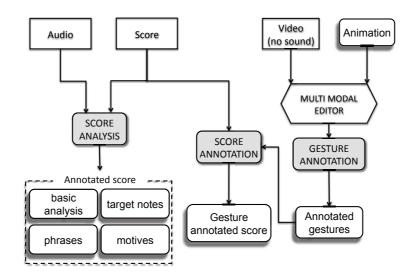
Figure 36. Path of the top-down analysis.

The top-down approach was performed by the player and consisted of the following steps:

1. First, an annotation by the performer was done and consisted of a basic interpretative analysis of the score. Second, target notes were added to the score based on the recall of the players' performance. These are the notes towards which phrasing is directed and very often constitute the

expressive culmination point of a phrase. Because these notes are linked to the formal structure of the piece, the performer also indicated phrases in the score. Furthermore, other elements such as principal notes (notes that support phrase) and recurring motives (upbeats, closing motives, passing notes) were annotated and added to the score, resulting in an annotated score (Figure 37).
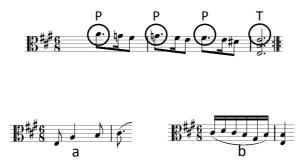
Figure 37. Examples of annotations on the score. T = Target note, P = Passing note. (a) = 'upbeat' and (b) 'closing' are examples of recurring motives.

2. A second annotation (gesture annotation) by the performer dealt with the video recording of the performance (Figure 38). To avoid being biased by the music, annotations were executed without sound in ELAN. Not all classifications were annotated due to the frontal perspective of the observation. Therefore, annotations were limited to the "clarinet bell" (circle), "head up-down", "arms flapping", "bending knees" and "feet stepping". The player also added "feet together", because it occurred very often, and "breathing" (mouth open and closed around the mouthpiece).
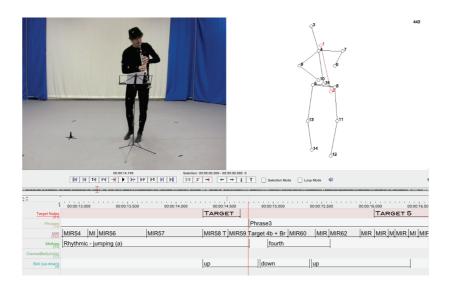
Figure 38. Example of player annotations in the multi-modal editor ELAN.

3. Gesture annotations based on the non-audio video analysis were added onto the score, resulting in a gesture annotated score. This enabled us to see, in a simple and comprehensive way, whether some gestures coincide with the elements (target notes, principal notes, recurring motives) that were annotated in the score based on the performer's descriptive analysis.

5.7 Establishing the framework

5.7.1   Concepts of subjective top-down analysis

In our analysis framework, we believe that an objective bottom-up approach to movement feature extraction (segments) needs to be further clarified with the help of a subjective top-down approach. This subjective top-down approach comprises both a qualitative as well as a quantitative methodology. In the qualitative approach, we aim at an intuitive understanding of certain specific relationships between the objective bottom-up and subjective top-down features of the musical gestures. In the quantitative approach, we aim at an explicit

measurement of particular relationships between the objective bottom-up and subjective top-down features of the musical gestures.

The above viewpoint is much inspired by the theory of embodied music cognition, as explained in Leman (2007). The theory states that our interaction with music is embodied, and thus mediated by the human body. It is through corporeal articulations that experiences can be encoded in musical sound (through music playing). It is through corporeal imitations that musical sounds make sense. The concept of a repertoire of gestures, or corporeal articulations, also called the action-oriented ontology, is an essential component of this theory. It is assumed that the notions of musical expression and intentionality can be linked with this gesture repertoire. Gestures are assumed to facilitate the interaction with music.

In order to integrate the concepts of the subjective top-down approach, the relationship between a musician's expressive body movements and locally intended targets and the supporting generative structure in the musical score were investigated. The latter can be conceived as the performer's focal points within the expressive discourse of music playing. We, thereby, assume that, based on these locally intended targets, the performer's gestural skills subsume a grouping of movements into larger movement patterns or gestures (Godøy and Leman, 2010). As such, the locally intended targets can be understood as reference frames for musical motility, which reduces the cognitive load during the musical performance. Without such targets and grouping, one would focus on all notes equally as much, and therefore it is likely that there would be more cognitive load, which would make a musical performance more difficult and certainly less musical (as it is with novice players). Therefore, a focus on targets and grouping in relation to musical information and gesture is a first step towards a better understanding of how possible reference frames for musical motility may facilitate music playing.

In music analysis, local targets are typically related to different parameters of the musical structure, such as long notes, high notes, notes related to the tonal structure, or rhythmic structure and so on. However, in order to identify the local targets of a particular performance, it is necessary to rely on the performer's introspective analysis. Moreover, one should keep in mind that local targets are present in any expressive performance style, no matter whether this style is 'light', 'heavy', or 'exaggerated expressive'. This approach thus requires the collaboration of a performer who assesses his or hers own performance in terms of local intentions, and chunked gestures. This approach supports the general idea that musical targets can be understood as intended goals that result in sound through the mediation of expressive gestures. This idea is inspired by theoretical accounts on embodied music cognition (Leman, 2007) and musical gesture (Godøy & Leman, 2010). It provides the basis to establish a framework that focuses on how body movements relate to local musical targets. This analysis of the intended musical targets (the subjective method) can be combined with motion capture and statistical analysis (the objective method).

5.7.2 Linking the top-down and bottom-up approach

Based on the results of the analysis of the *Clarinet player data* (Data 5), it was possible to realize the proposed framework of this thesis. In order to establish the link, it was important to fulfill the following conditions: (i) a communicative collaboration, (ii) a methodology for annotations, (iii) a tool that enables us to edit multi-modal data and (iv) a methodology for comparing the top-down and bottom-up approach.

Communicative collaboration was possible due to the fact that the performer is a collaborator at IPEM. It was not only possible that the player himself annotated his performance, but he was also able to discuss and change ideas about his viewpoints of the subjective top-down strategies from his personal background in philosophy, musical education (conservatory) and

musical teaching. This improved the realization of the framework to a great extend. The necessity to establish a multi-disciplinary collaborative way of working was found to be vital in order for the realization of to the framework to happen.

Annotations are based on (i) a descriptive analysis of musical structural interpretations and intentions, (ii) a frame-by-frame video analysis of gestures, using the categorization of clarinetists' expressive movements as defined by Wanderley et all (Wanderley et al., 2005) and (iii) a visualization of some annotated gestures on the score.

The solution for multi-modal data editing was found by using ELAN (http://www.lat-mpi.eu/tools/elan). Although the tool was developed in the field of language technology it proved to be very useful for the objectives in this research. Not only this software enables us to combine video, animations, audio for manual annotations but it also gives us the opportunity to import extracted features (segmentation data) and kinematic data (as time series) using the objective bottom-up method.

The methodology to compare the two approaches was based on a qualitative and a quantitative comparison. First, the qualitative method was realized by merging the ELAN files of the top-down and bottom-up methods. By selecting either an annotation made by the performer or a segment from the bottom-up feature extraction, using begin and end frame of the selection, a visualization is made in matlab and compared with the corresponding part in the score (see example in Figure 39).
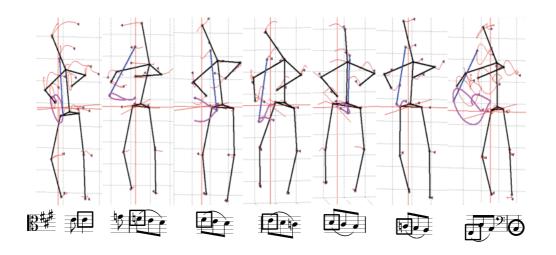
Figure 39. Example of selected sequence of segments compared with the corresponding part of the score. P = Passing note, T = Target note.

The same sequence can then be compared with annotations of the player in order to evaluate and compare similarities and differences between the top-down and bottom-up approach (see Figure 40).
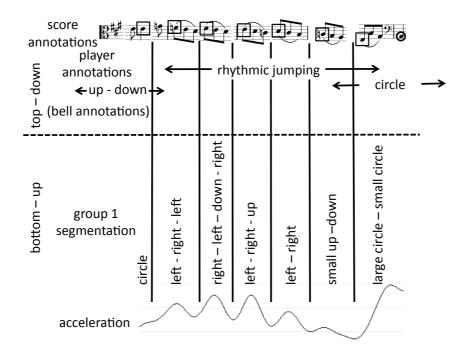
Figure 40. Example of qualitative comparison Top-Down and Bottom-Up annotations

Second, a method that enables us to develop a quantitative comparison was established. A method was developed to quantify the degree of correspondence between the player annotations and the extracted gestures from the movement data. From the start and end points found in the annotations, a Gaussian distribution is constructed where the mean of the distribution is the mean of the start and end points and the width is defined by the 95% confidence interval defined by the length of the gesture. The same procedure is applied to the gesture segments found in the movement data. The resulting sets of Gaussian distributions are tested for similarity by performing a t-test. Based on this test, a gesture segment found in the movement data is accepted or rejected in the comparison with those found through annotation (Figure 41).

Figure 41. Quantitative comparison of segmentation and player annotations.

5.8 Summary

In this chapter, the methodology used in this work has been presented. An overview of the evolution of hardware and software has been presented. The importance of exploratory data analysis in and the key concept of variability were discussed and selected examples of methods used in this work were presented. By means of statistical flow charts, the methods needed to investigate the research questions corresponding the data sets the methodology was further

deployed. Finally, the establishment of the framework was given. The next four chapters contain texts of the resulting four papers.

6 Sharing musical expression through embodied listening - a case study using Chinese guqin music.[1]

Abstract

In this study we report on the result of an experiment in which a guqin music performance was recorded and individual listeners were asked to move their arm along with the music that they heard. Movement velocity patterns were extracted from both the musician and the listeners. The analysis reveals that the listeners' movement velocity patterns tend to correlate with each other, and with the movement velocity patterns of the player's shoulders. The findings support the hypothesis that listeners and player share, to a certain degree, a sensitivity for musical expression and its associated corporeal intentionality.

Keywords: embodied music cognition, music, perception, expression, guqin music

6.1 Introduction

Musical signification practices involve high degree of corporeality and, therefore, it is interesting to study music perception from the viewpoint of body movement. In a recent survey by Lesaffre et al. (2008), 95% of participants ($n = 663$) reported that they move spontaneously along with music while listening. Apparently, when people enjoy music, there is often body movement that goes along with it.

In the past, music perception often has been linked with movement. For example, Aristotle assumed that music is reflecting "men in action," and that music, through melody and rhythm, expresses different characters with which our soul may attune during the perception of music (Aristotle & McKeon, 2001). Hume suggested that a kind of sympathy is essential in our perception of art (Hume, Selby-Bigge, & Nidditch, 1975), while Kant (1790/1924) introduced the notion of *sensus communis* as a kind of inter-subjective sensitivity for beauty in art. In the late 19th century, Hanslick (1891) introduced the idea that the perception of moving forms ("tönend bewegte Forme") is sufficient for signification. However, it was Lipps (1903) who, in a more explicit way, spoke about the corporeal articulation of expression ("Ausdrucksbewegungen"). Through moving along with the forms in art (with our body and/or our mental imagery), he claimed it is possible to enter into an empathic relationship with (musical) forms and, consequently, to fully signify them. Later on, steps were undertaken to identify different types of body movement in relation to music perception, such as expressive movements (Truslit, 1938; see also Repp, 1993), repetitive movements related to conducting (Becking, 1928; see also Nettheim, 1996), or movements that express emotional content (Clynes, 1977).

In recent work, the relationship between music perception and movement has been studied from different viewpoints, such as musicology and philosophy (e.g., Cumming, 2000; Hatten, 2003), behavioral studies about music

performance, walking on music, activation of emotions (e.g., Friberg & Sundberg, 1999; Friberg, Sundberg, & Fryden, 2000; Gabrielsson, 2003; Juslin & Laukka, 2004; Palmer 2005; Styns, Van Noorden, Moelants, & Leman, 2007; Sundberg, 2003), cognitive neuroscience (e.g., Brown, Martinez, & Parsons, 2006; Chen, Zatorre, & Penhune, 2006; D'Ausilio, Altenmüller, Belardinelli, & Lotze, 2006; Zatorre, Chen, & Penhune, 2007), and computational modeling (e.g., De Poli, 2004; Camurri,Volpe, De Poli, & Leman, 2005, Mion & De Poli, 2008; Widmer & Goebl, 2004). Movement is now seen as an important factor that contributes to perceptual disambiguation (Phillips-Silver & Trainor, 2008). Movement simulation also accounts for the better recognition of and better synchronization with one's own audio-recorded piano performance (Keller, Knoblich, & Repp, 2007; Repp & Knoblich, 2004). The link between music perception and movement has furthermore been supported by evidence for shared neuronal codes of perception and action (e.g., Berthoz, 2000; Hommel, Musseler, Aschersleben, & Prinz, 2001; Jeannerod, 1994;Wilson & Knoblich, 2005), with important implications for our understanding of social musical interaction (Clayton, 2007; De Bruyn, Leman, Moelants, & Demey, 2009; De Bruyn, Leman, & Moelants, 2008; Keller, 2008).

In short, in our understanding of music perception, movement often has been at the focus of attention, through notions that imply the sharing of musical expression and the coupling of perception and action. Thanks to recent developments in technology a direct study of expressive body movement in relation to music perception now is possible (Leman & Camurri, 2005), although measurement and direct evidence still is hindered by a number of methodological difficulties, such as the identification of reliable movements, and their distinction from unreliable or arbitrary movements (Wanderley, Vines, Middleton, McKay, & Hatch, 2005). In addition, there are particularities of data analysis that deserve special attention, because movement data are time series and the data-analysis of

time series is known to be difficult especially when human movement is involved (Desmet, Leman, Lesaffre, & De Bruyn, 2010).

The aim of the present paper is to explore whether music expression can be studied and understood within the paradigm of embodied music cognition (Leman, 2007), using empirical methods that comply with the particularities of music-driven human movement. In the embodied music cognition approach it is assumed that the human body is the natural mediator between experience and physical reality. Music-driven movements of the listener are seen as corporeal articulations that reveal a mirroring (in movement) of the perceived sonic moving forms. Through this mirroring, it is assumed that musical expression can be captured more fully, experienced, and understood.

Correlations of music-driven movements among listeners would support the hypothesis that their embodied perception of music expression is shared. Correlations between listeners and the player would further support the hypothesis that there is a link between what listeners perceive as expression, and what the player produces as expression. A correlation would support the idea that listeners can somehow decode (and reveal through music-driven corporeal articulations) the expression that the player has encoded in the musical audio.

A major research question is how the above hypothesis of embodied music cognition can be tested and further explored. In the present paper, we use fragments of Chinese guqin music as stimuli in an experiment that measures and compares music-driven corporeal articulations.

The guqin (pronounced *ku-chin* in English) is interesting for our purpose. The instrument belongs to the family of the zither, a plucked stringed instrument. It consists of a long and narrow hollow wooden box that functions as a sound box and whose upper part functions as a fretless fingerboard on top of which there are seven strings attached, each about 110 cm long. The strings are tuned to a pentatonic scale. The way in which guqin music is played, namely by plucking

the string with the right hand and moving the finger of the left hand over the string, makes it suitable for a detailed study of the relationship between movements of listeners and movements of the player. Indeed, the guqin has no bow, and fingering is directly related to sound, as there are no frets to interfere within the sliding. Although the playing technique is rather complex, the sliding tones in guqin music can be conceived as sound patterns that reflect aspects of the playing movement, without much intermediate technology. Guqin music is thus of particular interest to embodied music cognition research because the encoding of playing gestures into sound patterns proceeds in a relatively direct way, which facilitates the study of possible relationships between movement and sound (see Henbing & Leman, 2007).

The methodology that we developed for this study has a focus on the registration and extraction of the movement velocity. The data analysis is based on a technique that combines the correlation of time series with cost functions that account for local time shifts of the recorded time series (Desmet et al., submitted).

6.2 Method

In this study, we made an audio, video, and movement recording of three fragments of guqin music. Then we asked listeners to listen to these fragments and respond by spontaneously moving a telescopic stick along with the music. The movements of the listeners were recorded.

6.2.1 Stimuli

The musical stimuli used in this experiment were three musical fragments (P1, P2, P3) of the traditional Chinese piece "Missing an old friend" played on the guqin by Henbing Li, an experienced guqin player educated at the Central Conservatory of Music in Beijing. Each fragment had a duration of about 30 s. Of the three musical fragments, P1 and P2 have a rather fluent melodic line that is

clearly structured. In contrast, P3 has a more narrative character with a less fluent melodic line. The fragments contain pitch slidings (in glissando and vibrato patterns), which is typical for guqin music. None of the fragments has a clear regular beat. The audio and video of this performance was recorded on hard disk[2]. In addition, a motion capturing system with infrared cameras was used to monitor the movements of 11 different markers attached to the joints of the musician (Figure 42). The movements were recorded at a sampling frequency of 100 samples/s using the Qualisys Motion Capture System (http://www.qualisys.com).

Figure 42. Guqin player and motion capturing system

6.2.2 Task

During the experiment, listeners were asked to move a joystick while listening to the music through headphones. Each fragment was preceded by five short beeps in order to announce the next fragment. The joystick was attached horizontally to a stand. A telescopic stick, which allowed free movement of the arm in a considerable space in front of the participant, was attached to the joystick and participants were asked to move this stick with the right or left arm (see Figure 43). The adapted joystick (Saitek ST 90) was connected to a computer via USB,

---

[2] See www.ipem.ugent.be/2005/Guqin/video/P1P2P3.wmv

using the musical software AnalogBox (http://www.andyware.com). The movements were recorded in a two-dimensional plane, using horizontal and vertical coordinates. The movement data were stored on hard disk at a sampling frequency of 100 samples/s. After each session, listeners were asked to fill in a questionnaire where they gave an assessment of their performance.
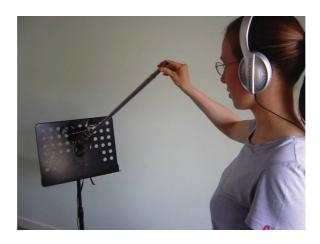
Figure 43. Recording of embodied music listening via joystick

6.2.3 Participants

Thirty participants (13 men, 17 women) participated in the experiment. They were selected from a database of participants that was set up in the context of a musical audio mining project (MAMI) at Ghent University (Lesaffre et al., 2008). None of the participants knew Chinese guqin music. All the participants explicitly agreed that their data might be used for scientific research. The age of the participants varied from 21 to 35 years with an average of 27.6 years. Fifteen participants played an instrument. On average, the participants listened to music 19.6 hours ($SD = 13.1$) per week. Participants were paid 10 Euro for their participation to the experiment.

6.2.4 Procedure

Participants were asked to fill in a questionnaire about gender, age, musical background, listening hours per week, absolute pitch perception. In order to become familiar with the equipment before the start of the experiment, participants were asked to move the joystick while listening to a musical fragment by Chopin (Preludes in C major—24 Preludes Opus 28). The fragment was chosen to acquaint participants with the joystick and the task of moving along with music. Participants could practice as many times as they wanted. The experiment itself then consisted of four sessions that followed each other. In between sessions, participants received information about guqin music. After the first session, participants could listen to the music as much as they wanted. After the second session, participants could inspect a graphical score, and after the third session, participants could watch the video of the player. After each session, participants reported on their own performance (the processing of these data is not contained in the present paper). The four sessions enter into the analysis in order to check consistency and possible learning effects.

Each session contained the same three pieces (P1, P2, P3) of guqin music, played in the same order. The participant had to listen and move to the same piece twice in a subsequent repeated trial as in: P1-P1—P2-P2— P3-P3. The repeated piece was always separated by a short break. Between pieces, a slightly longer break was used. Each participant took part in 12 trials (three pieces in four subsequent sessions).

6.3 Data analysis

The movement velocity pattern is a time series whose values give the amount of displacement over subsequent time units. Given a lack of frequencies above 2 Hz, the sampling rate of the movement velocity patterns could be reduced to 4 samples/s, which corresponds to subsequent time units of 250 ms. The

displacement is based on three dimensions in case of the infrared camera system, and on two dimensions in case of the joystick. However, in the movement velocity pattern, three or two dimensions are reduced to one single dimension, so that a comparison between movements recorded with both systems becomes possible.

In calculating the correlation between any pair of movement velocity patterns, both the raw and the warped patterns are studied. Raw patterns are just the original movement velocity patterns, while warped patterns are based on dynamic time warping, which is a technique that allows local time shifts of values in time series. The warped time series compensates for the fact that movements of the listener can be locally time shifted as compared to the player's movements, due to anticipation or delay. In this study, the dynamic time warping envelope was restricted to a maximum of 750 ms (which corresponds with three movement velocity samples). Figure 44 gives an example of the dynamic time warping technique. Correlation of warped movement velocity patterns can produce better results than correlation of raw movement velocity patterns because there can be a better match between the peaks of the patterns. However, this improvement is not guaranteed. The number and nature of the local shifts can be calculated as a cost of the dynamic time warping, which can be taken into account in the analysis. For example, in the section on listener-player correlations, this cost is taken into account in order to determine the body part of the player that best correlations with the arm movements of the listener.
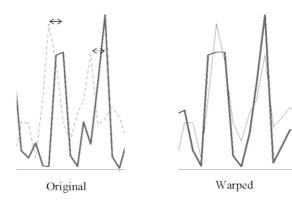
Figure 44. The effect of dynamic time warping. Left: comparison between unwarped (solid line) and reference (dotted line), right: comparison between warped (solid line) and reference (dotted line).

Data analysis is based on intra-listener correlations, inter-listener correlations, and listener-player correlations: (i) Intra-listener correlations address each participant's ability to repeat movement velocities during a subsequent repeated listening of a piece (e.g., as in P1- P1 or P2-P2). The number of significant correlations ($p < .01$) reflects the extent to which the population can perform this task in a reliable (that is, repeated) way. For that reason, these performances are also called *reliable* performances. The choice for $p < .01$ as a criterion for making the distinction between reliable performances and unreliable performances was based on a detailed inspection of the number of correlated movement velocities. (ii) Inter-listener correlations address the relationships among the movement velocity patterns of listeners. (iii) Listener-player correlations address the correlations between movement velocity patterns of listeners (one variable) and the movement velocity patterns of the different joints of the player (11 variables).

6.4 Results

Repeated measures ANOVA revealed that there were no effects of gender, age, musical background, listening hours per week, and absolute pitch perception.

This is in agreement with the fact that all participants were unfamiliar with the music. Therefore, these effects were discarded from further analysis.

6.4.1 Intra-listeners Correlations

6.4.1.1 Analysis of raw movement patterns.

Figure 45 shows the relative amount of reliable performances as a mean value for all participants, using the unwarped patterns. The amount increases for the three fragments P1, P2, and P3 over the sessions S1, S2, S3, S4. This increase works best for fragment P1 and fragment P2, where it goes from above 40% in session S1 to about 70% in session S4. Instead, for fragment P3, there is only a slight increase, as the number goes from 40% to about 45%. The optimal fit was calculated by selecting the maximum R^2 value for different regression types. The best fit was obtained using a second order regression. GLM (General Linear Modeling) for Repeated Measures (Tables 6a and 6b) revealed a significant difference between pieces ($p < .05$), a significant effect of session for a linear model ($p = .005$) and no significant interaction between session and piece. This means that, for reliable performances, participants tend to improve in their performance for all pieces over sessions.
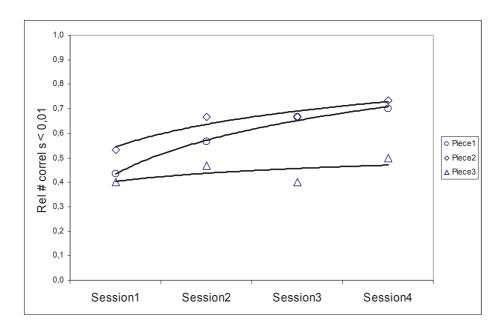
Figure 45. The relative amount of reliable performances as a mean value for all subjects, using the warped patterns.

(A) Test for Differences Among Pieces

Source	Type IV Sum of Squares	df	Mean Square	F	p	Partial Eta Squared
Intercept	113.34	1	113.34	278.69	.000	.76
Piece	2.77	2	1.39	3.41	.04	.07
Error	35.38	87	0.41			

(B) Test for Differences Among Sessions

Source	Session	Type IV Sum of Squares	df	Mean Square	F	p	Partial Eta Squared
Session	Linear	1.50	1	1.50	8.14	.005	.09
	Quadratic	0.04	1	0.04	0.25	.62	.003
	Qubic	0.11	1	0.11	0.56	.46	.006
Session* Piece	Linear	0.33	2	0.17	0.91	.41	.020
	Quadratic	0.07	2	0.04	0.20	.82	.005
	Qubic	0.09	2	0.04	0.23	.80	.005
Error (session)	Linear	16.06	87	0.19			
	Quadratic	15.38	87	0.18			
	Qubic	16.90	87	0.19			

Table 6. GLM for Repeated Measures applied to the data of Figure 4: (a) test for differences among pieces, (b) test for differences among sessions.

6.4.1.2 Analysis of warped movement patterns.

The results obtained with the dynamic time warping technique applied are shown in Figure 46, and Table 7a and 7b (for the GLM analysis). The number of reliable performances over session is from 65% to 85% for P1, and from 60% to 70% for P2. With warped data, P3 gets 35% to above 40%. The results with warped data are not very different from the results with unwarped data, suggesting that participants do not anticipate or delay their own performances. In both cases, the major result is the number of reliable patterns, which indicates that listeners can repeat their own movements when listening to the same piece that is immediately repeated. However, this number depends on the character of the piece, as the differences between P1 and P2 with respect to P3 clearly illustrate.
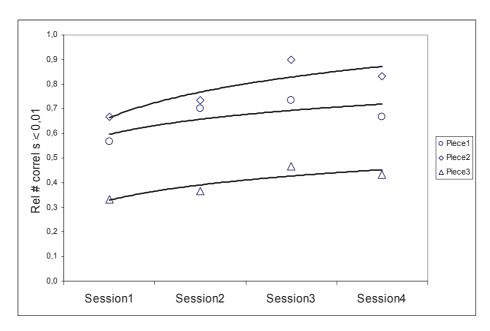
Figure 46. Relative amount of reliable performances as a mean value for all subjects, using the warped patterns.

(A) Test for Differences Among Pieces

Source	Type IV Sum of Squares	df	Mean Square	F	p	Partial Eta Squared
Intercept	138.14	1	138.14	441.16	.000	.84
Piece	9.37	2	4.69	14.97	.000	.26
Error	27.24	87	0.31			

(B) Test for Differences Among Sessions

Source	Session	Type IV Sum of Squares	df	Mean Square	F	p	Partial Eta Squared
Session	Linear	0.93	1	0.93	15.38	.002	.06
	Quadratic	0.47	1	0.47	12.69	.01	.03
	Qubic	0.09	1	0.09	1.51	.05	.006
Session* Piece	Linear	0.11	2	0.05	0.31	.73	.007
	Quadratic	0.11	2	0.05	0.30	.74	.007
	Qubic	0.15	2	0.07	0.40	.67	.009
Error (session)	Linear	15.11	87	0.17			
	Quadratic	15.18	87	0.17			
	Qubic	16.11	87	0.19			

Table 7. GLM for Repeated Measures applied to the data of Figure 5: (a) test for differences among pieces, (b) test for differences among sessions.

6.4.2   Inter-listener Correlations

The number of inter-listener correlations based on the raw data of all reliable ($p <$ .01) performances were low (Figure 47), but with warped data inter-listener correlations were obtained in about 50% of the movements for fragments P1 and P2, and about 30% for fragment P3 (Figure 48). This seems to indicate that movement velocities of different participants display some anticipation and delay. GLM for Repeated Measures (Table 8a, related to Figure 48) revealed a significant difference between pieces ($p <$ .01). In addition, the analysis showed a significant effect over session ($p =$ .002) and a somewhat weak interaction between session and piece ($p <$ .01) for a cubic model. The latter may be attributed to the slightly deviant trend of fragment P3, compared to fragments P1 and P2. The rubato narrative character of fragment P3 seemed to have a less pronounced degree of inter-listener agreement, but a higher degree of personal interpretation as sessions proceed.
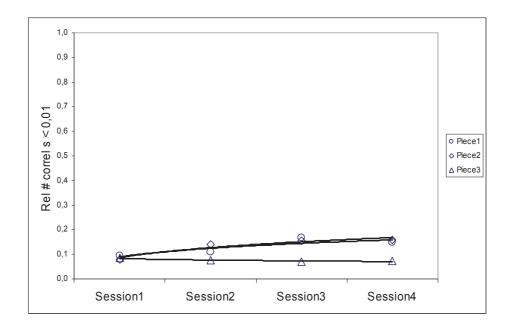
Figure 47. The number of inter-subject correlations, based on all reliable ($p < 0.01$) performances, unwarped.
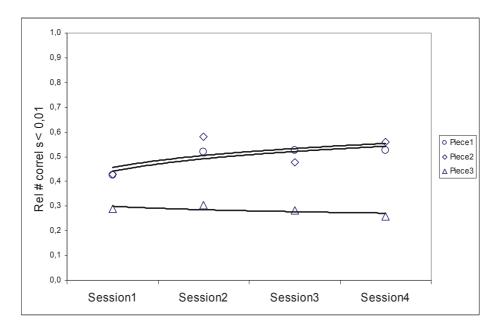
Figure 48. The number of inter-subject correlations, based on all reliable ($p < 0.01$) performances, warped.

(A) Test for Differences Among Pieces

Source	Type IV Sum of Squares	df	Mean Square	F	p	Partial Eta Squared
Intercept	225300.10	1	225300.10	172.73	.000	.67
Piece	13270.02	2	6635.01	5.09	.008	.11
Error	113476.38	87	1304.33			

(B) Test for Differences Among Sessions

Source	Session	Type IV Sum of Squares	df	Mean Square	F	p	Partial Eta Squared
Session	Linear	386.42	1	386.42	3.88	.05	.04
	Quadratic	321.11	1	321.11	3.32	.07	.04
	Qubic	520.57	1	520.57	10.61	.002	.11
Session* Piece	Linear	564.67	2	0.17	2.84	.06	.06
	Quadratic	40.11	2	0.04	0.21	.81	.005
	Qubic	492.54	2	0.04	5.02	.009	.10
Error (session)	Linear	8659.21	87	0.19			
	Quadratic	8420.28	87	0.18			
	Qubic	4268.59	87	0.19			

Table 8. GLM for Repeated Measures applied to the data of Figure 7: (a) test for differences among pieces, (b) test for differences among sessions.

6.4.3 Listener-Player Correlations

In studying the correlation between listener and player we used the movement data from each listener and compared these with the movement data of the different joints of the player. The joints of the player were labeled as follows: J1 = head, J2 = left shoulder, J3 = left elbow, J4 = left wrist, J5 = left thumb, J6 = left ring finger, J7 = right shoulder, J8 = right elbow, J9 = right wrist, J10 = right middle finger, J11 = right index finger (see Figure 49). In this comparison, we used the dynamic time warping technique by considering the total number of local shifts as a cost. This cost was then taken into account with the correlation analysis.
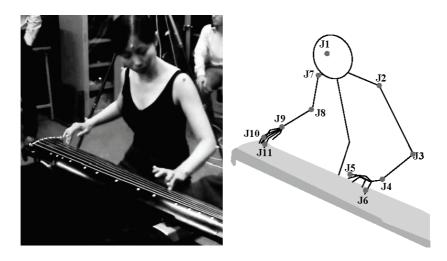
Figure 49. Picture of the player and related stick figure with joints labeled (see text).
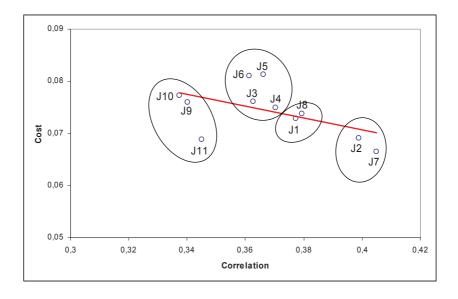
Figure 50. Correlation and cost between different joints of the player and the response of the listeners.

Figure 50 shows that there is an interesting relation between correlation and cost. The higher the correlation obtained (horizontal axis), the lower the cost it takes to achieve this (vertical axis). This analysis therefore clearly shows that the movement velocity of the right shoulder of the player (J7) has the highest average correlation and the lowest cost compared with the arm movement of the listeners. Based on the movement velocity patterns of 11 different joints, the player's

movements can be classified, using cladogram analysis, into four different groups, namely right hand, left elbow and left hand, head and right elbow, and shoulders (Figure 50). The cladogram analysis (using Phylip, http://evolution.genetics.washington.edu/phylip.html) aimed at organizing a given set of movement data according to character state changes in the corresponding encoded sequences. Based on the above results, we selected the right shoulder (J7) as the reference movement that will be used for further comparison with movements of the listeners.

The average number of correlations between unreliable performances and the player's right shoulder (J7) are below 40% for all fragments (Figure 51). GLM for Repeated Measures for unreliable movements revealed no significant effect over session. Figure 52 shows the results for reliable performances. The average number of significant listener-player correlations is between 40% and 55% of the total amount of reliable performances. For fragment P1 and P2, there was a trend towards an increasing amount of correlations over sessions, whereas for fragment P3, the trend decreased. GLM for Repeated Measures revealed a significant effect over piece ($p < .001$) (Table 9a) and a significant effect over session ($p = .005$) (Table 9b). This result corresponds to previous findings that the movements depend on the nature of the musical pieces.
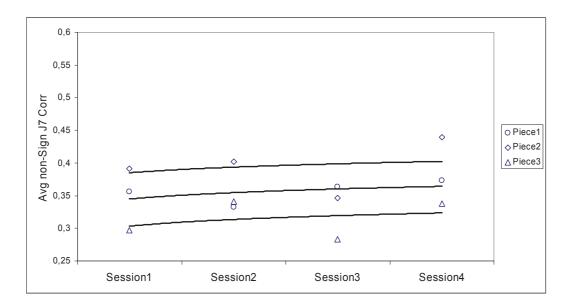
Figure 51. Average correlations between listeners' movements and J7 *non-reliable* performances over the four sessions.
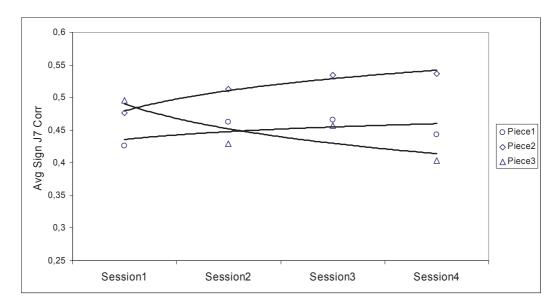
Figure 52. Average of correlations between listeners' movements and J7 for *reliable* performances over the four sessions.

The number of correlations, still up to about 40% for P1 in the case of unreliable performances (Figure 51), may be explained by the fact that these performances may still contain movements that are in tune with those of the player. Indeed, unreliability is merely related to the fact that the listener did not repeat the

movement in the subsequent trial. However, it is not excluded that one out of the two movement velocities of the repetitive trial has indeed some correspondence with the movement velocities of the player. The distinction between reliable and unreliable performances is indeed rather robust. However, in general, the percentages of the unreliable performances are significantly lower than the ones of the reliable performances. These results contribute to the hypothesis that the movement velocities of the listeners' arm tend to correlate with the movement velocities of the player's shoulder.

(A) Test for Differences Among Pieces

Source	Type IV Sum of Squares	df	Mean Square	F	p	Partial Eta Squared
Intercept	67294.68	1	67294.68	386.32	.000	.82
Piece	3522.34	2	1761.17	10.11	.000	.19
Error	15154.98	87	174.20			

(B) Test for Differences Among Sessions

Source	Type IV Sum of Squares	df	Mean Square	F	p	Partial Eta Squared
Session	1.50	1	1.50	8.14	.005	.09
Session* Piece	0.33	2	0.17	0.91	.41	.02
Error (session)	16.06	87	0.19			

Table 9. GLM for Repeated Measures applied to the data of Figure 11: (a) test for differences among pieces, (b) test for differences among sessions.

6.5 Discussion

6.5.1 Methodology

From a methodological point of view, we believe that the repeated listening design (using musical excerpts of 30 s) provides an interesting method for determining whether a music-driven movement is reliable or not. It was assumed that listeners who found a proper motor strategy to express their movement *in tune* with music (during these 30 seconds) would not radically change their strategy during a second performance that immediately followed the first one, while listeners who found that their motor articulations were *out of tune* with

music would be inclined to change strategy and therefore give a different performance during the second trial. Thus, high correlation between successive performances (e.g., P1—P1) would indicate *reliable performance*, while low correlation would indicate *unreliable performance.* The measurement of the correlation between the two repetitive performances thus offers an objective criterion for accessing movement reliability. The analysis of the movement data further reveals that better results can be obtained with warped movement patterns than with unwarped movement patterns. Improved results with warped data account for the fact that participants may show different adaptations (reacting faster or slower) to the presented musical excerpts.

Data about the listeners' subjective evaluation of their own performance (based on a questionnaire after each session, but not reported in detail in the present paper) are consistent with the objective observation of reliable performances. The participants were asked to evaluate the degree of difficulty, discrepancy, and satisfaction of their performances over the pieces and the sessions. Using cross tabulation (chi-square) it was found that the evaluation changed primarily over the pieces and slightly over the sessions. Although further details go beyond the scope of the present paper, it is of interest to mention that, for reliable listener performances, a significant correlation was found between an objective measure (based on the repeated listening design explained above) and a subjective measure (based on the self-assessment of the listener's own performance). No such relationship was found for the unreliable performances.

6.5.2 Inter-listeners and Listeners-Player Correlations

The results for inter-listener correlations support the hypothesis that listeners share the movement velocity patterns of music-driven corporeal expressions, and thus the embodied perception of expression during listening to music. In addition, the results for listener-player correlations support the hypothesis that there may be a link between what listeners perceive as expression, and what the player

produces as expression. This link is established through corporeal articulations that accompany music playing and music listening. The data seem to support the idea that listeners can somehow decode the expression that the player has encoded in the musical audio. The results also contribute to the hypothesis that the embodiment of music expression has grounding in social communication.

Obviously, the two hypotheses, namely inter-listener correlation, and listener-player correlation, are not completely independent. If the movement velocity patterns of the listeners correspond with the movement velocity patterns of the player ('s shoulder), then it can be expected that the movement velocity patterns of the listeners are also related with each other. Our analysis seems to confirm this relationship between the two hypotheses.

6.5.3   The role of Music

The number of inter-listener and listener-player correlations could depend on the nature of the musical fragment. The fragments with a melodic pitch line (P1 and P2) scored better than the fragment with a broken melodic, and thus more declamatory, pitch line (P3). It was noticed that fragment P3 displays an interesting difference between intra and inter-subject analysis. The intra-subject analysis suggests that listeners tend to improve their embodied perception for P3 over session, while the inter-subject analysis suggests that listeners tend to drive away from a common interpretation. The two trends can be explained by the fact that listeners develop their own embodied listening solution to P3 in a consistent way, but their embodiment solution differs from each other. In this particular example, the lack of a shared inter-subjective embodied listening solution may be due to the narrative character of the piece. The results suggest that melodic fragments may induce a higher degree of inter-subjective share among listeners than narrative fragments, as the latter have more appeal to individual interpretation. In other words, listeners are more consistent among each other in the interpretation of melodic fragments than in the interpretation of narrative

fragments, and this increases with session. This indicates that the listeners' shared music-driven corporeal articulations depend on the character of the music, that is, on musical parameters such as melodic line and prosody that have the power to activate movement.

Similar results have been found in De Bruyn et al. (2008). Although in this study, the group condition outperformed the individual condition, there was a major effect of musical piece, in the sense that some pieces, although similar in tempo, activated body movement better than other pieces. We believe that the differences between P1, P2, and P3 can be attributed to a similar effect of musical piece. The melodic lines of P1 and P2 are easier to express, and hence, their expression is easier to share. However, the precise relationship between acoustical features and musical structure on the one hand, and the movement features of its co-articulated corporeal expressions on the other hand needs further investigation.

6.5.4   The Role of Learning

The present study did not focus on learning, although the use of four sessions, mainly intended to test the consistency among the listeners' movements, but revealed some interesting aspects. The significant improvement of the results over the four sessions for P1 and P2 seems to indicate that aspects of learning may be involved. The more the participants got acquainted with the guqin music, the more they shared each other's velocity of co-articulated corporeal expressions. This is in line with the idea that embodied listening responses may involve implicit learning, and hence, that the shared expression among participants may be enhanced in response to repeated exposure to music. However, learning may depend strongly on the character of the music, since P3 shows no such learning effect.

6.5.5 The Role of Shoulders

Analysis of the relationship between the movement velocity patterns of the joints of the player (Desmet, Leman, Demey, & Lesaffre, submitted) reveals that the head is often in advance of the actual playing gesture. Studies on the kinematics of human movement of arms and hands show that the elbow and wrist are loosely coupled, whereas the shoulder and elbow are tightly coupled (Bosga, Meulenbroek, & Swinnen, 2003; d'Avella, Fernandez, Portone, & Lacquaniti, 2008; Lacquaniti and Soechting, 1982). Large-scale movements are carried out primarily using the proximal joints of the upper limb (shoulder and elbow), whereas small-scale movements are carried out by the most distal joints (wrist and fingers). Playing the guqin can be seen as the combination of both large-scale (shoulder—elbow) and small-scale (wrist—finger) movements. While playing, the head seems to indicate movement intention, and the movement of the shoulder can be considered as a large-scale movement that is related to the small-scale control of sound production gestures. The left arm is involved in large-scale transversal (or horizontal from the viewpoint of the player) movements from one position to another position on the guqin strings, and in small-scale (also transversal) movements that focus on pitch sliding. The large-scale movements are technical and not directly related to the sonic image, whereas the small-scale movements leave a direct sonic trace in terms of pitch modulation (such as glissandi and vibrato, as shown in Henbing & Leman, 2007). The right arm is mainly moving in the sagittal plane, which is perpendicular to the string (or vertical from the viewpoint of the player), for plucking purposes.

An analysis of the player's movements (Desmet et al., submitted) shows that the movement of the right shoulder (J7) and the right elbow (J8) correlate better (mean correlation coefficient for the three pieces = .61) than the movement of the left shoulder (J2) and the left elbow (J3) (mean correlation coefficient for the three pieces = .25). This is consistent with the finding that the listeners'

movements correlate more with the right elbow than with the left elbow (see Figure 50), which seems to imply that the listeners' movements share patterns with the player's large-scale movements of expressive plucking. The fact that the arm and hand movements of the listeners are constrained by holding and moving a stick may constrain small-scale movements. However, movements of the wrist are possible, similar to conducting movements.

We further believe that the movement of the player's shoulder has an intentional component, though its precise goal may be vague. This hypothesis is based on the fact that plucking (by the right hand) provides energy to the string, which is then used for pitch sliding (by the left hand). However, during plucking, the player's attention is often focused on the pitch sliding because that action requires very precise and detailed movements. Hence, the shoulder articulation that goes together with plucking may well be an expression of the movement pattern that has to be executed in a later stage after plucking. In that sense, the movement of the shoulder has an intentional character, rather than an executive character. Since listeners tend to move in correlation with the player's shoulder, we believe that it is justified to say that the shared expressive patterns among listeners and player contain an aspect of intentionality, although the precise goal of the movement may remain vague. Kant (1790/1924) already drew attention to the notion of 'sensus communis' as the capacity for sensing goal-directness without goal. He conceived it as the core capacity for aesthetic appreciation, which human participants share among each other.

The above findings seem to support the hypothesis (i) that listeners can decode to a certain degree the expression of the musical stimulus through corporeal articulations, and (ii), that, in some cases, these corporeal articulations may reflect the expressive articulations of the performer.

6.6 General Discussion

The ability to decode the expressive code from music through embodiment can be linked with a social component and a language component.

Indeed, one of the major effects known in social anthropology is that music contributes to the development of a personal self of young people, and that it fosters social bonding and group identity (Freeman, 2000; Gregory, 1997; Hargreaves & North, 1997). Although this social anthropological aspect has not been addressed explicitly in the present study, our methodology of monitoring embodiment may be highly relevant in view of studies that address social music interaction. Expressive social behavior often relies on non-conscious imitation or mimicry, which shows itself in movement synchrony, behavior matching, as effectuated in speech patterns, facial expressions, emotions, moods, postures, gestures, mannerisms, and idiosyncratic movements (Bargh & Chartrand, 1999; Lakin, Jefferis, Cheng, & Chartrand, 2003; Niedenthal, Barsalou, Winkielman, Krauth-Gruber, & Ric, 2005). Within a social group setting, a participant is more likely to get along harmoniously with others in the group if it is behaving similarly to them, compared with being "out of sync" and behaving differently. Also in the visual domain, observers are most sensitive to their own movements, compared to those of friends, and then those of strangers (Loula, Prasad, Harber, & Shiffrar, 2005). In that context, music may be considered a particular case in which expressive behaviors are driven by a sonic source. Our study suggests that these behaviors may be partly shared, and that embodiment forms the core of this *sensus communis* (Kant, 1790/1924) for music expression. Our study suggests that music, through embodiment, can be a strong driver of social interaction. This finding is in line with recent studies on empathy and social interaction, in which the 'mirror system' plays a central role (see e.g., Gallese, 2003; Knoblich & Sebanz, 2006).

As to language, McNeill (2005) and Gallagher (2005) have argued that hand gestures that accompany speech can be seen as a motor component of speech, necessary to create the narrative space that is shared in a communicative situation (see also Brown et al., 2006; Sparing et al., 2007). In this approach, expressive (co)-articulations of the hand are not just seen as a supplement of language, but as a component that is fully integrated with language. We believe that the result of the present study is in agreement with the hypothesis that spoken language and music have a common motor component that shows itself in expressive gesture. In both music and language, it is likely that embodiment creates a shared space for the communication of expression.

Our study suggests that the communication of an expressive code through music forms part of a mechanism where listeners, in order to bridge the semantic gap between sonic cues and meaning, embody music and thus engender a plausible or virtual action to music based on their own action repertoire (Leman, 2007). In line with the embodied music cognition paradigm, this embodiment is based on a mirror system that is rooted in a shared neuronal basis for perception and action. Through embodied listening, music perception can be said to be grounded in the listener's subjective action repertoire and its perceived expression can be shared and signified by a community.

6.7 Conclusion

The present study developed a methodology to study the hypothesis that expressive patterns in music can be generated, perceived, and shared among player and listeners. The methodology is based on the measurement of movements, using a design in which listeners have to repeat their movement patterns in order to be sure that their co-articulations were regular or reliable.

The epistemological background for this study is rooted in the paradigm of embodied music cognition, which assumes a close relationship between perception and action through a mirroring mechanism. In this paradigm, the

human body is seen as a natural mediator between subjective experience and physical reality. Corporeal co-articulations of music playing and music listening can be seen as expressions of this natural mediation.

The comparison of the movement velocity patterns suggest that listeners share their sensitivity for musical expression and that they are able to display this sensitivity through music-driven corporeal articulations. The study also revealed that the movement velocity patterns of the player's shoulder correlates with the movement velocity patterns of the listeners' arms, which suggests that their corporeal co-articulations are shared to a certain degree. Further analysis of the shoulder patterns let us believe that the sharing of musical expression has an intentional character

7 Statistical analysis of human body movement and group interactions in response to music.[3]

Abstract

Quantification of time series that relate to physiological data is challenging for empirical music research. Up to now, most studies have focused on time-dependent responses of individual subjects in controlled environments. How ever, little is known about time-dependent responses of between-subject interactions in an ecological context. This paper provides new findings on the statistical analysis of group synchronicity in response to musical stimuli. Different statistical techniques were applied to time-dependent data obtained from an experiment on embodied listening in individual and group settings. Analysis of inter group synchronicity are described. Dynamic Time Warping (DTW) and Cross Correlation Function (CCF) were found to be valid methods to estimate group coherence of the resulting movements. It was found that synchronicity of movements between individuals (human–human interactions) increases significantly in the social context. Moreover, Analysis of Variance (ANOVA) revealed that the type of music is the predominant factor in both the individual and the social context.

Keywords: Embodiment, Human body movement, Music research, Social interaction, Statistical

---

7.1 Introduction

The analysis of human body movement is relevant for a number of research areas such as therapy and rehabilitation (Nayak, Wheeler, Shiflett, & Agostinelli, 2000), sports (Martin, 2008), bioinformatics (Buldyrev, Goldberger, Havlin, Mantegna, Matsa, et al., 1995) and neurology (Machulda, Ward, Borowski, Gunter, Cha, et al., 2003). Also in empirical music research, there is a growing interest in how the human body moves and responds to music (Castellano, Bresin, Camurri, & Volpe, 2008; Bernhardt & Robinson, 2008; Thaut, Mcintosh, Rice, Miller, Rathbun, et al., 1996). However, the study of music-driven human body movement is complex because it has to deal with several factors that introduce variability on top of music driven time varying data, such as the neural–muscular–skeletal variability among subjects, the variability in response patterns of single subjects due to learning and training, or the subjects' background (gender, culture) (Stergiou, 2004). The present study relies on Leman's model of music communication, which is based on the notion of embodiment (Leman, 2007). The human body is thereby considered as a natural mediator between mind and physical environment. In this paper, we focus on Leman's social factor (Leman, Desmet, Styns, Van Noorden, & Moelants, 2007) of the above music communication model by studying music-driven body movement of a group of people whose social interaction is taking place in ecological conditions. Music is thereby seen as a social phenomenon and the quantification of social interaction is considered to be a key factor for the development of future electronic mediation technologies and applications. So far, most studies on music-driven body movement have been carried out in controlled laboratory conditions, often with single subjects, limited to simple motor tasks (Toiviainen & Snyder, 2003; Boone & Cunningham, 2001). Although some studies have focused on group behavior, few studies have studied music-driven body movement in real life (ecological) environments (Clayton, Sager, & Will, 2005). In this paper we test

the hypothesis that humans move more synchronous to the beat of the music and with each other in a group than in an individual setting. To test the above hypothesis we rely on different methods yet all methods share a common approach, namely, the definition of similarity measures, which is then applied on the Multivariate Time Series (MTS) matrices. This paper is organized as follows: in Section 2 the experimental design and data considerations are reviewed, Section 3 deals with the analysis, discussion and conclusions are given in Sections 4 and 5.

7.2 Experimental Design and Data Considerations

The experiment was carried out during the Accenta 2007 exhibition in Ghent, where groups of four subjects moved remote Wii sensors while listening to (recorded) music. Sixteen groups of four adolescents participated in the experiment (mean age sixteen). Audiences could watch the performances of these groups. Each group had to perform the task in two conditions, namely, an individual condition, where the participants were blindfolded, and a social condition, where the participants could see each other. Each group had to move in response to six pieces of music. Each piece lasted about 30 s. For a more detailed description of the experimental setup see De Bruyn et al. (2008); Demey, Leman, Bossuyt, & Vanfleteren (2008). The acceleration data from the remote Wii sensors were registered wireless on-line on a laptop computer via the Bluetooth protocol in a PD patch and sampled at a 100Hz rate. Given the design of the experiment (16 groups, four participants per group, six musical excerpts, two conditions, three axes), this resulted in 2304 time series with length $N = 3000$ (30 s / 100 samples s^{-1}). In order to avoid the influence of hesitations and confusions at the beginning and the ending of the task, a 5–25 s interval of the time series was chosen for further analysis, instead of the recorded 0–30 s. The x, y, z dimensions of the accelerometer of the Wii sensor were further reduced to one single dimension, using the formula:

$$a_{t_i} = \sqrt{a_x^2(t_i) + a_y^2(t_i) + a_z^2(t_i)}$$

where a_{t_i} is the global value at time i, and $a_x(t_i)$ the acceleration value for dimension x. Inspection of the resulting time series revealed differences in the range of the accelerations (strong and weak responses). As the occurrence in time of acceleration changes is of interest in this analysis rather than the intensity the amplitude of the calculated accelerations was rescaled to an [0,1] interval. Due to the definition of the total acceleration the minimum value of the series is 0 hence rescaling was based on the division of the values by the maximum value in the corresponding time series.

7.3 Analysis

Before analysis, the time series were tested for stationarity, as it is well known that this condition has a great influence on the stability of correlation coefficients (Yang & Shahabi, 2005). The Unit Root test was used to investigate possible deviations from stationarity. It was found that this assumption could not be accepted in the majority of the MTS. Possible explanations for this observation are drift of the Wii sensors due to the end of the lifetime of the batteries or failing Bluetooth connectivity. Therefore, trend removal was used to obtain stationary MTS. Dynamic Time Warping (DTW) was then used in order to deal with small anticipations and delays in human movement (Parsons, 1987). In this study, we apply a multivariate DTW and a similarity measure based on the cost function. Constraints were introduced in order to speed up DTW calculations. A Sakoe–Chuba band (Sakoe & Chiba, 1978) with a width of 100 was selected which accounts for a 2.5% range. A cumulative distance matrix was then used to find the optimal path, by applying dynamic programming. DTW was calculated for all six possible combinations in the MTS (4 × 2,001) series within each group. Figure 53 shows a fraction of the time series of 1 group (four subjects) before (left panel) and after warping (right panel).

The warped MTS were then inspected for normality of the residuals of the differences between subjects within the group. It was found that, after warping, normality of the residuals could be accepted (Kolmogorov–Smirnov, $\alpha = 0.05$), which was not the case for the original series. The benefit of DTW was investigated by comparing the cross correlations of the original and the corresponding warped series. In the example it can be seen that in the individual context Wii1, Wii3 and Wii4 move in a similar way whilst Wii2 shows a different pattern.
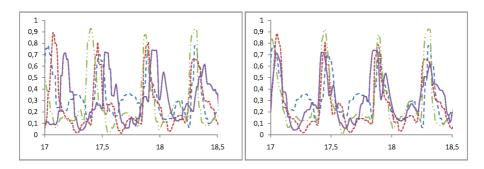
Figure 53. Original (left) and corresponding warped (right) accelerations (Group 1 Song 1).

Cross correlation on the original series show moderate to very low correlations with two non-significant values. The DTW data show higher and significant correlations in all inter subject combinations (Table 10). Each cell in this table represents the correlation (upper row) and the corresponding significance (bottom row) between the subjects. We define the similarity between subject movements as $S_{ij} = (1 - |Corr_{ij}|)$ with S_{ij} between 0 and 1, low values indicating a high (closer related) inter subject synchronicity.

Original (Individual)

	Wii1	Wii2	Wii3	Wii4
Wii1	1	0.012 / 0.589	0.565 / 0.000	0.368 / 0.000
Wii2		1	0.030 / 0.182	-0.144 / 0.000
Wii3			1	0.434 / 0.000
Wii4				1

Warped (Individual)

	Wii1	Wii2	Wii3	Wii4
Wii1	1	0.228 / 0.000	0.711 / 0.000	0.558 / 0.000
Wii2		1	0.258 / 0.000	0.224 / 0.000
Wii3			1	0.592 / 0.000
Wii4				1

Warped (Social)

	Wii1	Wii2	Wii3	Wii4
Wii1	1	0.752 / 0.000	0.782 / 0.000	0.644 / 0.000
Wii2		1	0.767 / 0.000	0.670 / 0.000
Wii3			1	0.613 / 0.000
Wii4				1

Table 10. Example of correlation and significance values in individual and social conditions.

To obtain a measure for group coherence, correlations were averaged for each group, song and condition. A geometric representation of S_{ij} can be used as a tool to classify groups. The plot is constructed by positioning the four participants of a group so that the distances (length of the lines) are proportional to the similarities. In the example (Figure 54, left), the subjects handling Wii1, Wii3 and Wii4 move synchronous, while the subject that is handling Wii2 shows a different pattern in the individual condition. However, in the social condition (Figure 54, right) the coherence improves (Wii2 moves along with the other participants).

Figure 54. Geometric representation of within group similarities.

A plot of the obtained values (individual vs. social) indicates a possible nonlinear trend (Figure 55). The coherence of a group seems to be proportional to the degree of difficulty of the song. For instance, song 4 was well known by the participants and had a clear beat. Even in the individual condition, subjects were able to synchronize very well with the music. Hence the improvement of the

group coherence of the social condition was low in this case. On the other hand, song 3 was found difficult and unknown by the participants, resulting in a higher effect of the social condition on the group coherence. Univariate ANOVA analysis was used to investigate the effect of condition and song. Homogeneity of variance (Modified Levene, $\alpha = 0.05$) and normality (Kolmogorov–Smirnov, $\alpha = 0.05$) could be accepted.
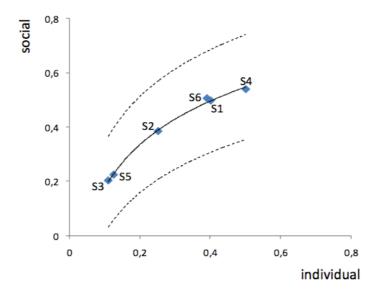
Figure 55. Social versus individual correlations (dashed lines represent 95% interval).

The test indicates that the type of song was the dominant factor while the effect of condition is rather weak but significant. No interaction (condition - song) was observed. A Tukey analysis reveals that the songs can be grouped in three subsets (S3, S5), (S2) and (S1, S6, S4). Finally the DTW cost function was evaluated. Several cost functions have been proposed. For this study the total cost was based on the Euclidean distance of the corresponding warped (x_i, y_j) pairs. Univariate ANOVA was used to estimate the effect of song and condition on the DTW cost. It was found that the warp cost depends mainly on the song and that there is a small decrease in the social condition except for song 4, which has the lowest cost. In order to test the validity of the above method a separate

experiment was set up. The subjects of this experiment were bachelor students of musicology (average age 22) who did the same experiment as the Accenta setup in the laboratory at IPEM. The group coherences were calculated and compared with the results of the experiment. As only three groups were involved the results are only indicative but nevertheless they reveal some interesting information. First of all it can be seen that the social vs. individual coherences of the students are comparable with the results of the Accenta experiment (Figure 56).
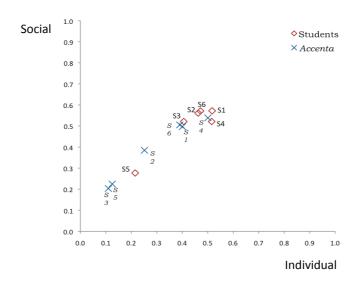
Figure 56. Comparison of Accenta and student correlations.

On the other hand, there are differences over the songs. Song 3 has much higher levels for subjects from musicology, than for the subjects of the Accenta experiment. This can be explained by taking into account that the students all have a musical education background and hence familiar with baroque music. Songs 1, 2, 4 and 6 all show a high coherence for the students indicating a possible effect of age. Song 5 shows an improvement but has low values when compared to the other songs. Song 5 had the most complex rhythm and was influenced by oriental elements. An effect of cultural background may be a possible explanation.

Alternatively, the human–music interaction was studied based on the amount of seconds the participants synchronized correctly with the nominal tempo of the music. This is calculated from the norm of the raw data for each block of 2 s by applying a fast Fourier transform (FFT) over a 4-s moving window with a 2-s over lap. The dominant peak in the Fourier transform is identified and compared with the nominal beats per minute (BPM) of the excerpt for deciding on the correctness of the synchronization. Also the half and double of the nominal BPM were considered as correct. For more detailed information about this method see De Bruyn et al. (2008), Demey et al. (2008). Based on the obtained scores, the impact of a social context on synchronization was studied using ANOVA analysis. Homogeneity of variances (Modified Levene test, α = 0.05) and normality (Kolmogorov–Smirnov, $p < 0.05$) could be accepted. Results show that synchronization results of the participants are significantly higher in the social condition compared to the individual condition (Anova, α = 0.05). The main effects are visualised in an interaction plot in Figure 57.
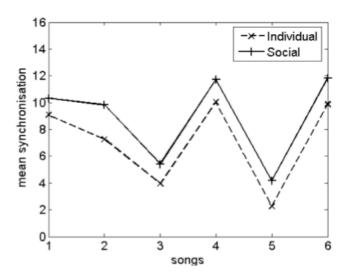
Figure 57. Visualization of the mean synchronization results per song in the individual and social condition.

As can be seen in Figure 57, the songs themselves have a great impact on the synchronization results. A multiple comparison Tukey analysis shows that

participants score significantly lower for songs 3 and 5 than for songs 1, 4 and 6, while the results of song 2 are somewhere in between. This can be explained by the rhythmical complexity of the songs: songs 1, 4 and 6 are pop songs with a very clear beat, songs 3 and 5 can be interpreted either binary or ternary, whereas song 2 can only be interpreted binary but has an unclear beat.

7.4 Discussion

For the MTS data in the experiment presented here, in which subjects were asked to synchronize with the beat of the music, identical results are obtained with the analysis of human–human synchronization based on DTW and the analysis of human–music synchronization based on FFT (Fast Fourier Transform) (De Bruyn et al., 2008; Demey et al., 2008). This indicates the validity of the DTW-based method to analyze movements to music of multiple subjects, which can now be applied to the study of human–human interaction in more generic movements to music. Although the data were collected in an ecological setting with several unknown sources of variance, it was shown that the effect of song and condition can be quantified. The impact of the characteristics of the music is the predominant factor and this is in agreement with the model of musical communication on which this study is based. Humans can indeed decode the intentionality of the music and translate the energy input into movements as a function of musical con tent. Both in the individual and social context this can be quantified. In the social condition there is a benefit as a consequence of imitation effects during the social interaction. The results clearly show that group coherence improves when people move together. Whether this is only the result of direct human–human interaction or that other factors such as presence of public play a role, is not yet clear. The results of the group coherence measure are in good accordance with the analysis of the other variables derived from the collected data. Up to now it is not possible to separate the human–human interaction from the human–environment one in an ecological setting. In order to

improve this type of experiment a new experimental design was proposed. In this design the subjects are not blindfolded in the individual condition, but separated using screens, and 10 songs with carefully selected properties are chosen. A drawback from the Accenta data was the lack of consistent information of the participants' background. The pre-survey could not be used for analysis due to unbalanced results and information about the experience of the subjects during the experiment (post-survey) was not available. In future experiments the use of properly designed surveys need to be included in the experiments. Large-scale user studies for use of the analysis of human–music relationships has been proven to be of great importance (Lesaffre, De Voogdt, Leman, Baets, Meyer, et al., 2008). Finally additional measurements such as video analysis and more sensors per subject are recommended in order to refine the analysis. On the analytical level improvement of the applied techniques and the use of new methods will be investigated. At this moment Correlation optimized DTW based on PCA analysis (Tomasi, Berg, & Andersson, 2004) and quantification of complexity and determinism of the movement data are tested (Sarkar & Barat, 2006). Another important issue is the reduction of calculation time of DTW. Several methods will be tested in the nearby future (Dixon, 2005).

7.5 Conclusion

In this study different statistical techniques were tested for the analysis of human movements to music. Using DTW in combination with CCF and ANOVA it was found that the type of music is the dominant factor of the inter group movements as a response to music stimuli. The effect of condition is low but significant, even in the ecological setting of the experiment. Using DTW and CCF it is possible to quantify interactions and to classify groups by coherence. The outcome of this study enables also to define a statistical path as a tool for researchers and as a guideline for appropriate experimental designs for future research.

8 Gesture as coordinated action in Guqin playing[4]

Abstract

This paper presents the results of a study of the movement velocities of a guqin player using motion capture of 11 different joints during 3 different songs. The aim of this research is to define appropriate statistical strategies as a suitable method to process movement data related to a music performer's gestures, with a focus on uncovering measures that reflect coordinated action. The velocity data of the player was classified using Principal Components Analysis (PCA) and significant differences in playing style were observed using Analysis of Variance (ANOVA). We concluded that the measured joints could be classified in 3 subgroups according to the technique used by guqin playing.

Keywords: guqin, movement, coordinated action, gesture, PCA

---

[4] Desmet, F., Leman, M., Demey, M. & Lesaffre, M. (2011). Gesture as coordinated action in Guqin playing. *Case study, IPEM University of Ghent*.

8.1 Introduction

Playing a music instrument is a complicated activity, which requires years of training and daily practice. The coordination of different body parts covers thereby an essential aspect of the actions that generate and control sound with respect to musical expression and communication. Examples of coordinated body activity are: plucking a string of a guitar with the right hand and shortening the length of the string with the left hand, or lip pressure and attack and key control on the trumpet.

The coordination of body parts during music playing can be viewed in terms of body movement, that is, changes in physical position. However, these movements form part of an action-oriented ontology of the musician: this is a set of patterns that are linked to frames of reference (body schemata) for motor effectuators, and to representations (body images) of actions, posture, and body, in relation to the musical task (Gallagher, 2005; Henbing & Leman, 2007; Leman, 2007). If viewed from this perspective, coordination in music playing is purposeful, and therefore, related to action sequences that aim at having an expressive character, in one way or another (Rolf et al., 2009). The goal-directed nature of such movements, and the fact that they may be accessible as action chunks ('out of time') to the human mind makes such movements peculiar. Therefore, it is straightforward to call such a body movements: gestures (Godøy & Leman, 2010).

The study on how humans coordinate their movements has been investigated in view of different tasks (Cordo et al., 1994; Stergiou et al., 2001), including musical tasks (Large, 2000; Li et al., 1999; Wanderley et al., 2005). Thereby, the focus has been on technical playing (Baader et al., 2005; Shan & Visentin, 2003), on the understanding of spontaneous movement in response to music (De Bruyn et al., 2008; De Bruyn et al., 2009; Desmet et al., 2010; Toiviainen et al., 2009), and on the gross coordination within a broader

framework of gesture-based musical communication, in which the sharing of expression is a key element (Naveda & Leman, 2010). In addition, musical expression has been studied with regards to purposive actions (Schogler et al., 2008). However, despite the recent interest in music and body movement, less attention has been devoted to the coordination of different body parts of the performer.

8.2 Background

8.2.1 Embodied music cognition

The present study is carried out within the research framework of embodied music cognition, in which coordination of body parts is considered within a framework of mediations between the mental world of intended actions, experiences, feelings, emotions, expressions, and the world of moving body parts, or energetic patterns (as sound, visual information) (Leman, 2007). A general communicative framework for music is adopted, in which the player is assumed to encode expressive patterns into sound, which the listener decodes on the basis of an embodied engagement with sound patterns.

The focus of this study is on the coordinated movements of a single musician, namely a guqin player. The Chinese guqin music has been the topic of several studies that focused on playing gestures (Henbing & Leman, 2007), guqin sound synthesis (Penttinen et al., 2006), and shared expression (Leman et al., 2009).

8.2.2 Guqin music

The guqin (pronounced ku-chin in English) proves to be interesting for our research goal. The instrument belongs to the family of the zither, a plucked stringed instrument, which consists of a long and narrow hollow wooden box and functions as a sound box. The construction of the guqin is shown in Figure 58.

The upper part functions as a fretless fingerboard on top of which there are 7 strings attached, each about 110 cm long (Penttinen et al., 2006). The way in which guqin music is played, namely by plucking the string with the right hand and moving the finger of the left hand over the string, makes it suitable for a detailed study of the body movement in relation to sound and the musical communicative context. Indeed, the guqin has no bow, and fingering is directly related to sound, as there are no frets to interfere within the sliding. Although the playing technique is rather complex, the sliding-tones in guqin music can be conceived as sound patterns that reflect aspects of the playing movement, without much intermediate technology. Studying guqin music is thus particularly interesting for embodied music cognition research because the encoding of playing gestures into sound patterns are directly linked with each other, which facilitates the study of possible relationships between movement and sound.
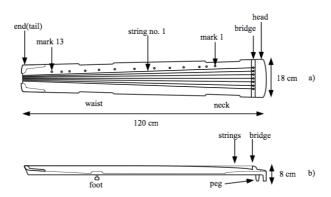
Figure 58. Construction of the Guqin a) top view b) front view

The seven strings are tuned as a pentatonic scale. The basic tuning of the open strings is C2;D2; F2;G2;A2;C3, and D3 ranging from the lowest string (no. 1) to the highest (no. 7). Zhang Jianhua in Beijing constructed the guqin used in this experiment in 1999. The boards are made of fir, and the roughcast consists of deer horn powder and raw lacquer. Shangy steel-nylon strings are used. For a more detailed description of the construction of the instrument see (Henbing & Leman, 2007).

Nowadays, the Guqin is played on a table while placed on anti-slip mats. The neck of the instrument is positioned at the right side. The right hand plucks the strings (between bridge and first mark) whilst the left hand is used to press the strings against the top plate of the body, which produces smooth sliding tones on the fretless instruments. The most frequently used compound glides are based on a variation of movement upwards or downwards from the targeted note but there are some less used techniques that have glides on both sides of this note. Finally, there are a number of vibrati, i.e. small variations on either side of the main note. In the glides and vibrati, a lot of attention is paid to the speed, rhythm and size of the movements. This does not imply that one of them may be fast or slow or big or small. It may start fast and slow down or vice versa; the acceleration or deceleration must occur in the right place and, as a result, there may be variations in the attack at the beginning and firmness of the close of the technique. A vibrato, for example, may start with a quick, fairly large movement, but may gradually slow down and becomes smaller. For more details, see (Henbing & Leman, 2007).

8.2.3 Research questions

The following hypotheses are considered here: (i) the movements of a music playing body are driven by a small number of action strategies, (ii) a clear difference in playing style between different songs can be observed.

The coordination of movements is explored using Principle Component Analysis (PCA) to obtain a non-redundant set of variables for a compact description of certain processes or phenomena (dimensionality reduction) (Daffertshofer et al., 2004). The difference in playing style is analyzed using Analysis of Variance (ANOVA).

This paper is further organized as follows: in section 8.3 methods used are described, section 8.4 is devoted to the analysis process, principles and

techniques, and in section 8.5 the results of the study are presented followed by a discussion and conclusion in sections 8.6 and 8.7.

8.3 Methods

8.3.1 Musical stimuli

Three musical fragments (P1, P2 and P3) of a traditional Chinese piece, called "Missing an Old Friend", were played on the guqin by Henbing Li, an experienced guqin player educated at the Central Conservatory of Music in Beijing. Each fragment was about 30 seconds in length. Of all the chosen musical fragments, P1 and P2 have a rather fluent and clearly structured melodic line. In contrast, P3 has a more narrative character with a less fluent melodic line.

8.3.2 Motion capture

A motion capturing system with infrared cameras and reflective markers was used to monitor the movements of 11 different markers attached to different joints of the musician (see Figure 59). The system used is a Qualisys Motion Capture System from Sweden at a sampling rate of 100 Hz.

Marker	Position
1	Center forehead
2	left shoulder
3	left elbow
4	left wrist
5	nail left thumb
6	nail left pink
7	right shoulder
8	right elbow
9	right wrist
10	nail right thumb
11	nail right pink
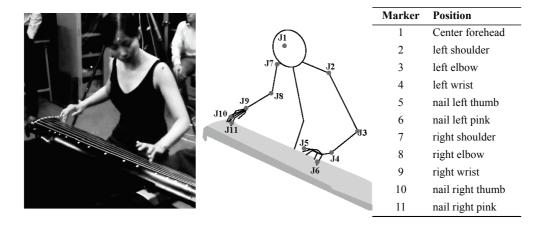
Figure 59. Experimental setup: Position and labeling of the joints

8.3.3 Raw data

The raw data delineate a 3-dimensional vector function of time. As the aim of this study is to investigate and compare movement patterns, bias in favor of particular directions should be avoided and the vector function should be reduced to a single scalar function of time. For each marker position in each piece the magnitude of velocity was calculated over intervals of 250 ms. In this way 33 time series of 109 samples were obtained corresponding to 11 marker positions in 3 musical pieces. The authors are aware of the relatively large time interval resulting in a sampling frequency of only 4 Hz, however, the movements under study are slow and fluent.

Due to the nature of the measure (magnitude of the velocity) normality cannot be assumed (*KS* test, $p < 0.05$). This is quite common in datasets of human movement on music (Desmet et al., 2009). Moreover, if the sample size is small, there is a degradation of the correlations between poorly distributed variables. Variable transformation is a commonly used method to overcome these problems and enhance the results of statistical techniques. In this study, a square root transformation of the variables was applied to overcome this problem. An illustration of the original and transformed magnitude of the velocity corresponding to the head of the player in the first piece is shown in figure 61. The transformed data allow the assumptions of normality (*KS*, $p > 0.05$) and homogeneity of variance (*Levene*, $p > 0.05$) to be met. Grubb's test was used to check for outliers and extreme values. The H_0 hypothesis (no outliers present) could be accepted for the transformed data. The results prove the usefulness of statistical methods in which the assumptions of normality, homogeneity of variance and absence of outliers are used as prerequisites.
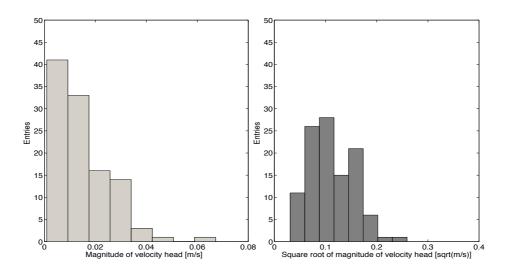
Figure 60. Example of histograms before (left) and after (right) transformation (Head - Piece 1)

8.4 Analysis

In this section we describe the analysis process, principles and techniques used for testing the three stated hypotheses. Firstly the use of PCA for classification of movement velocities is explained. Secondly the difference in playing style for the different songs is studied using ANOVA.

Principal Components Analysis (PCA) was used to investigate the first assumption that the proposed action strategies can indeed be extracted from the registered movement velocities. PCA is often used to describe the relationships between variables in terms of underlying, but invisible, quantities called factors (Johnson & Wichern, 2002). In the present study, PCA is used to test the hypothesis that the movement velocity patterns can be divided in a small number of action strategies.

The appropriate tests were used to evaluate this type of analysis. The tests to determine the criteria for the minimum amount of values, the absence of biased variables, the normality and a sufficient heterogeneity of the data show that PCA can be used. The PCA analysis was done for each piece separately ($n = 109$, 11

variables). No missing values were present in the dataset. A cutoff of eigenvalues >1 was used as the criteria for component extraction and Varimax rotation with Kaiser normalization was selected. Kaiser-Meyer-Olkin (KMO) and Bartlett's tests were used to check for sampling adequacy and sphericity. For all 3 pieces, the KMO values are larger than 0.5 indicating a sufficiently high proportion of variance caused by underlying factors. The Bartlett's test results were significant in all cases proving that the correlation matrix is not an identity matrix and that the strength of the relationships is high ($p < 0.05$). For all 3 pieces, 3 principal components are found with an eigenvalue higher than 1. The explained total variance is 76.25 % (Piece 1), 75.83 % (Piece 2) and 77.61 % (Piece 3) and each component contributes with comparable (20 - 30%) proportions to the total variance after rotation. This can be seen in the scree plots in the first column of Figure 61 where the components with an eigenvalue larger than 1 are indicated.

		Components piece 1			components piece 2			components piece 3		
		1	2	3	1	2	3	2	1	3
body part	Joint									
head	1		0.432	0.514	0.545					0.623
left shoulder	2			0.687		0.324	0.528			0.651
right shoulder	7			0.715		0.590	0.524			0.923
left arm – hand	3	0.824			0.820			0.828		
	4	0.939			0.998			0.989		
	5	0.657			0.689		0.523	0.639		
	6	0.794			0.849		0.421	0.751		
right arm – hand	8		0.667	0.565	0.849				0.742	0.502
	9		0.789		0.736				0.925	
	10		0.804		0.687				0.854	
	11		0.823		0.585				0.775	

Table 11. Rotated component matrices

The rotated factor matrices enable to define the components as shown in Table 11. The output can be interpreted better by removing the clutter of low

correlations (< 0.3) and by rearranging the joints and swapping component 1 and 2 of Piece 3 (Table 11). The factor in the first column is constituted from the movements of the left arm and hand, the factor in the second column contains movements from the right arm and hand while third column contains the head movement and the movement of the shoulders for Piece 1 and 3 while for Piece 2 the movement of these body parts are also in the second column. The grouping of the body parts in these 3 factors can also be seen in the component plots shown in the middle column of Figure 61.

The scatter plot of the calculated versus predicted correlations shown in the right most column of Figure 61 shows that the principal components can be regarded as good predictors for the determination of correlations between the joints. The dashed line represents the ideal maximum correlation (slope = 1). The solid line represents the calculated regression of the predicted versus observed data. The results of the linear regression for each piece are summarized in table 2. The linear model can be accepted (high and significant F-values) and the small deviation of the significant values of the slopes (high and significant t-values) from 1 quantifies the validity of the data reduction into the proposed components. In the regression plots (third column of Figure 61) the dashed line represents the ideal line with a slope of 1 and the solid line is the actual regression. The results of the regression analysis (Table 12) show that the linear model can be accepted and that the slopes of the fitted lines can be considered as equal and close to 1.
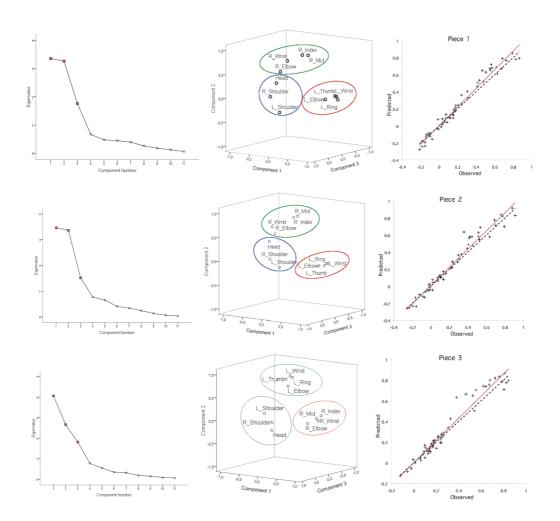
Figure 61. Scree plot, Component plots and predicted versus observed correlations for the 3 pieces.

piece	R^2	*Adj.* R^2	*F* (model)	*p*	slope	*t* (slope)	*p*
1	0.972	0.971	1828.28	< 0.01	1.093	42.76	< 0.01
2	0.960	0.960	1284.53	< 0.01	1.067	35.84	< 0.01
3	0.952	0.951	1045.13	< 0.01	1.076	32.33	< 0.01

Table 12. Regression analysis of Predicted versus Observed Correlations.

When relating the movement of the body parts to the actual playing of the guqin it is clear that the movement of the left hand and arm enables the gliding, while the right hand and arm performs the plucking of the strings. The other body

134

parts, namely both shoulders and the head, can be related to a more corporeal expressive type of movement. Therefore the gliding can be attributed to Factor 1 (Piece 1 and 2) and Factor 2 (Piece 3), the plucking to Factor 2 (Piece 1 and 2) and Factor 1 (Piece 3) and Factor 3 can be related to the expressive corporal component. The connection between action strategy (gliding, plucking, expressing) and body part (left arm, right arm, head/shoulders) suggests that for Piece 1 and 3, the movement of the right elbow (joint 8) consists of a combination of the plucking and corporal component. The shoulder movements used in Piece 1 and Piece 3 are not related with the gliding or plucking components, while in Piece 2 they are connected to the plucking component. The use of the head movement is different in all 3 pieces. Piece 1 reveals that the movement of the head is defined by a combination of the corporal expressive and the plucking component. In Piece 3 the movement of the head is only the result of the expressive corporal component. Keeping all this in mind, the movement patterns in Piece 2 are quite interesting: the movement of the head is only defined by the plucking component. The latter is also reflected in the shoulder movements.

In short, PCA confirms the hypothesis that the movement velocity patterns of the joints are determined by a small number of action strategies for the three different pieces and that action strategies are associated with body parts. PCA makes it possible to reduce the initial 11 variables to 3 new variables, namely, Corporal/Intentional (Head, Shoulders), Gliding (Left arm, left hand) and Plucking (Right arm, right hand). The resulting time series for the 3 new variables are presented in figure 62.
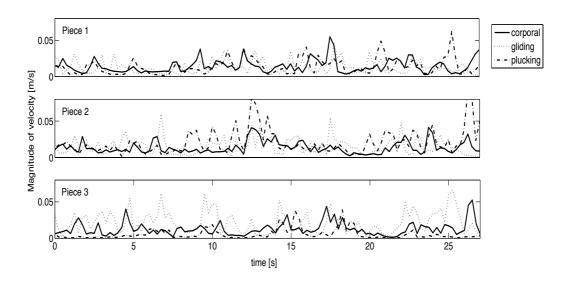
Figure 62. Component velocities time series plots.

ANOVA of the velocities was used to look for significant differences between the extracted components of the 3 pieces. The results from these analyses are shown in Figure 63 and Table 13. The movement velocities of the corporal component are the same for all pieces. The use of the gliding component is significantly different in Piece 3 than in the other two pieces. The use of the plucking component is significantly different in all pieces. Piece 2 has a significantly higher value and Piece 3 has a significantly lower value when compared to Piece 1.
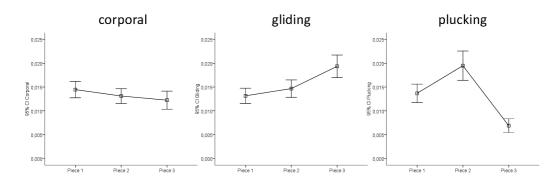
Figure 63. Mean plots and 95% confidence intervals of the components for the 3 pieces (ANOVA).

Component	*F*	*p*
Corporal	1.65	0.195
Gliding	10.66	< 0.01
Plucking	30.78	< 0.01

Table 13. Results of the ANOVA analysis

8.5 Results

The results provide evidence for the stated hypothesis, namely, that (i) the movements of the joints of a music playing body are determined by a small number of action strategies, (ii) a clear difference in playing style between different songs can be observed.

By using PCA, it was shown that the movements of the joints can be reduced to 3 components, which correspond with the technique of playing the Guqin. These are Corporal/Intentional (Head, Shoulders), Gliding (Left arm), Plucking (Right arm). The extracted components reveal differences between the 3 pieces. It was found that the plucking component contributes the most to different action strategies.

By combining PCA and ANOVA analysis, detailed information about the movement of a musician can be obtained. If considered from this perspective, coordination in music playing is related to the notion of gesture. This is movement and likely also intended (i.e. goal-directed) movement, which is related to actions that aim at generating sounds that are expressive in one way or another. This analytical strategy opens doors for further research. It would be interesting to investigate performers with different skill levels and to analyze other instruments. This can also be a promising technique in order to define new instruments or to visualize gesture induced audio, based on audience movement capture.

8.6 General discussion

Within the research framework of embodied music cognition (Leman, 2007) the human body is seen as a mediator between physical environment (where musical actions are seen as body movements) and subjective experience (where musical actions are seen in terms of values, intentions, expressions, feelings). Associated with this is a musical communication model in which the player encodes expressions through corporeal articulations, and in which the listener decodes the transmitted information through corporeal imitations (in an overt or covert way) (Leman et al., 2009; Overy & Molnar-Szakacs, 2009). Although recent studies have contributed to a more detailed concept of the human body as mediator in music playing (e.g. (Palmer et al., 2009)), only a few studies have actually focused on the coordination of different body parts during music playing. As music playing is based on purposive gestures, involving time-dependent musical targets related to the expressive domain, it is generally assumed that the coordination of different body parts plays an essential role (see e.g. Williamon, 2004). The viewpoint of the human body as mediator between physical environment and subjective experience provides a framework that may clarify the role of the coordination of different body parts, in particularly in relation to purposive gestures.

However, the present study is limited by the fact that gestures are considered as a continuous movement. Hence, the resulting observations are defined by general action patterns, such as gliding and plucking during a (rather long) melodic phrase. A more fine-grained approach could consider the segmentation of the continuous movement into gestural units. As shown in (Henbing & Leman, 2007), the most elementary gestures involve the movement of a finger from point A to point B on the string. Typically, a tone in guqin music can be conceived as a concatenation of such elementary gestures, and a motive can be conceived as concatenations of tone-related gestures. If seen from this

perspective, the elementary gestures, as studied in (Henbing & Leman, 2007), and the general gestures, as studied in the present study, provide two different scopes on gestures, namely, a fine grained one, and a global one. Clearly, further work on the intermediate level, where gestures pertain to expressive action units, such as whole tones or sequences of tones (called motives or figures), is needed. From this perspective, (Godøy & Leman, 2010) has suggested that gestures have a hierarchical and nested nature, with target points. In order to further determine target points, it may be of interest to examine head movements in relation to the coordination of body movements in more detail, using physiological data (e.g. electromyography).

Although coordinated body movements in musical playing can be considered within a broader theoretical framework of action perception couplings e.g. (Friston, 2010; Schogler et al., 2008), many studies have reported on specific brain activities, such as the motor control of a single part of the body (Fuchs et al., 2000; Haueisen & Knösche, 2001; Watson, 2006), or the brain activity during a limited non-musical task (Chen et al., 2008; Magescas et al., 2009; Zatorre et al., 2007). Recent studies aim at the continuous monitoring of auditory-motor interactions (feed forward or feedback) during playing (Lindenberger et al., 2009). The ability of this fine-grain correction of individual movements is unique in music and may be controlled by hierarchical auditory-motor frames of reference that guide actions (Feldman & Levin, 2009).

Important for our study is that specific regions in the brain correlate strongly with movement velocity, independent of movement direction and mode of coordination (Gross et al., 2002; Kelso et al., 1998). Consequently the analysis of movement velocity can provide information on brain activities stimulated during music performance. Furthermore, understanding the mechanism of coordinated action is of importance in musical education (learning) in order to develop skills to control delay and anticipation and to enhance performance. Within the domain of medical treatment and rehabilitation, knowledge about this

mechanism can prove to be of significance. We can for instance try to improve the recovery of patients with movement disorders due to illness (Parkinson) or trauma (stroke patients).

8.7 Conclusion

By studying coordinated action in music playing, it is possible to define statistical strategies that process movement data in relation to a music performer's gestures. In the present paper, a combination of statistical techniques is used, with a focus on uncovering measures that reflect coordinated movement of body parts in relation to intentionality. The methods involve Principal Components Analysis (PCA) used as data-reduction method followed by Analysis of Variance (ANOVA) to indicate significant differences between musical pieces. It was found that the movements of 11 body parts can be reduced to 3 groups and that these groups can indicate different playing styles between different music pieces.

9 A statistical analysis framework for assessing a clarinet player's performer gestures in relation to locally intended musical targets[5]

Abstract

Musicianship is known to display high-level cognitive skills, which involve different aspects of mental processing and corporeal control. Of particular interest is the match between the musician's mental focus on musical targets (the so-called musical intentions) and the expressive (or so-called auxiliary) body movements. To what extend are these related to each other? And what does this relationship reveal about mind-body connections? To approach these questions, a case study was set up around a clarinet solo performance played from score, covering a style of music unfamiliar to the player. The clarinetist's movements were recorded with an optical movement tracking system. A statistical analysis method was developed, to account for movement data in relation to the potential musical intentions and targets. The bottom-up movement analysis method was validated with the performer's annotations of targets in the musical score and the performer's annotations of musical gestures in the performance video. The results reveal that the mental focus on musical targets is related to bodily expression. This finding supports the idea of an embodied model of musical syntax processing, which is strongly related to corporeal gestures.

*Keywo*rds: musical expression, performance, clarinet, statistical analysis, segmentation, annotation, gestures, intentionality, kinematics

---

9.1 Introduction

Musical performance depends on the ability of a player to display high-level cognitive skills, whereby different aspects of mental processing and corporeal control are needed (Parncutt & McPherson, 2002; Zatorre, Chen & Penhune, 2007). Music can thereby be understood as an art that is related to "moving sonic forms" (Leman, 2007). Music playing, in particular can be understood as the encoding of a musical idea into moving sonic forms, while listening can be understood as the decoding of these moving sonic forms into something meaningful. The intended origin of the encoding, and the result of the decoding should not necessarily address the same semantic content (which is rather typical for verbal communication). Instead, it is assumed that a successful communication of music is highly dependent on whether moving sonic forms appeal to a repertoire of corporeal articulations that humans tend to use in social contexts. In that respect, moving sonic forms can be seen as sound traces of body movements, or more particularly of expressive sensitivities and articulations, that are subtle corporeal nuances that express emotional states, drives, attitudes, motility, agency. Therefore, the display of high-level cognitive skills in musicianship is a matter of mental and corporeal control over a musical instrument through which expressive sensitivities are translated into moving sonic forms or patterns.

In order to better understand the nature of the mental and corporeal control that constitutes the musical performance, several studies have addressed musical playing in relation to musical intentions or targets (Davidson, 2007; Repp & Keller, 2010; Thompson & Luck, 2008; Kirke, Miranda & Zhang, 2010). Indeed, in the theory of musical communication mentioned above (Leman, 2007), it is assumed that moving sonic forms in music (i) originate from musically intended actions on a musical instrument, (ii) appeal to our sensitivity to

corporeal nuances (iii) in order to stimulate listeners intended actions while listening to music.

In several studies, this musical intentionality has been conceived in relation to a particular expressive performance style that is associated with an entire (often short) musical piece. Typically, musicians are asked to perform a musical piece in a number of targeted expressive sensitivities, such as 'light', 'heavy', 'expressive', 'immobile' and so on. Recordings of these performances provide us with a context for comparative analysis of subtle differences in audio and movement patterns. The goal is then to extract objective features from audio and body movement, and to correlate them with the expressed sensitivity (De Poli, Mion & Roda, 2009; Mion, De Poli & Rapana, 2010).

Using this paradigm for the study of clarinet performances, Canazza et al. 1997 found that audio features of clarinet performances played in a "light", "heavy" or other style could be connected to a control space that would allow the reconstruction of the characteristic expressive style from a neutral example. In a number of studies, it was shown that reduced information of one of these modalities (such as movement, without audio) might be sufficient to perceive the performance style (Dahl & Friberg, 2007; Davidson, 1993; Nusseck & Wanderley, 2009). Wanderley et al. (2005) studied body movements during standard clarinet performances (as played in concert) opposed to exaggerated expressive and immobile performances and found that general movement features may characterize the intended expressive sensitivities. Similar studies exist in relation to the expression of a piano performance (Castellano, Mortillaro, Camurri, Volpe & Scherer, 2008; Clarke & Davidson, 1998). These studies indicate the important role of body movement as a corporeal mediator of expressive sensitivities that are encoded in audio structures.

However, these studies reflect on expressed sensitivities in terms of general stylistic targets. Less attention is thereby given to the way in which particular local sensitivities, or local musical targets, are communicated. In line

with the general idea that musicians seem to move in correlation with local sensitivities in moving sonic forms, Davidson (2005; 2007) observed that the performer uses particular movement shapes that refer to specific and identifiable expressive locations within the context of a whole performance. When studying pianists, these shapes are accounted for by the notion of an "expressive centre of moment", which is the physical centre within the body through which the musically expressive information is produced by more extended body parts. The study suggests ways of addressing intended expressive sensitivities at a local structural level, rather than a global level that is associated with a performance style. However, many questions regarding the precise relationship between expressive sensitivities, musical intentions, and body movement need to be clarified and investigated further, especially also in view of a refinement of the musical communication model.

In order to address these questions, the present study focuses on the relationship between the musician's expressive body movements and locally intended targets and the supporting generative structure in the musical score. The latter can be conceived as the performer's focal points within the expressive discourse of music playing. We thereby assume that, based on these locally intended targets, the performer's gestural skills subsume a grouping of movements into larger movement patterns or gestures (Godøy & Leman, 2010). As such, the locally intended targets can be understood as reference frames for musical motility, which reduces the cognitive load during the musical performance. Without such targets and grouping, one would focus on all notes equally as much, and therefore it is likely that there would be more cognitive load, which would make a musical performance more difficult and certainly less musical (as it is with novice players). Therefore, a focus on targets and grouping in relation to musical information and gesture is a first step towards a better understanding of how possible reference frames for musical motility may facilitate music playing.

In music analysis, local targets are typically related to different parameters of the musical structure, such as long notes, high notes, notes related to the tonal structure, or rhythmic structure and so on. However, in order to identify the local targets of a particular performance, it is necessary to rely on the performer's introspective analysis. Moreover, one should keep in mind that local targets are present in any expressive performance style, no matter whether this style is 'light', 'heavy', or 'exaggerated expressive'. This approach thus requires the collaboration of a performer who assesses his / her own performance in terms of local intentions, and chunked gestures. This approach supports the general idea that musical targets can be understood as intended goals that result in sound through the mediation of expressive gestures. This idea is inspired by theoretical accounts on embodied music cognition (Leman, 2007) and musical gesture (Godøy & Leman, 2010). It provides the basis to the present study that focuses on how body movements relate to local musical targets. This analysis of the intended musical targets (the subjective method) can be combined with motion capture and statistical analysis (the objective method).

To sum up, we made the hypothesis that there are mutual relations between a clarinet player's locally intended musical targets, his / her expression as a corporeal nuance, and its trace as a sonic moving form. It is assumed that these relations manifest themselves through an objective (bottom-up) observation of corporeal articulations and a (top-down) subjective assessment of intended targets and their gestural expression. The goal of this paper is to clarify this relationship through the development of a framework for statistical analysis of musical gestures.

This paper is organized as follows: first, a general description of the experimental setup and analysis approach is given. Second, an objective bottom-up approach is developed for corporeal segment extraction. Third, we consider further clarifications of segment extraction using a subjective top-down approach whereby gestures are annotated. Fourth, a comparison between the objective and

subjective methods is presented. Finally, an overview of the obtained results is given. This is followed by a general discussion and conclusions.

9.2 Experimental setup and analysis approach

The development of an analysis framework for musical gestures is based on an experiment in which a musical performance by a skilled clarinet player was recorded on a motion capture, audio and video system. A distinction was made between two analysis approaches. First, an objective analysis, which aimed at extracting segments from recorded motion capture data, video and audio data. Second, a subjective analysis that aimed at annotating musical structural interpretation, target intentions, and gestural types from the viewpoint of the clarinet player. The goal was to see to what extend it is possible to relate those two viewpoints. The experiment was conducted at the laboratory of IPEM, Ghent University.

9.2.1 Participant and task

The clarinet player who participated in this study is a skilled clarinetist with a musical training of about 15 years. He was asked to perform four musical pieces with different styles, namely Schumann, Fantasiestücke op 73 (first part); Bach, orchestral Suite nr 3 BWV 1068 (air); Mozart, Clarinet quintet K581 (trio 2 from minuet); Marais, Basse de viole and BC, Troisième livre (gigue). Compared to W.A. Mozart's Clarinet Quintet and Schuman's Fantasiestücke, which belong to the clarinetist's repertoire, the clarinet player had no experience in playing baroque music. Furthermore, to make sure that the playing would be as unselfconscious as possible the clarinetist was told that the aim of the performance recording was to provide material to do a test of the motion capture system's possibilities. No additional performing instructions were given and the clarinet player received the musical scores just a couple of days before the recording took place. At the time of playing, the clarinetist was not aware of his

part in the experiment. He understood the playing to be a try-out of the Optitrack system. Therefore, there only was a limited amount of time to prepare the pieces, which contributed to a rather intuitive playing session.

The Marais piece was chosen as exemplary for developing our analytical model. Although the music of Marais is not considered to be that technically difficult, the piece is quite challenging for a player who is unfamiliar with the specific ornaments in French baroque music. Given the spontaneous nature of the experiment, the melodic interpretation was determined by an intuitive way of playing, and described as "gay" at the beginning of the score. This indication could be considered as the generally intended expression. In the end, the player decided to play long and sustained phrases while making a distinction between the more rhythmic and fluent melodic passages.

9.2.2 Recordings

The recordings of the movements of the clarinet player were done with an OptiTrack motion capture system, audio recording system, and video camera. Thirty-three infrared reflective markers covered the musician's full body, including the clarinet. Twelve infrared sensitive cameras were used; enabling us to get a three-dimensional positional readout of infrared reflective markers at a sampling rate of 100 Hz. Together with the OptiTrack recording, audio was recorded using a microphone (Shure Bèta 87A – mono) in combination with a Max/MSP patch. The entire performance was recorded on video using a handy cam (Canon Legria HFS100) positioned on a tripod in a fixed frontal view.

9.2.3 Analysis framework

A basic distinction can be made between an objective bottom-up approach which is based on the analysis of raw motion capture data, and a subjective top-down approach that starts from the analysis of the musical score, and a manual annotation of video and audio data. The bottom-up approach results in segments

of the kinematic data derived from the mocap data while the annotations of the player result in gestures related to expressiveness and intentionality.
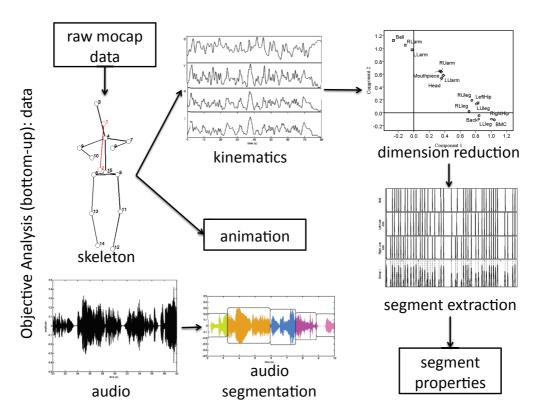
Figure 64. Framework for bottom-up data pre-processing and segmentation.

The bottom-up objective approach is shown in Figure 64. The raw data consist of three-dimensional time series of marker motions; audio, video, and musical score information. In this approach, audio, video, and musical score are basically used to complement the analysis of the motion capture data. The raw motion capture data are first converted into a skeleton. By means of the resulting skeleton data displacement, velocity, acceleration and jerk time series and their norms are calculated and the corresponding skeleton animation is generated. The next step consists of the dimensional reduction by means of a principal components analysis (PCA). By combining the groups as a result of the PCA, a new data set for movement flow is calculated and this data set is used for

comparing the segments of the reduced data set with the annotated gestures of the top-down approach. Finally, each gesture is characterized by a set of properties. The mocap data were segmented using a segmentation algorithm written in Matlab, while the audio data were segmented using the MIR toolbox (Lartillot & Toiviainen, 2007). The audio segmentation was added to enable a faster detection of a selected annotation in the score.

Figure 65 illustrates the subjective top-down approach, which, in our current approach, is carried out by the player (see the task description below). It consists of score analysis, video annotation, and audio annotation.
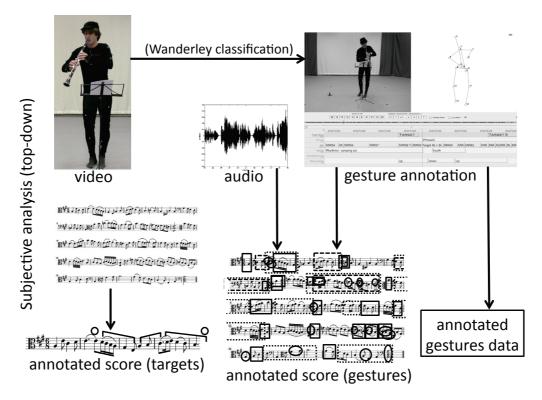
Figure 65. Top-down subjective analysis and annotation workflow.

As illustrated in Figure 66, the resulting data sets from the objective bottom-up analysis are then combined with the subjective top-down annotations for further comparative analysis. In this comparative analysis, we use ELAN in combination with Matlab. ELAN provides a multi-modal browser for annotations

of selected gestures (Wittenburg, Brugman, Russel, Klassmann & Sloetjes, 2006), while Matlab is used to extract features from a selected gesture. In addition, Matlab enables us to rotate a selected three-dimensional skeleton representation in combination with the additional properties.
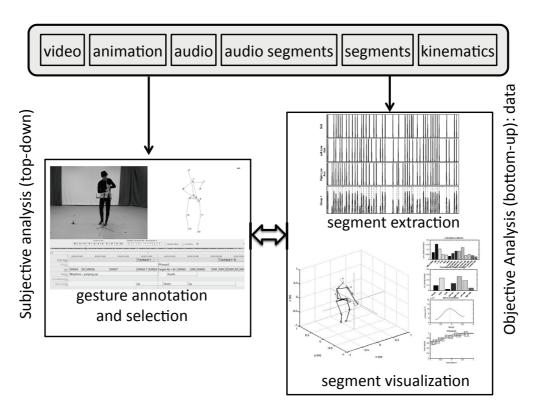
Figure 66. Comparison of objective and subjective data.

In what follows we provide more detailed information on each approach: (1) objective bottom-up, (2) subjective top-down and (3) comparison between both approaches.

9.3 Corporal movement segment extraction – an objective bottom-up approach

The goal of this section is to describe the methodology that is used for the segmentation of the motion capture data.

9.3.1 Mocap data preparation

The raw data of the motion capture recording consisted of a file in c3d data format and the corresponding audio file. These files were imported in Matlab for further analysis, using the Mocap toolbox import function (Toiviainen & Burger, 2010).

The Euclidean positions of the markers were first converted to the international system of units (SI units) (Taylor, 1995) and an additional marker for the body mass center of the player was added. The body mass center was defined at a position of ten cm above the centroïd of the polygon formed by the four hip markers. The choice of this position was based on the human engineering standards of NASA (NASA-STD-3000, 1995). The initial body mass center position was calculated at the beginning of the experiment when the player was asked to stand still with stretched arms for calibration of the motion capture system (T-pose). This position was used as the (0,0,0) position in the Euclidean space by translation of all marker vectors. By adding this additional body mass center marker other body movements such as the body weight displacement, back curl, up and down movement of the clarinet bell as defined by Wanderley et al. (2005), could be extracted from the data.

The data were synchronized with the audio and then smoothed to remove the noise related to the motion capture system. The latter is of significant importance especially when calculating the derivatives of the positional information like the acceleration, which is used as a key kinematic variable for the study presented here. From several smoothing techniques (for an overview see Bowman & Azzalini, 1997), LOWESS (Cleveland & Devlin, 1988) gave the best result. This technique is based on a locally weighted smoothing with a polynomial order of 1 or 2. The smoothing parameter corresponds with the window size.

The optimum value of the window size was obtained as shown in Figure 67. We are well aware of the fact that robust differentiators (e.g. Savitsky-Golay) can be used to reduce noise but the problem of selecting an optimum choice for the window size remains the same. First, a reference data set was recorded while the player maintained the T-pose (standing upright with arms stretched horizontal outward (a)). Second, the distribution of this recorded data was calculated for the 3D positional data of all markers on the body of the player after subtracting the mean and trend removal. The result is a Gaussian distribution around zero resembling the noise of the system with a variance related to the calibration process (T-pose) (b). Third, a similar distribution from the movement data during the music performance was obtained by applying a series of smoothing thresholds and subtracting the smoothed data from the corresponding data before smoothing (c). When applying too much or too little smoothing to the motion data the resulting distribution after subtraction with the original data will be respectively too wide or too narrow (d). As a criterion for selecting the smoothing value an F-test for equality of variances was performed on the distributions obtained from the T-pose (b) and that from the movement data (e). The F-ratio of the reference and motion data noise with a value closest to 1 is then used to select the corresponding threshold value for the optimal smoothing (f).

The same procedure was applied to the calculation of the derivatives of the positional data namely the velocity, acceleration and jerk.
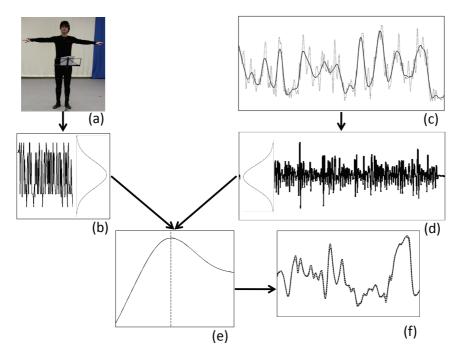
Figure 67. Smoothing optimization procedure (a) T-pose, (b) system noise, (c) original data and smoothed data for a selected window size, (d) difference between original and smoothed signal, (e) F-ratio between system noises and (f) resulting smoothing at optimum window size.

Finally, by means of this data set the player skeleton was calculated (see Table 14). A Matlab function was written which requires only the input of a c3d file and the corresponding audio file, to automate the data preparation process.

marker	skeleton part
1	clarinet mouthpiece
2, 3	clarinet bell
4, 7	right hip
5, 6	left hip
8,9, 10	back
11, 12 13	head
14, 15, 16	left upper arm
17, 18, 19	left lower arm
20, 21, 22	right upper arm
23, 24, 25	right lower arm
26, 27	left upper leg
28,29	left lower leg
30, 31	right upper leg
32, 33	right lower leg
(4, 7, 5, 6)	body mass center (calculated)

Table 14. Marker - skeleton conversion.

9.3.2 Segmentation and filtering

Given the kinematic variables it was then possible to extract a set of segments (primitives), which were used for further analysis. With that aim in mind we developed a method that is based on statistics.

The segmentation and filter method can be described as follows: during a first step of the segmentation, the time series of the acceleration were normalized between zero and one, after which all possible minima were located by determining those points where the jerk (derivative of the acceleration) changed sign from negative to positive.

During a second step, each minimum in the acceleration was inspected and accepted or rejected on the basis of two thresholds. A first threshold was applied to eliminate those parts where little (below a certain signal to noise ratio) movement occurred. A second threshold was applied to disregard segments due to the so-called 'shoulders' or 'spikes' in the movement data. These thresholds were applied to the difference between the minimum acceleration value under inspection and the maximum acceleration value occurring before and after the minimum value under inspection.

With each threshold value, there is a corresponding number of segments found in the time series for each skeleton part considered, which results in a distribution for each skeleton part. By varying both threshold values, the variance of these distributions changes. We found that when the thresholds were low only very small segments were selected and when the thresholds were high only very large segments were selected. By taking the maximum variance of the bivariate distributions as a selection criterion, we obtained the most acceptable threshold values for each skeleton part of this specific data set.

In short, in order to obtain applicable segments, a filter method was applied based on the rejection of segments, which had a low range of acceleration (elimination of segments with no pronounced acceleration maximum). In addition, small perturbances in a larger segment, which were observed as "shoulders" or double peaks in the series, were removed. By changing both threshold values in a single function an optimum of the variance was then used to select the optimal parameters for the segmentation process (Figure 68).
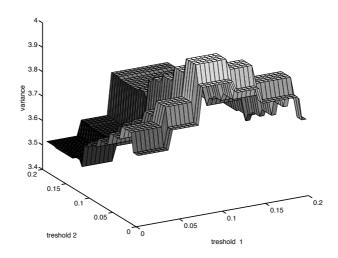
Figure 68. Threshold optimization of the segmentation method by changing the two threshold values and calculation of the variance in the number of extracted segments.

9.3.3 Dimension reduction using Principal Components Analysis (PCA)

In order to define segments in accordance with the top down annotation of gestures, using the classification by Wanderley et al. (2005) and the additional gestures added by the player, a descriptor based on the x, y, z displacement and derivatives is used.

The goal of using Principal Components Analysis (PCA) is to reduce the amount of data by looking whether markers that capture the movements of body parts are correlated. If they are correlated, then it is possible to combine these markers so that in the end a smaller number of data can be handled.

As a first step in this analysis, the magnitude of the acceleration of each joint is tested for normality (Kolmogorov-Smirnov (*KS*) test). As it is often the case in neuro-muscular signals, normality cannot be accepted due to the skewing to the right of the histograms (which is in this case the result of using the

magnitude of the acceleration). A square root transformation resolved this problem. Although some outliers were still present in some of the skeleton positions after transformation, normality for all data sets could be accepted (*KS, p* > 0.05).

As a second step in this analysis, PCA was applied to the extracted acceleration data. Due to the fact that the data sets are quite large Promax rotation was chosen as this is an alternative non-orthogonal (oblique) rotation method, which is computationally faster than the direct oblimin method and therefore is more useful for calculating large data sets (Abdi, 2003). The conditions needed for sampling (Kaiser-Meier-Olkin sampling adequacy index for comparing the magnitudes of the observed correlation coefficients with the magnitudes of the partial correlation coefficients and which should be greater than 0.5) and Bartlett's test for sphericity (to test whether strength of the relationship among variables is strong enough to use the method) were met. Two components were extracted which explained 74.3 % of the variance of the data (first component explains 44.7 % and second component 32.6 %). The two components are also highly correlated and can be related to either the part of the body (upper part, torso and lower part) or the acceleration range (high, mid and low).
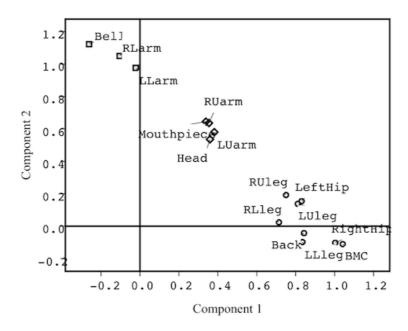
Figure 69. PCA component plot of the first two components of acceleration data.

As shown in Figure 69, the data set reveals three distinct groups. The movements of the bell and lower arms form the first group, which is obvious as the hands are connected to the instrument. The mouthpiece, the head and upper arms form the second group. The remaining skeleton positions, that is, back and hips, body mass center and legs form the third group.

Although PCA reveals that there are three clusters of joints, we decided that for further analysis, it is more convenient to consider four groups. We particularly make a further distinction within group three between the torso and the legs (see Table 15). By doing so, the results can be more easily compared with the manual annotations by the player for the movement of the legs (stepping and knee bending).

Group	PCA grouping	Calculated groups
bell, lower arms	PCA 1	clarinet
mouthpiece, head, upper arms	PCA 2	upper body
back, hips and body mass center	PCA 3	torso
upper and lower legs		legs

Table 15. Selection of groups based on principal components.

9.3.4 Combining marker segments

For each group of markers found in the PCA the segments (as defined in section 3.2) of the corresponding skeleton parts were combined, using a method where Gaussian distributions are defined for each start and end time of a segment. The use of Gaussian distributions around the start and end times was needed in order to be able to combine only those segments which showed enough overlap. The overlap was quantified by means of a t-test. The mean values of the Gaussians are defined by the time of start and end time of a segment. In order to define a value for the standard deviation a Mean Reaction Time of 200 ms was chosen as an estimate (Jensen, 2006). This results in a time series for each skeleton part containing Gaussian distributions located on the detected onsets with a width of 200 ms. The resulting separate skeleton time series of the group are combined using a weight based on the acceleration range and a curve fit of the best Gauss curve is then performed. The final result is a single time series for each group of markers containing the start and end points of segments found from the movement data as normal distributions. An example for group 1 (Clarinet Bell, Left Lower Arm and Right Lower Arm) is shown in Figure 70.
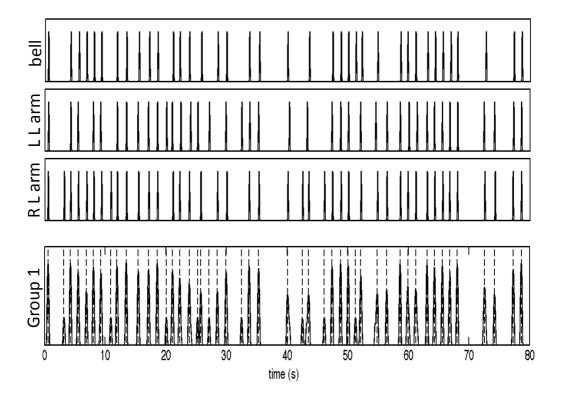
Figure 70. Example of Gaussian combination of group 1 (clarinet bell, left lower arm and right lower arm). The three figures on top represents the Gaussian fits of the different segment onsets of each marker of the group. The figure below shows the resulting onsets for the group as the resulting combined normal distributions.

9.3.5 Classification

It is tempting to see whether the extracted segments can be further grouped into different segment classes, as a successful grouping would facilitate the interpretation of the gestures. This analysis proceeds in two steps, namely, feature extraction and similarity comparison.

A set of properties (e.g. duration, curvature, arc length) of each extracted segment for each group is calculated (see Table 16). First, the properties were calculated for the separate markers after which the properties were combined corresponding the groups as defined in the previous section. A weighting based on the range of acceleration of the separate joints was applied in order to take into account the difference in contribution of the joints within a group. Second,
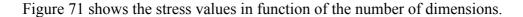
standard classification methods were applied. However, it was found that standard classification methods such as hierarchical clustering failed. It was found that no explicit partition into clusters occurred. Therefore, we adopted the approach proposed in Naveda (2011), namely to use Procrustes Analysis in combination with Multi Dimensional Scaling. The thus obtained properties can be considered as a set of landmarks of a shape (Goodall, 1991) enabling Ordinary Procrustes analysis to extract a similarity matrix.

Property	Definition
duration	length of the gesture (s)
arc length	parametric spline approximation to fit the curves, followed by numerical integration. (m)
curvature	based on Frenet-Serret space curve invariants (median of the resulting curvature vector)
median velocity	
median acceleration	median of kinematic parameters
median jerk	
begin to end distance	Euclidean distance between start and end point of the gesture
relative fraction X displacement ⎤	
relative fraction Y displacement ⎮ ⇒	fraction in x, y and z direction after mean removal
relative fraction Z displacement ⎦	

Table 16. Overview of selected segment properties.

This matrix was then further analyzed using non-metric Multi Dimensional Scaling with a stress minimization based on Kruskal's normalization, which is based on the sum of squares of the inter-point distances. In general, the goal of the analysis is to detect meaningful underlying dimensions in order to explain observed similarities or dissimilarities (distances) between the investigated objects. Multi Dimensional Scaling enables to analyze any kind of similarity or dissimilarity matrix, in addition to correlation matrices. The Multi Dimensional Scaling method was used in order to project the gestures from the

subjective approach onto the multi dimensional segmentation of the objective approach.

Figure 71 shows the stress values in function of the number of dimensions.
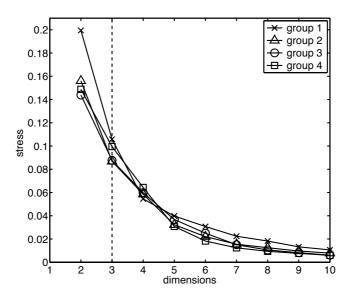
Figure 71. Scree plot of MDS stress over dimensions.

However no distinct 'elbow' effect occurs in the scree plot, although the plot seems to suggest that group 1 (clarinet bell and lower arms) has a higher stress value than the other groups, indicating more complexity (sources of variance) in the properties of this group. In this analysis, the MDS is based on a representation of the similarity data in three dimensions. The stress values for the groups in these three dimensions are between 0.08 and 0.11, which is considered to be a poor to fair fit. Therefore, we conclude that the classification approach did not enable a straightforward grouping of the segments. It seems that in the current state of our research, we are reaching the limit of the bottom-up approach. To better understand how segments are related to intentions, it is therefore necessary to rely on a comparison of the objective bottom-up approach with the more subjective top-down approach.

9.4 Understanding the relation between segments and gestures using a subjective top-down approach

In our analysis framework, we believe that an objective bottom-up approach to segment extraction can be further clarified with the help of a top-down approach. This method comprises both a qualitative as well as a quantitative approach. In the qualitative approach, we aim at an intuitive understanding of certain specific relationships between the objective bottom-up and subjective top-down features of the musical gestures. In the quantitative approach, we aim at an explicit measurement of particular relationships between the objective bottom-up and subjective top-down features of the musical gestures.

The player himself annotated his performance. Annotations are based on (i) a descriptive analysis of musical structural interpretations and intentions, (ii) a frame-by-frame video analysis of gestures, using the categorization of clarinetists' expressive movements as defined by Wanderley et al. (2005) and (iii) a visualization of annotated gestures on the score.

9.4.1 Descriptive score analysis

A first annotation by the performer consisted of a basic interpretative analysis of the score. The goal of this annotation was to gather data and information from the performer's point of view. We should be aware of the fact that, aiming at studying a performer's intuitive way of playing, he was not informed initially about the experimental character of his performance. Scores were not studied in depth and no interpretative annotations were added to the score prior to performance.

After the experiment was finished and after the performer was informed about the goals of the experiment, he was asked to indicate target notes in the score based on his recall of the performance. These are the notes towards which phrasing is directed and very often constitute the expressive culmination point of

a phrase (see Figure 72). Because these notes are linked to the formal structure of the piece, the performer also indicated phrases in the score. Furthermore other elements such as principal notes (notes that support phrase; see Figure 73), and recurring motives (upbeats, closing motives; see Figure 74) were annotated for later analysis of the objective movement data.

Figure 72. Example of target note (T).

Figure 73. Example of passing notes (P) and target note (T).

Figure 74. Examples of motives: (a) upbeat, (b) closing.

9.4.2 Frame-by-frame analysis

A second annotation by the performer dealt with the video recording of the performance. The goal of these annotations was to provide a reference frame for the objective analysis based on the mocap data and to make a comparison between a qualitative and quantitative approach in the analysis of performers' gestures possible.

To avoid being biased by the music, annotations were executed without sound. Not all classifications were annotated due to the frontal perspective of the observation. Therefore annotations were limited to the "clarinet bell" (circle),

"head up-down", "arms flapping", "bending knees" and "feet stepping". The player also added "feet together", because this posture occurred very often, and "breathing" (mouth open and closed around the mouthpiece).

9.4.3   Gesture annotation in the score

To visualize when in the music certain gestures occurred, gesture annotations based on non-audio video analysis were added onto the score (see Figure 75). This enabled us to see, in a simple and comprehensive way, whether some gestures coincide with the elements (target notes, principal notes, recurring motives) that were annotated in the score based on the performer's descriptive analysis. Furthermore, it enabled us to compare the objective bottom-up approach and the subjective top-down approach as adopted in this study.

First, phrases and recurring motives were added to the gesture annotations in ELAN, based on the sound recording of the performance. Next, three classes of gestures, i.e. "arms flapping", "bending knees" and "feet together", were added onto the score based on the combination of sound file and video annotation. The choice for these gestures was based on their salient occurrence during the performance.
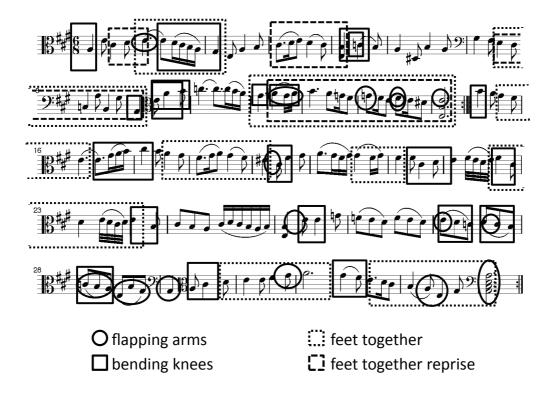
Figure 75. Gesture annotations of feet together, arm flapping and knee bending added to the score.

9.5 Comparison bottom-up and top-down

Up to this point, we have described the objective bottom-up and the subjective top-down approach. In this section, we provide a link between the two methods. This link will be provided on a qualitative and quantitative basis.

9.5.1 Score analysis

The goal of the score analysis is to identify the intended musical targets in connection with musical segments (also called motifs, figures) that support this intended target. The player, who marked the target points and supporting melodic sequences with a pencil on the score, as shown in Figure 76, carried out the annotation.

Figure 76. Annotated score by the player showing target notes, principle notes and recurring motives.

The goal of the subsequent comparative analysis was then to find evidence for the hypothesis that the player's local intended targets are related to body movements that express particular expressive sensitivities. However, in this approach, we do not aim to interpret the semantics of the intended target, as we believe that a proper verbal translation of the expressed sensitivity is not always possible. Thus, the annotations form part of a structural analysis, which is then related to particular types of corporeal articulations.

It is rather intuitive to look at the bell gestures as we believe these are the most important to support the theory. The bell has the largest variability of displacement and delivers a relatively high degree of freedom in spatial movement.

The technical arrangement of our analysis framework makes a straightforward direct comparison between annotated targets (subjective top-down) and segmented gestures (objective bottom-up) possible. To carry out this

comparison, the annotated musical phrase is selected and the extracted gestural features are inspected and compared with the annotations.

Figure 77. Score annotation of the 1st phrase.

We found that in general the bell gestures (defined by the bottom-up segmentation procedure) corresponded remarkably well with the segment annotations in the score. An example of segment annotation is shown in Figure 77 and 78.

Figure 77 shows how the score has been annotated in terms of target points and structural motives or segments (e.g. upbeat, passing notes) that support the target points.
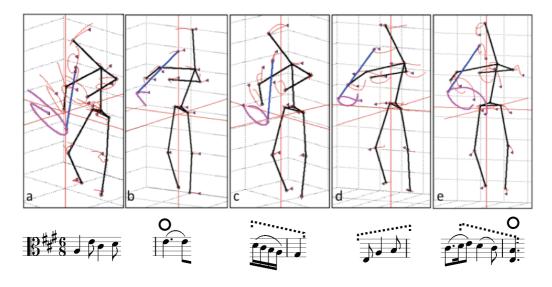
Figure 78. Example of score annotation and corresponding extracted segments of phrase 1.

Consider the first phrase (from 2.02 to 8.09 s). This part can be subdivided into five sections, as shown in Figure 78. The skeleton on top of this figure shows the bottom-up segmentation of the players' movement regarding the bell. The example shows that during the first part of this phrase, three

consecutive gestures (a, b and c) are found that correspond with three segments in the score. The skeleton figures show that the movements of the bell are somehow symmetric around the first target note (b).

Thus, the sequence starts with a clockwise expanding nearly planar spiral gesture of the bell, which corresponds with the first bar of the score (a). The gesture starts before playing the first note, which starts at the maximum of the acceleration in the segment. This gesture is followed by a displacement of the bell from left to right and down, which corresponds with the target note indicated by the player (b). Finally, a third gesture is again circular, which corresponds with the sequence of the 16th notes (c). The player also indicates the latter as motives.

The motive of 16th notes is then followed by a new motive, which was defined as an up-beat (d) by the player. This corresponds with a small circular movement of the bell and is immediately followed by the last part of the phrase (e), which ends on the target note and corresponds with a large circular movement of the bell. The 2 motives correspond with the small and the large circular movement.

The end of this sequence of gestures corresponds with the beginning of the next sequence of gestures as specified by the player. This indicates that gestures overlap. As a consequence of this the boundaries of the segments from the bottom-up analysis are not clear and distinct onsets of gestures. In some cases, this overlap can be understood as a co-articulation in the sense that the end of a gesture already contains a preparation for the next gesture (Zatorre et al., 2007).

In total, we considered seven phrases and found that in most cases the target notes as indicated by the player were indeed specific gestures extracted specifically in the movement. However, it was also found that the two successive movements frequently overlapped. Furthermore, it was found that the different

motives play an important role in the movements of the player. The motives can be explained as a 'preparation' of the gesture corresponding with the target note. It was also found that when inspecting the movements of the feet that the player followed the dancing movements of the gigue but ended with the feet closed on the target notes.

9.5.2 Video analysis

The goal of the video analysis was to enable the player to annotate different gestural tiers by combining video, skeleton movie without audio. An example is shown in Figure 79.
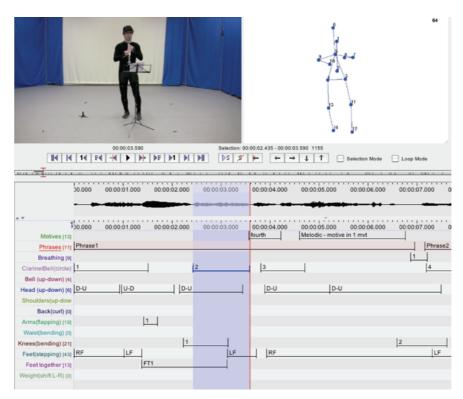
Figure 79. Example of video annotations of the player.

The annotations of the player show that only the most explicit gestures were detected. This can be explained by the Mean Reaction Time (MRT), which is about 160 ms to detect auditory stimuli, and approximately 190 ms to detect

visual stimuli at the age of 20. MRT values increase with age (Brebner & Welford, 1980). Annotating complex body movements is also difficult due to the fact that the observer has to try to stay focused on one skeleton position and make an abstraction of the other parts of the body. The process is also labor intensive and the detection of the limits (onset and offset) gesture is in some way subjective. Another problem arises from the fact that the video is a 2 dimensional image (fixed frontal view) so the observer has to decode this 2D image into a 3D image. Finally an observer looks at displacement in the Euclidean space; it is nearly impossible to 'look' at velocity, acceleration or jerk.

9.5.2.1 *Qualitative analysis*

The goal of this part of the analysis was to combine the kinematic time series, the annotations of the player and the extracted gestures to enable a qualitative comparison of the top-down and bottom-up approach.

When comparing the example in Figure 80 with the segments extracted from the data (clarinet bell circle 9, knee bending 9, and the movements of the feet (LF, RF) it was found that all the gestures could be identified as segments extracted from the data. Moreover, additional information could be added to the annotations of the player especially in the movements of the feet, which were found to have a more complex pattern due to separate left or right feet steps in different directions and the closing of the feet. It should be noticed the onset detection for the movement of the feet could not be determined exactly. The reason for this lies in the fact that during the experiment no markers were positioned directly on the feet. The markers on the lower legs are used as an approximation of the movement of the feet resulting in some deviations.
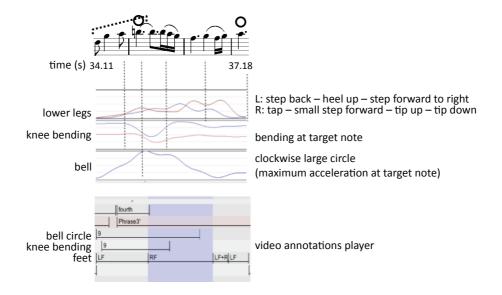
Figure 80. Combining player annotations with score (first part of phrase 3).

In a second example (phrase 6 of the score), the advantage of combining the bottom-up approach and the top-down approach is illustrated. The acceleration of Group 1 reveals a sequence of distinct segments in the instrument related movements. The beginning of the phrase is unclear due to an overlap with the ending of phrase 5. This is followed by 5 "swinging" movements that correspond with 5 consecutive groups of three 8th notes in the score. The last sequence starts with a large circle corresponding with the last set of three 8th notes and ends at the target note with a small circular movement of the instrument which is immediately followed by a larger circular movement at the beginning of phrase 7 (the coda of the piece, see Figure 81).
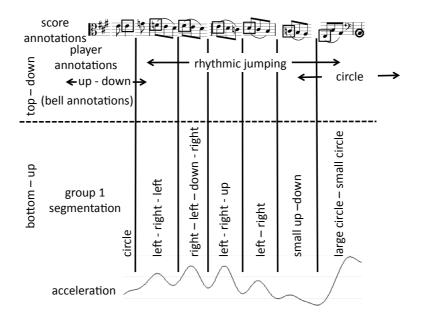
Figure 81. Example of qualitative comparison of top-down and bottom-up annotations of phrase 6. Squares in the score indicate principle notes, the circled note is the target note.

The whole sequence before the target note was annotated by the player as "rhythmic jumping" due to the nature of the piece (gigue). The target note is annotated as a circular bell movement and starts before the actual note. The player indicated also that the circular movement of the bell at the target note continues directly in the beginning of the next phrase (Figure 82).
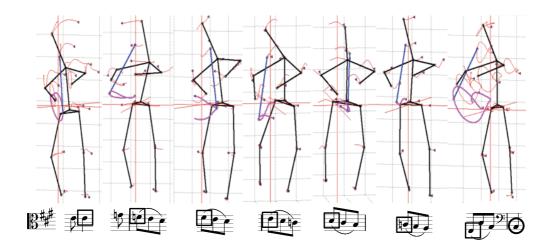
Figure 82. Corresponding segment gestures of phrase 6.

9.5.2.2 Quantitative analysis

A method was developed to quantify the degree of correspondence between the player annotations and the extracted gestures from the movement data. From the start and end points found in the annotations, a Gaussian distribution is constructed where the mean of the distribution is the mean of the start and end points and the width is defined by the 95% confidence interval defined by the length of the gesture. The same procedure is applied to the gesture segments found in the movement data. The resulting sets of Gaussian distributions are tested for similarity by performing a t-test. Based on this test, a gesture segment found in the movement data is accepted or rejected in the comparison with those found through annotation (Figure 83).

Figure 83. Example of comparison between the extracted segments (Group 1) and the corresponding video annotations of the player (clarinet bell and up-down).

The t-test ($\alpha = .05$) to test for equality of means between the annotations of the player and the closest segment extracted from the data was performed using random sampling. Type I and type II errors of the t-test are used to test for precision and recall of the method. Random sampling size was based on a power of 0.8 (error II). Table 17 shows the results for group 1.

player annotation	significance ($\alpha = 0.05$)	Corresponding extracted segment
circular	.0037	1
circular	.0000	3
circular	.0006	4
circular	.0399	7
up	.0357	10
down	.0091	10
up	.0022	11
*circular**	*.6378*	*17*
circular	.0314	21
circular	.0396	22
circular	.0129	23
circular	.0000	24
circular	.0131	25
up-down	.0218	32
circular	.0157	34
*circular**	*.6457*	*35*
*circular**	*.2575*	*36*
circular	.0027	38
circular	.0145	44
*circular**	*.4991*	*44*

Table 17. Example of quantitative comparison between player annotations and extracted segments. * Annotations that were not detected as segments.

For the annotations of the clarinet bell (circular and up and down movements) the player defined twenty gestures (sixteen were indicated as circular, four as up or down movements). Figure 83 shows a not detected annotation (a), partial but significant detections (b) and a perfect match between segmentation and annotation (c). Sixteen gestures (clarinet bell and up and down) annotated by the player correspond to a segment extracted from the data. This

provides strong evidence for the stated hypothesis of this work, namely that the intentionality of the player results in gestures, which can be extracted as segments from movement capture data.

9.6 General results

In this paper, we aimed at a better understanding of the nature of body movements in a musical performance. This understanding is inspired by two approaches, one called objective bottom-up approach, and another called top-down approach, the latter being more subjective. Main focus of the research was on the development of an analysis framework that supports linking between both approaches. In what follows, firstly, an overview is given of the analysis components of the framework, and secondly the important findings using the framework are reported. The analysis framework presented in this paper consists of the following analysis components:

- *Data pre-processing:* starting from raw mocap data, an algorithm was developed that leads to a normalization and bottom-up segmentation of the continuous movement data flow. The bottom-up segmentation of time series of acceleration was validated with a top-down analysis that is based on the performers intended targets.
- *Feature extraction:* starting from the segmented gestures, an algorithm was developed that extracts the features discussed in section 3.5, such as duration, arc length, curvature, velocity, acceleration, and jerk.
- *Objective data analysis* by means of two trajectories, namely, a trajectory that aims at gesture classification through similarity matching, and a trajectory that aims at gesture identification and understanding through comparison with a subjective top-down approach.
- *Statistical tools:* the use of PCA to reduce data, the use of a statistical method to optimize smoothing of the kinematic data, the use of feature

extraction, similarity matching, and visualization techniques (Procrustes, MDS) to better comprehend the data.

- The use of *a multimodal browser* ELAN that integrates the following data: 3D motion capture, video, audio, musical score, plus tiers that describe these data in terms of segmented units. Segments identified in this browser are further analyzed with Matlab tools that have been developed for this purpose.

- A method that makes the use of subjective top-down information to be used as a basis for data comprehension possible. This method has both a qualitative and a quantitative component to it. The qualitative component aims at an intuitive understanding of how gestures relate to musical targets, the quantitative component aims at an exact comparison between segmented gestures, both objective and subjective.

In bringing these different domains together, we developed our analysis framework that can be applied to other recordings of musical performances, especially to better understand the relationship between body movement, intended targets, and moving sonic forms. We are well aware of the fact that this should be proven with other data sets.

While using the above-mentioned framework we found that:

- The clarinetist's movements can be segmented into gestures, or movement segments that can be associated with particular gestures types, such as the circular movement of the clarinet bell, that are related to intentional (musical target related) functionalities. It was found that an objective bottom-up approach based on statistical criteria might provide a useful first step towards the segmentation of these movements.

- A classification of segmented gestures, using an objective bottom-up similarity-matching approach (based on Procrustes analysis) was not successful, and therefore, the analysis was complemented by a subjective

top-down approach that provided more information about intended targets.

- Expressive movements are related to musical targets. While some gestures provide access to the target, other gestures highlight the target, and other gestures guide the target. The gestures thereby function as corporeal articulations that express a musical discourse.

- The musical discourse can be understood as the encoding of moving sonic forms, using gestures that accompany the encoding of intended musical targets, the preparation for targets, and the post-processing of musical targets.

9.7 General discussion

9.7.1 An analysis framework for locally intended targets

Up to now, the most frequently applied methods for analyzing performers' body movement have been the use of video recording and movement tracker systems e.g. Wanderley et al. (2005).

In this study it was shown that a musician's body movements provide us with an interesting way to study the goal directed characteristics of a performance.

Motions capture, kinematic analysis and gesture extraction (segmentation) reveal differences in body movements in relation to different identified musical intentions and targets.

In general, the different types of annotations mentioned by the player and the primitives of gestures extracted from the movement analysis correspond with each other. Moreover, it was found that the presented segment extraction method enables us to make a more detailed description of the annotations as it was found that annotated gestures are a often a combination of a sequence of segments. Not

only is it possible to refine the typology of the gestures but also a detailed investigation of the profile of the gesture by means of the quantified gesture properties becomes possible as well.

9.7.2 Studying spontaneous performance gestures rather than induced performance gestures

Most studies of human gesture examine predefined and limited repetitive motor tasks (Alon, Athistos, Yuan & Sclaroff, 2009; Huang & Fu, 2009), with a setup of simple geometrical gestures (Field, Gordon, Peterson, Robinson, Stahovich & Alvarado, 2009) and/or concentrate on the active part of the body involved such as movement of the hand (Elmezain, Al-Hamadi, Appenrodt & Michaelis, 2009; Liu & Kavakli, 2010). Often, these studies rely on a theoretical model of a predefined "ideal" outcome of the gesture, which can be matched to the experimental data. Typical examples are hand-gesture recognition of simple structures such as a square, a circle or a triangle (Kratz & Rohs, 2010), the recognition of gestures when using a cell phone (He, Jin, Zen & Huang, 2009), and limited tasks in sports and health in for example gait analysis (Celis, Pipinos, Scott-Pandorf, Meyers, Stergiou & Johanning, 2009).

In contrast with these studies, our study considers the whole body in a research domain where no predictable parametric gesture functions or limitations linked to the task are available. The player was free to move (according to his personal intentions) when playing the music. Moreover, the musical piece was not based on clear repetitive measures or a clear beat. This is in line with other studies on musical expression in which the movement of the entire body is studied from the viewpoint of gestural expression (Cadoz & Wanderley, 2000; Camurri, De Poli, Leman & Volpe, 2001). However, most of these studies are induced performance studies, while our study focuses on locally intended targets. Therefore, a method was needed to define a detailed data pre-processing technique to overcome problems of missing values, outliers, smoothing,

differentiation, peak/valley detection and threshold parameters avoiding intuitive addition of parameters. Moreover, this method led to a framework of tools that can be used for future research.

9.7.3 Gesture comparison and classification

In an attempt to better understand the relationship between gestures and music, several authors have proposed classifications of movement types. Some approaches are theoretical, while other more practical. (Jensenius, Wanderley, Godøy & Leman, 2010) have proposed a theoretical framework in which five gesture categories are distinguished that build further on the framework proposed in (McNeill, 2000; Zhao, Costa, & Badler, 2000). However, that framework is very general and needs to be refined further. Wanderley and Vines (2006) have proposed categories that are used in the present study (Section 4.2) but that have been further complemented by other categories such as "feet together" and "breathing". The main difference between our research and that of Wanderley et al. (2005) is that we start our research from the raw motion capture data and from the general idea that a performers gesture is determined by the way a musician moves towards an intended target in the musical score, rather than towards a general intended expression of the entire piece. Intended targets are the points in the musical score that indicate performers focus of attention.

However, our attempt to classify clarinet gestures using a bottom-up approach based on similarity matching and classification (supervised assignment of gesture classes) was not very successful. In contrast, Naveda (2011) used the same method to compare 30 samba dances, with promising results. In his approach, repetitive movements characterized the dances, and each repetitive movement can be reduced to a basis gesture. All gestures of all dancers have the same duration and a closed form. The Procustes analysis calculates the cost of matching these forms, using rotation and other transformation processes. Compared with these dance gestures, our clarinet gestures differ in duration, and

they do not have a closed form. Therefore, rather than working with shapes, it was necessary to work with extracted features, which did not lead to a good similarity matching and classification result.

Apart from the fact that the clarinet gestures are different in duration and have an open form, it may well be the case that the adopted bottom-up approach reaches the limits of what is possible. Indeed, what we are aiming at is a categorization that is linked with intentionality, which is a subjective property that may be hard to trace when looking only at movements. Although there is evidence of the fact that aspects of intentionality are reflected in bodily articulations e.g. there is no direct bottom-up property in body articulation that can predict musical targets. For that aim, it was necessary to combine the bottom-up approach with a top-down approach. This approach implied a shift from prediction to understanding. Nevertheless, it is likely that with the help of a thorough analysis of the musical audio, new cues can be obtained, which, in combination with movement data, might make prediction possible. Whether this can be done on the basis of extracted features or whether more sophisticated simulations modeling is needed, remains a question to be examined in further research. Humans are good in understanding the intentions of other humans and theories of embodied cognition currently assume that this understanding, which comes down to anticipation and prediction, is based on a simulation model that is built up with the help of perceptive cues (Leman, Desmet, Styns, van Noorden & Moelants, 2007; Wilson & Knoblich, 2005). At this moment, however, it is necessary to rely on a comparison between the objective bottom-up approach and the subjective top-down approach. We believe that this approach is currently needed in our understanding of the communication model that underlies music and therefore, that a proper framework for statistical analysis that combines objective bottom-up with subjective top-down is useful.

9.7.4 Gesture segmentation

Various methods can be found in the literature to detect gestures from neuromuscular signals (Alon et al., 2009; Bayazit, Couture-Beil & Mori, 2009; Celis et al., 2009; Winter, 2009). We used a method based on the occurrence of maxima and minima in the norm acceleration and its first derivative (jerk). It was found that most methods found in literature can detect peaks and valleys in time series as long as there is a fluid sequence (sine like function) of the series or that there is a repetitive pattern in the gesture (such as stepping or playing the same short musical sequence several times) and autocorrelation or FFT techniques can be applied. In most cases segmentation is well defined due to either limited movements of objects such as the movement of an object such as a robot, or predefined human gestures associated with a clear meaning (pointing to a direction of a stimulus on a screen). In other cases, theoretical objects such as predefined figures (square, triangle, circle) or the recognition of handwriting (Kratz & Rohs, 2010; Liu & Kavakli, 2010; Zhang, Chen, Wang, Yang, Lantz & Wang, 2009) available. In present study, there is no obvious predictable model to define gestures as the player's movements were based on personal interpretation of the score and intentionality while performing. Furthermore, it was the aim to look at the whole body movement rather than focusing on a specific skeleton position. Therefore, several factors can contribute to the variance of gestures. First of all, there are factors, which are associated with the player such as musical skills, background, knowledge of the musical piece, etc. Secondly, the piece itself contains information based on the score such as musical period, indications by the composer, tempo, etc. Thirdly, the begin- and endpoint of a musical expressive gesture are not a simple sequence of independent gestures. The onset and offset of a gesture show an overlap between the previous and next gesture due to dynamic reparameterization of the player's brain activity in order to cope with anticipation, delay and intentionality. A previous study showed that anticipation and delay are important factors in musical performance analysis

(Desmet et al., 2011). Finally, there is also an effect due to the other gestures of the body (e.g. moving the clarinet bell while stepping forward or in combination with knee bending). Most methods fail when there are interruptions or leading or trailing tails in peaks and also when there are small spikes in the signal or when the signal is more complex. Combining the values of the maxima in the acceleration magnitude and the position where the sign of the first derivative of the norm acceleration was changing solved this problem. In this way every peak and valley position in the acceleration data was detected. It is obvious that not all the extracted local minima and maxima can be translated into a significant or meaningful gesture. Hence an appropriate filter was needed to eliminate unimportant minima within certain conditions.

9.7.5 Locally intended targets as reference frames for musical motility

Every musical piece has its own general expressive style expressed through the musician's interpretation. The interpretation of the music is captured in the so-called "big picture" (Chaffin, Imreh, Lemieux & Chen, 2003) characterized by a series of musical intentions or expressive sensitivities that are related to more local levels in the generative structure of the music. It is at these local levels that musicians will differ in the way they interpret the music and construct and communicate musical ideas. Therefore, studying the local expressive intentions of the musician is necessary to see how musicians deal with musical intentionality on the individual level.

The construction of musical intentions leads to a segmentation of the score into a series of units. These units or 'chunk's' are characterized by the use of certain musical elements as specific points where decisions are made related to intended expressive sensitivities. Some of these elements will serve as expressive culmination points (target notes) towards which the construction of a segmented unit is directed; other elements will support this directedness towards target points (e.g. principal notes). Knowing that musical intentions are reflected in the

movements of the musician (Davidson, 1993, Leman, 2007), it is important to study how movements are related to these more local levels of musical intentionality.

The way the music is segmented into musically meaningful units and the choice of important musical elements within this unit is determined by the global expressive style. For example, when the piece is to be played "gay" and it is a dance, it is likely that the music will be segmented differently (e.g. emphasizing smaller chunks by giving small accents on principal notes) then when the music is more "melancholic" (e.g. more gradual building up of the phrase without emphasizing smaller chunks). Therefore, we could say that the global expressive style leads to an intentional grid or matrix that is projected onto the generative structure of the music. This matrix consists of specific elements in the score and a series of performance cues that trigger the executive strategies connected to the expressive sensitivities or musical intentions at the more local level.

The intentional matrix and the accompanying executive strategies determine the musician's bodily behavior and therefore it is reasonable to assume that the movements of the musician reflect aspects of his musical intentions. According to this viewpoint, intentions may lead to specific goal-points, i.e. goal postures at certain points in time (Godøy & Leman, 2010) Whereas target points in the score would then coincide with *key frames* in the musician's bodily behavior, principal notes and recurring motives for example might influences *interframes,* i.e. movements between key frames (Rosenbaum, Cohen, Jax, Weiss & van der Wel, 2007). Therefore the intentional matrix and the accompanying executive strategies might be conceived of as determinants of the musician's spatiotemporal and bodily frame of reference that is in synchrony with musical cues and constitutes the musicians basic gestures (Leman & Naveda, 2010).

9.7.6 Gesture-related co-articulations

The global interpretation of the musical piece and its relation to the musician's gestures can be approached from the viewpoint of co-articulation. Generally speaking, co-articulation is the process in which smaller unites are meaningfully assembled into a higher-order level. In music, it is related to the levels of the generative structures. Notes are assembled into motives, motives into phrases, until the top-level of the generative structure, namely the piece as a whole, is reached. Musical expertise involves the communication of the higher levels within this structure.

If music and gestures correlate, then we can assume that way the basic gestures of the musician transition into one another is related to the way the music is structured according to the musician's musical intentions. With regard to the musician's bodily behavior, co-articulation can be defined as a "fusion of micro-gestures into more super-ordinate gestures, where there is a textual smearing of individual sounds and gestures so that they are no longer perceived as individual events". Music performance entails the generation of complex movements by combining basic gestures. But these combinations do not involve a mere serial concatenation of basic gestures. Co-articulation involves anticipatory modifications (Engel, Flanders & Soechting, 1997; Godøy & Leman, 2010). That is, past and future events influence present events. In music performance, this means that playing is always influenced by what is already played and what has yet to be played. Investigating the relationship between a musician's intentions and bodily behavior therefore needs to consider both segmentation and co-articulation.

9.8 Conclusion and future research

In this study, we aimed at understanding the relationship between a clarinetist's expressive intentions and body movements. To achieve this goal a new analytical

model has been developed that uses a strategy that combines a top-down and a bottom-up approach. The results of analyzing a clarinetist's expressive movements from those two perspectives are innovative because possible relationships between expressive musical intentions and physical movements in the performance could be identified. Mutual relations between a clarinet player's locally intended musical targets, his expression as a corporeal nuance, and its trace as a sonic movement were found. Considering the limited scope of the study, which is based on the data of one single performer, these results provide preliminary information. Further investigation aims at consolidating these findings by applying our novel analytical model to more performers, instruments and musical pieces.

An interesting area for possible exploration and further development could be to look at impact of performer's familiarity with the music. A comparative analysis between different clarinet players could also be an interesting field of further research. Another area for further research could be to develop a typology of gestures (Henbing & Leman, 2007) for the clarinet movements to use in machine learning techniques e.g. Hidden Markov modeling or Kalman filtering. Finally, it would be interesting to look at the movements and gestures of different players in order to try to find relations between the players' background, skills, age, personality, etc. Another area of interest is the development of tools for musical teaching based on the gestural behavior of students.

10 Discussion

In this thesis, we have developed a statistical framework that aims at supporting empirical research about embodied music cognition. The framework has been developed on the basis of several case studies, which culminated in a study about expressiveness and intentionality in clarinet performance. In that study, we show that the statistical framework can handle different data sets coming from different sources (e.g. motion capture, manual annotations). It has been shown that a detailed description of the data sets in combination with exploratory data analysis enables the researcher to prepare the data for further analysis. Not only this enables the researcher to define a logical structure of the data, making it possible to store them in a relational database but it also gives us the means to cope with errors, outliers and missing values. By focusing on the sources of variability, it is then possible to derive statistical paths to investigate hypothesis. In what follows, the motivation of the need to combine the top-down and bottom-up approach is first discussed followed by the need to establish a structured approach of data analysis in the field of systematic musicology. The next paragraph deals with description and prediction and the state of the art in this field up to now. This is followed by comments on intentionality and expression and how these key concepts have been investigated in this work. Finally, the limitations of this work are discussed.

10.1 Closing the semantic gap

The main contribution of the framework is that it closes the semantic gap between objective measurements of musical gestures and high-level aspects of musicality. Apparently, in this stage of the research, musicality cannot yet be straightforwardly extracted from the objective data that are obtained from movement recordings. A combined approach is needed in order to be able to

handle the data in relation to high-level musical concepts, such as expressiveness, and intentionality.

We are fully aware of the fact that some researchers may believe that these aspects of musicality should in fact be handled by a straightforward objective bottom-up approach, rather than a combined objective-subjective approach. However, we believe that the state-of-the-art of the research does not yet allow such a full-blown objective bottom-up approach. A good example is the clustering of the segmented clarinet gestures. Procustes and multidimensional scaling did not lead to any interpretable results so far, yet in combination with the annotations, it was possible to get a deeper insight into the nature of the gestures. The limitation of the objective bottom-up approach has to do with the fact that it is very difficult to extract intentions from signals, although we, humans, are good at it. At this moment, the semantic gap between objective measurement and subjective experience is too large. Therefore, we conceived the statistical framework as a tool that would help us to close the semantic gap, using a combined objective bottom-up and subjective top-down approach. We believe that this is a fundamental step towards future advanced analyses of data sets related to studies in embodied music cognition. In this context, it should be mentioned that the term 'subjective' is somewhat misleading. The subjective approach is based on the experience and skills of the performer. Knowledge about musical score analysis, the instrument and the relation with intentionality and expression are thereby essential and expert performers with a solid background in musical analysis are necessary. Therefore, the difference between the so called subjective top down method and the objective bottom up approach can be seen as the difference between two languages: the former is expressed in words, sentences and descriptions while the latter is expressed in numbers. In other words, it is mainly a difference in the way experience and knowledge is communicated. Throughout this work, it became gradually clear that our

understanding of human movements in relation to music relies on somehow merging both approaches.

10.2 A structured approach to data description and analysis

We believe that the framework offers a structured approach to the description and analysis of complex data sets in a way that has not been used very often in music research. This structured approach, starts from a model of variability, and it comprises a didactic component for the description of data sets and statistical paths that is useful as a way to structure data handling in an interdisciplinary research context, so that it can be stored in a database (e.g. SQL). Thus, the statistical framework offers a structure to researchers coming from different fields, such as musicology, engineering, movement sciences, medicine. This structure is complementary to structures that have been previously proposed but that are limited to structured data files to establish a standard for data interchange. For example, Jensenius et al. (2006) presents a gesture description interchange format (GDIF) for storing, retrieving and sharing information about music-related gestures. The main idea was that it should be possible to store all sorts of data from various input devices, such as controllers and motion capture systems, in a coherent and consistent data format, so that it can be used for interchange. We believe that our structured approach to data and data bases, based on the description schemes and the statistical paths that reveal the inherent logic of the data, is perfectly compatible with the idea of a gesture description interchange format. In fact, it would be useful to import or export data from GDIF format into an SQL database that is structured according to the proposed description and analysis schemes.

10.3 Description and prediction

In the past, music research was a descriptive science, which means that it focused on observations of things, and the description of how things happen and evolve.

In the last decades, music research has evolved into a predictive science, which means that observations can be related to their effects. Leman (2010) mentions that recent developments in music research indicate that it is a proactive science, which means that it has become a science that creates a situation by causing something to happen, rather than responding to it after it has happened. This becomes possible, for example, when a model or application is developed.

In view of the distinction between descriptive, predictive, and proactive music research, the present statistical framework can be considered a descriptive and predictive work. The state-of-the-art does not allow the development of a complete model, and in many cases are far from being able to predict outcomes on the basis of observed behavior. For example, we are not yet able to predict the target point in music solely on the basis of the observation of the player's movements. As humans, we may be able to do that, but we apparently use a model, which allows us to simulate the movement of the player, so that predictions can be made. Based on previous knowledge, identification of goal points is then rather straightforward. In fact, it resembles the identification of the goal point when somebody grasps an object. In short, our framework does not yet allow this kind of pro-action and prediction. Instead, what we believe to have achieved, is a framework that offers a structured approach for a descriptive analysis, forming the basis for concept clarification (e.g. concepts of body movement coordination, concepts of sharing expression, local target points in music and so on). At other points, we have been able to move into the direction of prediction (e.g. segmentation of gestures), but this research is limited. Further research is needed to elaborate on this.

10.4 Expressiveness and intentionality

Expressiveness and intentionality have been identified as core concepts of musicality. Our goal was to develop a method that made the empirical study of these concepts possible. By assuming that music is encoded expression of

intentions, which can be decoded by means of embodiment, which result in bodily expression (movement), understanding human movement is vital in order to propose a framework that contributes to a better understanding of music and its relation to humans. First, a method was proposed to find evidence for the capacity of humans to decode music into corporeal articulations, which are related to the expressive articulations of the performer.

In 'Sharing musical expression through embodied listening - a case study based on the movement responses of listeners to Chinese guqin music' (chapter 6), it was shown that by means of the correlation between time series, an analytical method was presented to support the idea that listeners can embody patterns of musical expressions.

In the second paper ('Statistical analysis of human body movement and group interactions in response to music', chapter 7) a method was proposed to quantify the social interactions between listeners when moving along to the beat of music in an ecological context. In this study, it was shown not only that movements on music depend on the type of music they hear but that there is also a social factor that plays a role.

The third paper focuses on the definition of appropriate statistical strategies to process movement data related to the gestures of a performer ('Gesture as coordinated action and intentionality in playing Guqin Music', chapter 8). It was found that the movements of the performer could be classified into clusters. This study showed also that there is a relation between the musical piece and the movements of the performer.

In order to clarify the viewpoint of the human body as mediator between physical environment and subjective experience when playing music, a framework was finally proposed. In 'A statistical analysis framework for assessing a clarinet player's performer gestures in relation to locally intended musical targets' (chapter 9), the methodology is explained into detail.

10.5 Limitations

The framework has a number of limitations of which we are fully aware.

A major limitation is caused by the fact that audio analysis is not included into the data processing. However, we believe that the framework allows a straightforward incorporation of audio analysis approaches. The next step in the development of this framework is certainly related to this issue, as we believe that a combined gesture-audio analysis may bring us a step closer to the prediction of musical intentions.

A second limitation is that, until now, only kinematic properties of movements have been investigated. We are convinced that also kinetics, energy and complexity are important topics. Some basic exploration of new methods such as directional statistics, phase analysis and measures for complexity has already been done, but it is too early to draw conclusions from the results obtained in this field up till now.

A third topic is that up to now the movements of the different body parts have been approached as single points rather than considering the whole body as a chain of rigid bodies instead of a set of isolated markers.

Fourth, up to now, it is too early to draw general conclusions as the framework focused on one performer and one musical piece. It is evident that the next step will be a confirmatory analysis based on other data sets and to expand and refine the proposed framework. A data set of full body movements, audio and video of nine participants in a large-scale experiment is available for analysis. In addition by means of questionnaires, information on personal background and the judgment of the performance are made available.

11 Conclusion

First of all, our research goal was to develop an empirical framework that allows the scientific study of the relationship between the subjective and the objective domains in which music manifests itself. Secondly, this framework is needed in order to achieve a better understanding of musical gestures as carriers of musical expressiveness and intentionality, which manifest themselves in corporeal articulations and responses of performers and listeners. The embodiment of high-level features of music into movement is a unique form of communication: what is being communicated is the quality of the experience.

11.1 The empirical framework

At the start of the project, empirical methods in musicology, and statistical tools in particular, were either used to process verbal reports of musical experiences, or from the viewpoint of audio analysis. Until the publication of Leman (2007), very little attention was paid to gesture analysis, particularly from an empirical viewpoint.

Data-analysis methods have been developed based on advanced multivariate statistics. This included the development of measurement devices, as well as tools for features extraction, correlation analysis and classification. An empirical methodology has been developed in view of an ecological setting. Recording and analysis have been based on lab-recordings as well as on recordings during public activities. In our research, we have acquired know-how about:

- the measurement of body movement using different camera systems (video and infra red cameras) and sensors (accelerometers). The performances of a guqin player and a clarinet player have been recorded

that way. In addition, listeners have been measured using the so-called embodied listening paradigm.

- extraction of features from the measurement, based on gesture segmentation, smoothing, data reduction techniques has been applied to the raw data, which results in descriptions that can be used for further processing.

- use of analysis of variance, correlation-analysis, and classification techniques for a large amount of different data (combining body movement features, audio features, bio-sensors, and verbal assessments).

We have made good progress in our scientific methodology. In particular, we have been able to acquire core know-how in (i) how the bottom-up objective analysis of musical gestures can be linked to a top-down subjective analysis of musical experiences, and (ii) how this approach can deal with a large range of data found in different measurement systems. The latter is particularly challenging and far from evident, given the large variability in human movement and its corporeal articulations to music.

That means that at this moment, we have achieved the core of a rather unique empirical framework for music analysis. This framework incorporates the most advanced sensing devices within a statistical framework that can handle a broad set of different data into our musicological analysis. A good deal of this methodology has been automated in the sense that processing routines are available in Matlab. Statistical paths have been developed that can be applied to future related studies. New toolboxes and ways of handling multi-modal information have been introduced, tested, and elaborated upon.

11.2 Understanding expressiveness and intentionality

As a result, with this empirical framework, we are now able to tackle very fundamental questions related to musical meaning formation. These questions are related to the foundations of the embodied music cognition paradigm that has

been introduced by Leman (2007). The questions have been handled in several papers along with the development of the empirical framework. The questions are related to the analysis of musical expressiveness, musical intentionality, and empathic listening.

While in the past, these questions have often been handled from a phenomenological perspective; my work has contributed to empirical perspective. The main difference is that the latter provides us with an explanatory model that can lead to a proactive model. The explanatory model can link several behaviors with mechanisms of action-perception coupling, thereby opening new connections to engineering and brain science. The results described in the papers show that the action-perception coupling forms the basis of the mechanisms that account for musical expressiveness and intentionality. Within the empirical framework, we now have tools to examine these issues in a scientific way, so that obtained results may provide possibilities for valorization in applications. Compared with the previous stages of research in systematic musicology, this implies a move from predictive science towards proactive science, which means that it offers a way of dealing with musical phenomena such that the outcome can have a real impact on the future developments, rather than merely describing what has happened before.

Although it is justified to say that on an international level, our work on empirical approach to music research is much appreciated, we still need to do more research in order to further consolidate the obtained results. It is very clear that the obtained results form the core of our future methodology in the field of empirical musicology. We strongly believe that our empirical framework forms an essential component for future proactive musicological research that will be able to support the development of the cultural and creative sector.

References

ABDI, H. (2003). Factor rotations in factor analyses. In Lewis-Beck M., Bryman, A., Futing T. (Eds.), *Encyclopedia for Research Methods for the Social Sciences* (pp. 792-795). Thousand Oaks (CA): Sage.

ALON, J., ATHITSOS, V., YUAN, Q., & SCLAROFF, S. (2009). A unified framework for gesture recognition and spatiotemporal gesture segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *31*,1685-1699.

ARISTOTLE, & MCKEON, R. P. (2001). *The basic works of Aristotle*. New York: Modern Library.

BAADER, A. P., KAZENNIKOV, O., & WIESENDANGER, M. (2005). Coordination of bowing and fingering in violin playing. *Cognitive Brain Research*, *23*, 436-443.

BAKER, A. (2010). Simplicity. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy (Spring 2010 Edition)*. Retrieved from: http://plato.stanford.edu/archives/spr2010/entries/simplicity/

BARGH, J. A., & CHARTRAND, T. L. (1999). The unbearable automaticity of being. *American Psychologist*, *54*, 462-479.

BAYAZIT, M., COUTURE-BEIL, A., & MORI, G. (2009). Real-time motion-based gesture recognition using the gpu. *Proceedings of the IAPR conference on Machine Vision Applications*, 9-12.

BECKING, G. (1928). *Der musikalische Rhythmus als Erkenntnisquelle* [Musical rhythm as source of knowledge]. Augsburg: B. Filser.

BERAN, J. (2004). *Statistics in musicology*. Boca Raton (Fla.): Chapman & Hall CRC Press.

BERNDT, D., & CLIFFORD, J. (1994). Using dynamic time warping to find patterns in time series. *AAAI94 workshop on knowledge discovery in databases (KDD-94)*, 359-370.

BERNHARDT, D. & ROBINSON, P. (2008). Interactive control of music using emotional body expressions. *CHI'08 extended abstracts on Human factors in computing systems*, 3117-3122.

BERTHOZ, A. (2000). *The brain's sense of movement*. Cambridge (MA): Harvard University Press.

BIRGÉ, L., & ROZENHOLC, Y. (2006). How many bins should be put in a regular histogram. *ESAIM: Probability and Statistics*, *10*, 24-45.

BOONE, R. T., & CUNNINGHAM, J. G. (2001). Children's expression of emotional meaning in music through expressive body movement. *Journal of nonverbal behavior*, *25*, 21-41.

BOSGA, J., MEULENBROEK, R. G. J., & SWINNEN, S. P. (2003). Stability of inter-joint coordination during circle drawing: Effects of shoulder-joint articular properties. *Human Movement Science*, *22*, 297-320.

BOWMAN, A. W., & AZZALINI, A. (1997). *Applied smoothing techniques for data analysis: The kernel approach with S-Plus illustrations*. New York: Oxford University Press, USA.

BOX, G. E. P., & COX, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society. Series B (Methodological)*, *26*, 211-252.

BREBNER, J. T., & WELFORD, A. T. (1980). Introduction: an historical background sketch. In *Reaction times* (pp. 1-23). New York (NY): Academic Press.

BROECKX, J. L. (1981). *Muziek, ratio en affect: Over de wisselwerking van rationeel denken en affectief beleven bij voortbrengst en ontvangst van muziek*. Antwerpen: Metropolis.

BROWN, M. B., & FORSYTHE, A. B. (1974). Robust tests for the equality of variances. *Journal of the American Statistical Association*, *69*, 364-367.

BROWN, S., MARTINEZ, M. J., & PARSONS, L. M. (2006). Music and language side by side in the brain: a PET study of the generation of melodies and sentences. *European Journal of Neuroscience*, *23*, 2791-2803.

BULDYREV, S., GOLDBERGER, A., HAVLIN, S., MANTEGNA, R., MATSA, M., PENG, C. K., SIMONS, M., & STANLEY, H. (1995). Long-range correlation properties of coding and noncoding DNA sequences: GenBank analysis. *Physical Review E*, *51*, 5084-5091.

CADOZ, C., & WANDERLEY, M. (2000). Gesture-music. In M.M. Wanderley & M. Battier (Eds.), *Trends in Gestural Control of Music* (pp. 71-93). Paris: IRCAM/Centre Pompidou.

CAMURRI, A., DE POLI, G., LEMAN, M., & VOLPE, G. (2001). A multi-layered conceptual framework for expressive gesture applications. *Proceedings International MOSART Workshop*, *Barcelona*. Citeseer.

CAMURRI, A., VOLPE, G., DE POLI, G., & LEMAN, M. (2005). Communicating expressiveness and affect in multimodal interactive systems. *IEEE Multimedia*, *12*, 43-53.

CANAZZA, S., DE POLI, G., DRIOLI, C., RODA, A., & VIDOLIN, A. (2004). Modeling and control of expressiveness in music performance. *Proceedings of the IEEE*, *92*, 686-701.

CANAZZA, S., DE POLI, G., RINALDIN, S., & VIDOLIN, A. (1997). Sonological analysis of clarinet expressivity. In M. Leman (Ed.), *Music, Gestalt, and Computing*, (pp. 431-440). Berlin, Heidelberg: Springer.

CARRUTHERS, P., & SMITH, P. K. (EDS.) (1996). *Theories of theories of mind*. Cambridge: Cambridge University Press.

CASTELLANO, G., MORTILLARO, M., CAMURRI, A., VOLPE, G., & SCHERER, K. (2008). Automated analysis of body movement in emotionally expressive piano performances. *Music Perception*, *26*, 103-119.

REFERENCES

CELIS, R., PIPINOS, I. I., SCOTT-PANDORF, M. M., MYERS, S. A., STERGIOU, N., & JOHANNING, J. M. (2009). Peripheral arterial disease affects kinematics during walking. *Journal of Vascular Surgery*, *49*, 127-132.

CHAFFIN, R., IMREH, G., LEMIEUX, A. F., & CHEN, C. (2003). "Seeing the Big Picture": Piano Practice as Expert Problem Solving. *Music Perception*, *20*, 465-490.

CHEN, J. L., PENHUNE, V. B., & ZATORRE, R. J. (2008). Listening to musical rhythms recruits motor regions of the brain. *Cerebral Cortex*, *18*, 2844-2854.

CHEN, J. L., ZATORRE, R. J., & PENHUNE, V. B. (2006). Interactions between auditory and dorsal premotor cortex during synchronization to musical rhythms. *Neuroimage*, *32*, 1771-1781.

CLARKE, E., & DAVIDSON, J. (1998). The body in performance. *Composition-Performance-Reception: Studies in the Creative Process in Music*, 74-92.

CLARKE, E. F., & COOK, N. (2004). *Empirical musicology: aims, methods, prospects*. New York: Oxford University Press, USA.

CLAYTON, M., SAGER, R., & WILL, U. (2005). In time with the music: The concept of entrainment and its significance for ethnomusicology. *European Meetings in ethnomusicology*, *11*, 1-42.

CLAYTON, M. R. L. (2007). Observing entrainment in music performance: Video-based observational analysis of Indian musicians' tanpura playing and beat marking. *Musicae Scientiae*, *11*, 27-29.

CLEVELAND, W. S., & DEVLIN, S. J. (1988). Locally weighted regression: an approach to regression analysis by local fitting. *Journal of the American Statistical Association*, *83*, 596-610.

CLYNES, M., & MENUHIN, Y. (1977). *Sentics: The touch of emotions* (1st ed.). Garden City (NY): Anchor Press.

CORDO, P., CARLTON, L., BEVAN, L., CARLTON, M., & KERR, G. (1994). Proprioceptive coordination of movement sequences: role of velocity and position information. *Journal of neurophysiology*, *71*, 1848-1861.

CUMMING, N. (2000). *The sonic self: Musical subjectivity and signification*. Bloomington (IN): Indiana University Press.

D'AUSILIO, A., ALTENMÜLLER, E., OLIVETTI BELARDINELLI, M., & LOTZE, M. (2006). Cross-modal plasticity of the motor cortex while listening to a rehearsed musical piece. *European Journal of Neuroscience*, *24*, 955-958.

D'AVELLA, A., FERNANDEZ, L., PORTONE, A., & LACQUANITI, F. (2008). Modulation of phasic and tonic muscle synergies with reaching direction and speed. *Journal of Neurophysiology*. *100*, 1433-1454.

DAFFERTSHOFER, A., LAMOTH, C. J. C., MEIJER, O. G., & BEEK, P. J. (2004). PCA in studying coordination and variability: a tutorial. *Clinical Biomechanics*, *19*, 415-428.

DAHL, S., & FRIBERG, A. (2007). Visual perception of expressiveness in musicians' body movements. *Music Perception*, *24*, 433-454.

DALIRI, M. R., & TORRE, V. (2009). Shape and texture clustering: Best estimate for the clusters number. *Image and Vision Computing*, *27*, 1603-1614.

DARWIN, C. R. (1872). *The expression of the emotions in man and animals*. London: John Murray.

DAVIDSON, J. W. (1993). Visual perception of performance manner in the movements of solo musicians. *Psychology of Music*, *21*, 103.

DAVIDSON, J. W. (2005). Bodily communication in musical performance. In D. Miell, R. Macdonald, & D. J. Hargreaves (Eds.), *Musical communication* (pp. 215-228). New York: Oxford University Press, USA.

DAVIDSON, J. W. (2007). Qualitative insights into the use of expressive body movement in solo piano performance: a case study approach. *Psychology of Music*, *35*, 381-401.

DE BRUYN, L., LEMAN, M., & MOELANTS, D. (2008). Quantifying children's embodiment of musical rhythm in individual and group settings. *Proceedings of the 10th International Conference on Music Perception and Cognition. Sapporo, Japan*, 662-665.

DE BRUYN, L., LEMAN, M., MOELANTS, D., & DEMEY, M. (2009). Does Social Interaction Activate Music Listeners? In R. Kronland-Martinet, S. Ystad, & K. Jensen (Eds.), *Computer music modeling and retrieval. Genesis of meaning in sound and music* (pp. 93-106) Berlin, Heidelberg: Springer-Verlag.

DE BRUYN, L., LEMAN, M., MOELANTS, D., DEMEY, M., DESMET, F. (2008). Measuring and Quantifying the Impact of Social Interaction on Listeners' Movement to Music. *Proceedings of the 2008 Computers in Music Modeling and Retrieval and Network for Cross-Disciplinary Studies of Music and Meaning Conference*, 298-305.

DE POLI, G. (2004). Methodologies for expressiveness modelling of and for music performance. *Journal of New Music Research*, *33*, 189-202.

DE POLI, G., MION, L., & RODA, A. (2009). Toward an action based metaphor for gestural interaction with musical contents. *Journal of New Music Research*, *38*, 295-307.

DEMEY, M., LEMAN, M., BOSSUYT, F., & VANFLETEREN, J. (2008). The Musical Synchrotron: using wireless motion sensors to study how social interaction affects synchronization with musical tempo. *Proceedings NIME 2008*, 372-373.

DESMET, F., LEMAN, M., & LESAFFRE, M. (2011). Gesture as coordinated action and intentionality in playing Guqin music. *Case study, IPEM Ghent University*.

DESMET, F., DEMEY, M., LEMAN, M., LESAFFRE, M., & DE BRUYN, L. (2010). Statistical analysis of human body movement and group interactions in response to music. In H-H Bock et al. (Eds.), *Advances in Data Analysis, Data Handling and Business Intelligence* (pp. 411-424). Berlin, Heidelberg: Springer-Verlag, 399-408.

DIXON, S. (2005). An on-line time warping algorithm for tracking musical performance. *Proceedings of the International Conference on Artificial Intelligence*, 1727-1728.

ELMEZAIN, M., AL-HAMADI, A., APPENRODT, J., & MICHAELIS, B. (2009). A Hidden Markov Model-based continuous gesture recognition system for hand motion trajectory. *19th International Conference on Pattern Recognition, 2008, ICPR 2008*. 1-4.

ENGEL, K. C., FLANDERS, M., & SOECHTING, J. F. (1997). Anticipatory and sequential motor control in piano playing. *Experimental brain research*, *113*, 189-199.

FELDMAN, A. G. & LEVIN, M. F. (2009). The equilibrium-point hypothesis--past, present and future. *Progress in Motor Control*, 699-726.

FIELD, M., GORDON, S., PETERSON, E., ROBINSON, R., STAHOVICH, T., & ALVARADO, C. (2009). The effect of task on classification accuracy: Using gesture recognition techniques in free-sketch recognition. *EUROGRAPHICS Symposium on Sketch-Based Interfaces and Modeling*, 499-512.

FOOTE, J. (2000). Automatic audio segmentation using a measure of audio novelty. *International Conference on Multimedia and Expo, 2000. ICME 2000*, *1*, 452-455.

FREEDMAN, D., & DIACONIS, P. (1981). On the histogram as a density estimator: L 2 theory. *Probability Theory and Related Fields*, *57*, 453-476.

FREEMAN, W. J. (2000). A neurobiological role of music in social bonding. In N. L. Wallin, B. Merker, & S. Brown (Eds.), *The origins of music* (pp. 411-424). Cambridge (MA): MIT Press.

FRIBERG, A., & SUNDBERG, J. (1999). Does music performance allude to locomotion? A model of final ritardandi derived from measurements of stopping runners. *The Journal of the Acoustical Society of America*, *105*, 1469-1484.

FRIBERG, A., SUNDBERG, J., & FRYDÉN, L. (2000). Music from motion: Sound level envelopes of tones expressing human locomotion. *Journal of New Music Research*, *29*, 199-210.

FRISTON, K. (2010). The free-energy principle: a unified brain theory? *Nature Reviews Neuroscience*, *11*, 127-138.

FUCHS, A., JIRSA, V. K., & KELSO, J. A. S. (2000). Theory of the relation between human brain activity (MEG) and hand movements. *Neuroimage*, *11*, 359-369.

GABRIELSSON, A. (2003). Music performance research at the millennium. *Psychology of Music*, *31*, 221-272.

GALLAGHER, S. (2005). *How the body shapes the mind*. Oxford: Clarendon Press.

GALLESE, V. (2003). The manifold nature of interpersonal relations: the quest for a common mechanism. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *358*, 517-528.

GALLESE, V., & GOLDMAN, A. (1998). Mirror neurons and the simulation theory of mind-reading. *Trends in cognitive sciences*, *2*, 493-501.

GODØY, R. I., & LEMAN, M. (EDS.) (2010). *Musical Gestures: Sound, Movement, and Meaning*. New York: Routledge.

GOLDSTEIN, H. (2010). *Multilevel statistical models*. Chichester: John Wiley & Sons.

REFERENCES

GOODALL, C. (1991). Procrustes methods in the statistical analysis of shape. *Journal of the Royal Statistical Society. Series B (Methodological)*, *53*, 285-339.

GREGORY, A. H. (1997). The roles of music in society: The ethnomusicological perspective. In D. J. Hargreaves & A. C. North (Eds.), *The social psychology of music* (pp. 123-140). Oxford: Oxford University Press.

GRICE, J. W., & ASSAD, K. K. (2009). Generalized Procrustes Analysis: A Tool for Exploring Aggregates and Persons. *Applied Multivariate Research*, *13*, 93-112.

GROSS, J., TIMMERMANN, L., KUJALA, J., DIRKS, M., SCHMITZ, F., SALMELIN, R., & SCHNITZLER, A. (2002). The neural basis of intermittent motor control in humans. *Proceedings of the National Academy of Sciences of the United States of America*, *99*, 2299-2302.

HANSLICK, E. (1891). *Vom Musikalisch Schönen ein Beitrag zur Revision der Ästhetik der Tonkunst* [The beautiful in music: A contribution to the revisal of musical aesthetics]. Leipzig: Barth.

HARGREAVES, D. J., & NORTH, A. C. (EDS.). (1997). *The social psychology of music*. Oxford: Oxford University Press.

HATTEN, R. (2003). Thematic gestures, topics, and tropes. In E. Tarasti (Ed.), *Musical semiotics revisited* (pp. 80-91). Helsinki: Hakapaino.

HAUEISEN, J., & KNÖSCHE, T. R. (2001). Involuntary motor activity in pianists evoked by music perception. *Journal of Cognitive Neuroscience*, *13*, 786-792.

HE, Z., JIN, L., ZHEN, L., & HUANG, J. (2009). Gesture recognition based on 3D accelerometer for cell phones interaction. *IEEE Asia Pacific Conference on Circuits and Systems APCAS-2008*, 217-220.

HENBING, L. & LEMAN, M. (2007). A gesture-based typology of sliding-tones in guqin music. *Journal of New Music Research*, *36*, 61-82.

HOMMEL, B., MÜSSELER, J., ASCHERSLEBEN, G., & PRINZ, W. (2002). The theory of event coding (TEC): A framework for perception and action planning. *Behavioral and Brain Sciences*, *24*, 849-878.

HONING, H. (2006). On the growing role of observation, formalization and experimental method in musicology. *Empirical Musicology Review*, *1*, 2-6.

HUANG, E.W., & FU, L. C. (2009). Gesture stroke recognition using computer vision and linear accelerometer. *8th IEEE Conference on Automatic Face & Gesture Recognition, 2008. FG'08*, 1-6.

HUME, D., SELBY-BIGGE, L. A., & NIDDITCH, P. H. (1975). *Enquiries concerning human understanding and concerning the principles of morals* (3d ed.). Oxford: Clarendon Press.

HUOVINEN, E. (2006). Varieties of musicological empiricism. *Empirical Musicology Review*, *1*, 12-27.

HYNDMAN, R. J. (1995). The problem with Sturges' rule for constructing histograms. *Monash University, Clayton Australia*. Retrieved from: http://robjhyndman.com/papers/sturges.ps

JEANNEROD, M. (1994). The representing brain: Neural correlates of motor intention and imagery. *Behavioral and Brain Sciences*, *17*, 187-202.

JENSEN, A. R. (2006). *Clocking the mind: Mental chronometry and individual differences*. Oxford: Elsevier Science.

JENSENIUS, R., KVIFTE, T., & GODØY, R. I. (2006). Towards a gesture description interchange format. *Proceedings of the 2006 reference on New interfaces for musical expression*. IRCAM-Centre Pompidou, 176-179.

JENSENIUS, R., WANDERLEY, M. M., GODØY, R. I., & LEMAN, M. (2010). Musical gestures: concepts and methods in research. In R. Godøy, & M. Leman (Eds.), *Music, Gesture, and the Formation of Embodied Meaning* (pp. 12-35). New York: Routledge.

JOHNSON, R., & WICHERN, D. (2002). *Applied multivariate statistical analysis*. New Jersey: Prentice-Hall.

JUSLIN, P. N., & LAUKKA, P. (2004). Expression, perception, and induction of musical emotions: A review and a questionnaire study of everyday listening. *Journal of New Music Research*, *33*, 217-238.

KANT, I. (1790/1924). *Kritik der Urteilskraft* [Critique of judgement] (6th ed.). Leipzig: F. Meiner.

KELLER, P. E. (2008). Joint action in music performance. In F. Morganti, A. Carassa, & G. Riva (Eds.), *Enacting intersubjectivity: A cognitive and social perspective on the study of interaction* (pp. 205-221). Amsterdam: IOS.

KELLER, P. E., KNOBLICH, G., & REPP, B. H. (2007). Pianists duet better when they play with themselves: On the possible role of action simulation in synchronization. *Consciousness and cognition*, *16*, 102-111.

KELSO, J. A. S., FUCHS, A., LANCASTER, R., HOLROYD, T., CHEYNE, D., & WEINBERG, H. (1998). Dynamic cortical activity in the human brain reveals motor equivalence. *Nature, 392*, 814-817.

KIRKE, A., MIRANDA, E. R., & ZHANG, Q. (2010). Artificial evolution of expressive performance of music: An imitative multi-agent systems approach. *Computer Music Journal*, *34*, 80-96.

KNOBLICH, G., & SEBANZ, N. (2006). The social nature of perception and action. *Current Directions in Psychological Science*, *15*, 99-104.

KRAEMER, H. C., & THIEMANN, S. (1987). *How many subjects?: Statistical power analysis in research*. Thousand Oaks (CA): Sage Publications.

KRATZ, S., & ROHS, M. (2010). A 3D gesture recognizer: simple gesture recognition for devices equipped with 3D acceleration sensors. *Proceedings on the 14th international conference on Intelligent user interfaces*, 341-344.

LACQUANITI, F., & SOECHTING, J. (1982). Coordination of arm and wrist motion during a reaching task. *The Journal of Neuroscience*, *2*, 399-408.

LAKIN, J. L., JEFFERIS, V. E., CHENG, C. M., & CHARTRAND, T. L. (2003). The chameleon effect as social glue: Evidence for the evolutionary significance of nonconscious mimicry. *Journal of Nonverbal Behavior*, *27*, 145-162.

LARGE, E. W. (2000). On synchronizing movements to music. *Human Movement Science*, *19*, 527-566.

LARTILLOT, O., & TOIVIAINEN, P. (2007). A matlab toolbox for musical feature extraction from audio. *Proceedings of the 10th International Conference on Digital Audio Effects, Bordeaux, France*, 1-8.

LATASH, M. L., SCHOLZ, J. P., & SCHÖNER, G. (2002). Motor control strategies revealed in the structure of motor variability. *Exercise and sport sciences reviews*, *30*, 26.

LEMAN, M. (2007). *Embodied music cognition and mediation technology*. Cambridge (MA): The MIT Press.

LEMAN, M. (2010). Some reflections on systematic musicology as proactive science. In R. Bader, C. Neuhaus, & U. Morgenstern (Eds.), *Concepts, Experiments, and Fieldwork: Studies in Systematic Musicology and Ethnomusicology* (pp.21-34). Frankfurt: Peter Lang.

LEMAN, M., & CAMURRI, A. (2005). Understanding muscial expressiveness using interactive multimedia platforms. *Musicae Scientiae, Special Issue 2005/06*, 209-233.

LEMAN, M., DESMET, F., STYNS, F., VAN NOORDEN, L., & MOELANTS, D. (2009). Sharing musical expression through embodied listening: a case study based on Chinese guqin music. *Music Perception*, *26*, 263-278.

LEMAN, M., LESAFFRE, M., & TANGHE, K. (2001). Introduction to the IPEM toolbox for perception-based music analysis. *The Online Contemporary Music Journal*, *7,* 25-50.

LEMAN, M., & NAVEDA, L. (2010). Basic gestures as spatiotemporal reference frames for repetitive dance/music patterns in Samba and Charleston. *Music Perception, 28*, 71-91.

LESAFFRE, M. (2005). *Music Information Retrieval. Conceptual framework, Annotation and User Behaviour*. PhD Thesis, Ghent University.

LESAFFRE, M., DE VOOGDT, L., LEMAN, M., DE BAETS, B., DE MEYER, H., & MARTENS, J. P. (2008). How potential users of music search and retrieval systems describe the semantic quality of music. *Journal of the American Society for Information Science and Technology*, *59*, 695-707.

LEVENE, H. (1960). Robust Tests for equality of variances. In I. Olkin, S. G. Ghurye, W. Hoeffding, W. G. Madow, W.G., & H.B. Mann (Eds.) *Contributions to probability and statistics: Essays in honor of Harold Hotelling* (pp. 278-283). Stanford: Stanford University Press.

LI, L., VAN DEN BOGERT, E. C. H., CALDWELL, G. E., VAN EMMERIK, R. E. A., & HAMILL, J. (1999). Coordination patterns of walking and running at similar speed and stride frequency. *Human Movement Science*, *18*, 67-85.

REFERENCES

LILLIEFORS, H. W. (1969). On the Kolmogorov-Smirnov test for the exponential distribution with mean unknown. *Journal of the American Statistical Association*, *64*, 387-389.

LINDENBERGER, U., LI, S. C., GRUBER, W., & MÜLLER, V. (2009). Brains swinging in concert: cortical phase synchronization while playing guitar. *BMC Neuroscience*. Retrieved from: http://www.biomedcentral.com/1471-2202/10/22

LIPPS, T. (1903). *Ästhetik: Psychologie des schönen und der kunst* [Aesthetics: Psychology of beauty in art]. Hamburg und Leipzig: L. Voss.

LIU, J., & KAVAKLI, M. (2010). Hand Gesture Recognition Based on Segmented Singular Value Decomposition. *Knowledge-Based and Intelligent Information and Engineering Systems*, 214-223.

LOULA, F., PRASAD, S., HARBER, K., & SHIFFRAR, M. (2005). Recognizing people from their movement. *Journal of Experimental Psychology: Human Perception and Performance*, *31*, 210-220.

LUAN, J. (2002). Data mining and its applications in higher education. *New directions for institutional research*, *113*, 17-36.

MACHULDA, M., WARD, H., BOROWSKI, B., GUNTER, J., CHA, R., O'BRIEN, P., PETERSEN, R., BOEVE, B., KNOPMAN, D., TANG-WAI, D., & OTHERS (2003). Comparison of memory fMRI response among normal, MCI, and Alzheimer's patients. *Neurology*, *61*, 500-506.

MAES, L. (2009). Sound sculptures and installations as potential new instruments. *American Musical Instrument Society Thirty-Eighth Annual Meeting*, 14.

MAGESCAS, F., URQUIZAR, C., & PRABLANC, C. (2009). Two modes of error processing in reaching. *Experimental Brain Research*, *193*, 337-350.

MARTIN, A. J. (2008). Motivation and engagement in music and sport: Testing a multidimensional framework in diverse performance settings. *Journal of Personality*, *76*, 135-170.

MASSEY, F. J. (1951). The Kolmogorov-Smirnov test for goodness of fit. *Journal of the American Statistical Association*, *46*, 68-78.

MCCULLOUGH, B., & HEISER, D. A. (2008). On the accuracy of statistical procedures in Microsoft Excel 2007. *Computational Statistics & Data Analysis*, *52*, 4570-4578.

MCNEILL, D. (ED.) (2000). *Language and gesture*. Cambridge: Cambridge University Press.

MCNEILL, D. (ED.) (2005). *Gesture and thought*. Chicago: University of Chicago Press.

MION, L., & DE POLI, G. (2008). Score-independent audio features for description of music expression. *IEEE Transactions on Audio, Speech, and Language Processing*, *16*, 458-466.

REFERENCES

MION, L., POLI, G. D., & RAPANA, E. (2010). Perceptual organization of affective and sensorial expressive intentions in music performance. *ACM Transactions on Applied Perception (TAP)*, *7*, 14.

MOENS, B., VAN NOORDEN, L., & LEMAN, M. (2010). D-Jogger: Syncing music with walking. *Proceedings of SMC Conference 2010, Barcelona*, 451-456.

NASA-STD-3000 (1995). Anthropometric Source Book: Anthropometry for Designers Anthropology Staff/Webb Associates, NASA. 1. Retrieved from:
http://msis.jsc.nasa.gov/Volume1.htm

NATTIEZ, J. J., & DALE, C. (1985). The Concepts of Plot and Seriation Process in Music Analysis. *Music Analysis*, *4*, 107-118.

NAVEDA, L. (2011). *Gesture in Samba: A cross-modal analysis of dance and music from the Afro-Brazilian culture*. PhD Thesis, Ghent University.

NAVEDA, L., & LEMAN, M. (2010). The representation of spatiotemporal music gestures, using Topological Gesture Analysis (TGA). *Music Perception, 28,* 93-111.

NAYAK, S., WHEELER, B. L., SHIFLETT, S. C., & AGOSTINELLI, S. (2000). Effect of music therapy on mood and social interaction among individuals with acute traumatic brain injury and stroke. *Rehabilitation Psychology*, *45*, 274-283.

NIEDENTHAL, P. M., BARSALOU, L. W., WINKIELMAN, P., KRAUTH-GRUBER, S., & RIC, F. (2005). Embodiment in attitudes, social perception and emotion. *Personality and Social Psychology Review*, *9*, 184-211.

NETTHEIM, N., & BECKING, G. (1996). How musical rhythm reveals human attitudes: Gustav Becking's theory. *International Review of the Aesthetics and Sociology of Music*, *2*, 101-122.

NEWELL, K. M., & SLIFKIN, A. B. (1998). The nature of movement variability. In Piek J. P. (Ed.) *Motor behavior and human skill: A multidisciplinary approach*, (pp. 143-160). Champaign (IL) USA: Human Kinetics Publishers.

NIJS, L., LESAFFRE, M., & LEMAN, M. (2009). The musical instrument as a natural extension of the musician. *Proceedings of the 5th Conference of Interdisciplinary Musicology*, 132-142.

NUSSECK, M., & WANDERLEY, M. M. (2009). Music and motion - How music-related ancillary body movements contribute to the experience of music. *Music Perception*, *26*, 335-353.

OVERY, K., & MOLNAR-SZAKACS, I. (2009). Being together in time: Musical experience and the mirror neuron system. *Music Perception*, *26*, 489-504.

PALMER, C. (2005). Time course of retrieval and movement preparation in music performance. *Annals of the New York Academy of Sciences*, *1060*, 360-367.

PALMER, C., KOOPMANS, E., LOEHR, J. D., & CARTER, C. (2009). Movement-Related Feedback and Temporal Accuracy in Clarinet Performance. *Music Perception*, *26*, 439-449.

PAN, W., & TANG, M. (2004). Examining the Effectiveness of Innovative Instructional Methods on Reducing Statistics Anxiety for Graduate Students in the Social Sciences. *Journal of Instructional Psychology*, *31*, 149-160.

PARNCUTT, R. (2007). Systematic musicology and the history and future of western musical scholarship. *Journal of Interdisciplinary Music Studies*, *1*, 1-32.

PARNCUTT, R., & MCPHERSON, G. (2002). *The science & psychology of music performance: creative strategies for teaching and learning*. Oxford: Oxford University Press, USA.

PARSONS, T. (1986). *Voice and Speech Processing*. New York: McGraw-Hill, USA.

PENTTINEN, H., PAKARINEN, J., VÄLIMÄKI, V., LAURSON, M., LI, H., & LEMAN, M. (2006). Model-based sound synthesis of the guqin. *The Journal of the Acoustical Society of America*, *120*, 4052-4063.

PHILLIPS-SILVER, J., & TRAINOR, L. J. (2008). Vestibular influence on auditory metrical interpretation. *Brain and Cognition*, *67*, 94-102.

POLLICK, F. E., PATERSON, H. M., BRUDERLIN, A., & SANFORD, A. J. (2001). Perceiving affect from arm movement. *Cognition*, *82*, 51-61.

QUETELET, A. (1835). *Sur l'homme et le développement de ses facultés ou essai de physique sociale*. Paris: Bachelier.

REPP, B. H. (1993). Music as motion: A synopsis of Alexander Truslit's (1938) *Gestaltung und Bewegung in der Musik*. *Psychology of Music*, *21*, 48-72.

REPP, B. H., & KELLER, P. E. (2010). Self versus other in piano performance: detectability of timing perturbations depends on personal playing style. *Experimental brain research*, *202*, 101-110.

REPP, B. H., & KNOBLICH, G. (2004). Perceiving action identity – How pianists recognize their own performances. *Psychological Science*, *15*, 604-609.

ROSENBAUM, D. A., COHEN, R. G., JAX, S. A., WEISS, D. J., & VAN DER WEL, R. (2007). The problem of serial order in behavior: Lashley's legacy. *Human Movement Science*, *26*, 525-554.

ROSNER, B. (1983). Percentage points for a generalized ESD many-outlier procedure. *Technometrics*, *25*, 165-172.

SAKOE, H. & CHIBA, S. (1978). Dynamic programming algorithm optimization for spoken word recognition. *IEEE Transactions on Acoustics, Speech and Signal Processing*, *26*, 43-49.

SAPATINAS, T. (2005). Discriminant Analysis and Statistical Pattern Recognition. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *168*, 635-636.

SARKAR, A., & BARAT, P. (2006). Multiscale entropy analysis: A new method to detect determinism in a time series. *Arxiv preprint physics/0604040*.

SCHOGLER, B., PEPPING, G. J., & LEE, D. N. (2008). TauG-guidance of transients in expressive musical performance. *Experimental Brain Research*, *189*, 361-372.

REFERENCES

SHAN, G., & VISENTIN, P. (2003). A quantitative three-dimensional analysis of arm kinematics in violin performance. *Medical problems of performing artists*, *18*, 3-10.

SNOW, C. P. (1993). *The Two Cultures, the Rede Lecture 1959*. Cambridge: The Cambridge University Press.

SPARING, R., MEISTER, I., WIENEMANN, M., BUELTE, D., STAEDTGEN, M., & BOROOJERDI, B. (2007). Task-dependent modulation of functional connectivity between hand motor cortices and neuronal networks underlying language and music: a transcranial magnetic stimulation study in humans. *European Journal of Neuroscience*, *25*, 319-323.

STEENKAMP, J. B. E. B., TRIJP, M., & VERHALLEN, T. (1989). Procrustes analyse (1). *Tijdschrift voor marketing*, *23*, 18-26.

STERGIOU, N. (ED.) (2004). *Innovative analyses of human movement*. Champaign (IL): Human Kinetics Publishers.

STERGIOU, N., JENSEN, J. L., BATES, B. T., SCHOLTEN, S. D., & TZETZIS, G. (2001). A dynamical systems investigation of lower extremity coordination during running over obstacles. *Clinical Biomechanics*, *16*, 213-221.

STURGES, H. A. (1926). The choice of a class interval. *Journal of the American Statistical Association*, *21*, 65-66.

STYNS, F. (2008). *Kenmerken van de muzikale basisbeweging in stappen op muziek - Een empirisch musicologisch onderzoek naar de grondslagen van de relatie tussen muziek en beweging*, Unpublished PhD Thesis, Ghent Unversity.

STYNS, F., VAN NOORDEN, L., MOELANTS, D., & LEMAN, M. (2007). Walking on music. *Human movement science*, *26*, 769-785.

SUNDBERG, J. (ED.) (2003). *Research in music performance (*Special issue of *Journal of New Music Research*). Lisse: Swets & Zeitlinger.

TAYLOR, B. N. (1995). *Guide for the Use of the International System of Units (SI): The Metric System*. DIANE Publishing.

THAUT, M., MCINTOSH, G., RICE, R., MILLER, R., RATHBUN, J., & BRAULT, J. (1996). Rhythmic auditory stimulation in gait training for Parkinson's disease patients. *Movement Disorders*, *11*, 193-200.

THOMPSON, M., & LUCK, G. (2008). Exploring relationships between expressive and structural elements of music and pianists' gestures. *Proceedings of the fourth Conference on Interdisplinary Musicology (CIM08)*, 1-10.

TOIVIAINEN, P., & BURGER, B. (2010). *MoCap Toolbox Manual*. University of Jyväskylä. Retrieved from: http://www.jyu.fi/music/coe/materials/mocaptoolbox/MCTmanual

TOIVIAINEN, P., LUCK, G., & THOMPSON, M. (2009). Embodied metre: hierarchical eigenmodes in spontaneous movement to music. *Cognitive Processing*, *10*, 325-327.

TOIVIAINEN, P., & SNYDER, J. S. (2003). Tapping to Bach: Resonance-based modeling of pulse. *Music Perception*, *21*, 43-80.

TOMASI, G., VAN DEN BERG, F., & ANDERSSON, C. (2004). Correlation optimized warping and dynamic time warping as preprocessing methods for chromatographic data. *Journal of Chemometrics*, *18*, 231-241.

TRUSLIT, A. (1938). *Gestaltung und Bewegung in der Musik* [Figuration and motion in music]. Berlin-Lichterfelde: C. F. Vieweg.

TUKEY, J. W. (1951). Components in regression. *Biometrics*, *7*, 33-69.

TUKEY, J. W. (1977). *Exploratory data analysis*. Reading (MA): Addison-Wesley.

VAN BENDEGEM, J. P. (2009). *Hamlet en entropie: de twee culturen een halve eeuw later. Een pamflettair essay* [Hamlet and entropie: the two cultures a half century later. A pamphlet]. Asp/Vubpress/Upa.

WANDERLEY, M. M., & VINES, B. (2006). Origins and functions of clarinettist's Ancillary Gestures. In A. Gritten, & E. King (Eds.), *Music and gesture* (pp. 165-191). Aldershot: Ashgate Publishing.

WANDERLEY, M. M., VINES, B. W., MIDDLETON, N., MCKAY, C., & HATCH, W. (2005). The musical significance of clarinetists' ancillary gestures: An exploration of the field. *Journal of New Music Research*, *34*, 97-113.

WATSON, A. H. D. (2006). What can studying musicians tell us about motor control of the hand?. *Journal of Anatomy*, *208*, 527.

WEIBULL, W. J. (1951). A statistical distribution function of wide applicability, *Applied Mechanics*, *18*, 293-297.

WIDMER, G., & GOEBL, W. (2004). Computational models of expressive music performance: The state of the art. *Journal of New Music Research*, *33*, 203-216.

WILLIAMON, A. (2004). *Musical excellence: Strategies and techniques to enhance performance.* New York: Oxford University Press.

WILSON, M., & KNOBLICH, G. (2005). The case for motor involvement in perceiving conspecifics. *Psychological Bulletin*, *131*, 460-473.

WINTER, D. A. (ED.) (2009). *Biomechanics and motor control of human movement*. Hoboken (NJ): John Wiley & Sons.

WITTENBURG, P., BRUGMAN, H., RUSSEL, A., KLASSMANN, A., & SLOETJES, H. (2006). Elan: a professional framework for multimodality research. *Proceedings of LREC 2006*.

YANG, K., & SHAHABI, C. (2005). On the stationarity of multivariate time series for correlation-based data analysis. *Fifth IEEE International Conference on Data Mining*, 4.

ZATORRE, R. J., CHEN, J. L., & PENHUNE, V. B. (2007). When the brain plays music: auditory-motor interactions in music perception and production. *Nature Reviews Neuroscience*, *8*, 547-558.

ZHANG, X., CHEN, X., WANG, W., YANG, J., LANTZ, V., & WANG, K. (2009). Hand gesture recognition and virtual game control based on 3D accelerometer and EMG sensors. *Proceedings of the 13th international conference of Intelligent user interfaces*, 401-406.

ZHAO, L., COSTA, M., & BADLER, N. (2000). Interpreting moverment manner. *Proceedings Computer Animation Conference. Philadelphia, PA, 2000*, 112-120.