Methods and Technologies for the Analysis and Interactive Use of Body Movements in Instrumental Music Performance

Federico Visi

Interdisciplinary Centre for Computer Music Research (ICCMR)
Plymouth University

This dissertation is submitted for the degree of
Doctor of Philosophy

March 2017

I dedicate this work to my loving mother.
Grazie infinite mamma, ti voglio tanto bene.

Declaration

At no time during the registration for the degree of Doctor of Philosophy has the author been registered for any other University award without prior agreement of the Graduate Sub-Committee.

Work submitted for this research degree at the Plymouth University has not formed part of any other degree either at Plymouth University or at another establishment.

This study was financed with the aid of a studentship from the School of Humanities and Performing Arts, Plymouth University.

Relevant scientific seminars and conferences were regularly attended at which work was often presented; external institutions were visited for consultation purposes and several papers prepared for publication.

A list of publications, performances, presentations, and workshops carried out as part of this doctoral research programme can be found in appendix B.

Word count of main body of thesis: 47,351 words.

Federico Visi
March 2017

Acknowledgements

Throughout this long journey, I had the luck to meet a number of wonderful people, to whom I am very grateful.

My supervisory team:

My Director of Studies Prof Eduardo Miranda. Thank you for your guidance, support, and trust. Your work has inspired me and showed me how groundbreaking research and compelling music making can live in synergy to one another.

My second supervisor Dr Lee Miller. Thank you for allowing me to see my research from a different perspective and for introducing me to Ashtanga Yoga. That not only has helped me finding the energies to make it through long periods of intense work, it also gave me greater awareness of how my body works and moves.

I am particularly grateful to the people I had the pleasure to work with in the past three years:

Dr Rodrigo Schramm, for all the work done together and particularly for the ideas that led to the definition of Periodic Quantity of Motion. Thank you for your generosity, kindness, warmth, and for all the nights spent in the lab together working hard to meet the next submission deadline.

Dr Duncan Williams, for his input in the side project in collaboration with UCLA and especially for the countless words of advice he gave me throughout my doctoral studies. Thank you so much for your expertise, wisdom, patience, altruism, and good sense of humour.

Esther Coorevits, for a wonderful one and a half years together spent writing, travelling, presenting our research at conferences, making music, performing, and enjoying ourselves. Thank you for all the energy and patience you put in bringing my ideas to life. Some of the research and musical visions presented in this dissertation would have not been possible without your talent.

Giovanni Dothel, for his key contribution to the side project at UCLA and for making that collaboration possible. I learnt a lot during those days in Los Angeles and also had a great time. Thank you for your time, energy, hospitality, and for putting up with my jet lag.

Dr Alexis Kirke, for his advice when I wrote my first grant application. That was a success, also, because of your input: thank you.

Pierre-Emmanuel Largeron, for his valuable input to the violin study. Thank you for your time and great musical skills.

The members of the fourMs - Music, Mind, Motion, Machines research group at the University of Oslo, Norway. Thank you for your hospitality and knowledgeable support during the motion capture sessions for the violin study presented in this dissertation.

Dr Katherine Williams, for her great contributions to the development of *11 Degrees of Dependence* and for performing the piece with me in Plymouth and in Leeds. Thank you so much for your patience and valuable insight as a musicologist and performer.

Dr David Bessell, for the physical modelling algorithm used in *11 Degrees of Dependence* and for all the long chats about music, research, and analogue synthesisers. Thank you, I hope we will find the time to make some music together.

Ana García Caraballos, for performing *11 Degrees of Dependence* in New York. Thank you so much for your commitment and for your terrific performances.

Andrew Telichan Phillips, for making the collaboration with NYU possible. Thanks man, best of luck with your projects.

Michael Mcloughlin, for his help with real-time motion data streaming and his contribution to the design of the multimodal data collection system. Thank you for all the ideas.

Marco Frattini and Lara Jones, for performing *Music of Motion and Presence* with me. Thank you so much for your talent and musical vision.

Collaborating with these people has been crucial for the development of this dissertation and for my growth as a researcher. I have learnt a lot from all of them.

I would also like to thank all the other fellow researchers at the Interdisciplinary Centre for Computer Music Research (ICCMR): Joel Eaton (without whom I certainly would not have completed Plymouth's Half Marathon twice, and run in the mud on a Saturday morning), Ed Braund, David Strang, Ben Payne, Aurelien Antoine, Jared Drayton, Dr Marcelo Gimenes, Nuria Bonet Filella. You have been my family in Plymouth, thank you for your help on several occasions and for the nice meals together.

I am very grateful to the School of Humanities and Performing Arts at Plymouth University for giving me the fantastic opportunity of a studentship; to the Doctoral

Training Centre for supporting my conference travels; to Santander Universities for funding the projects in Los Angeles and New York; and to the Roland Levinsky Memorial Fund for some additional travel support.

Thanks to Simon Ible and Peninsula Arts for giving me the opportunity to develop *Music of Motion and Presence* and perform it in The House theatre.

Thanks to the other people working in The House at Plymouth University: Victor Ladron de Guevara, Adam Benjamin, Joanne 'Bob' Whalley, John Matthews, Roberta Mock, Ruth Way. It has been great to work in such a stimulating interdisciplinary environment.

Thanks to the Plymouth University staff members for their technical support, in particular Melaine Le Bars and Stuart Billinghurst at The House theatre, Arron Tyler and Ben Payne at the Music department, and Richard Wood at the Brunel workshops. Thank you for making everything work during my performances and allow me to turn my ideas into a reality.

Thanks to Luigi Pizzaleo for his advice when I started outlining this research project. You probably did not realise it at the time, but by forwarding me an email about a PhD research studentship you changed the course of my life.

Thanks to my housemates, in particular Luca for giving me shelter in London and Sian for her advice while I was proofreading my dissertation.

Thank you Rønnaug for your encouragement and for having been next to me while I was writing the research proposal that led to this doctoral thesis.

Thank you Luisa for your hospitality and for your warmth while I was working on the final revisions. I am very happy we met each other again.

Finally, very special, heartfelt, thanks to my family, in particular my mother – to whom this thesis is dedicated – and my father. Thank you for everything you have done for me.

Much love to you all,

Federico

Abstract

Methods and Technologies for the Analysis and Interactive Use of Body Movements in Instrumental Music Performance

Federico Visi

A constantly growing corpus of interdisciplinary studies support the idea that music is a complex multimodal medium that is experienced not only by means of sounds but also through body movement. From this perspective, musical instruments can be seen as technological objects coupled with a repertoire of performance gestures. This repertoire is part of an *ecological knowledge* shared by musicians and listeners alike. It is part of the engine that guides musical experience and has a considerable expressive potential.

This thesis explores technical and conceptual issues related to the analysis and creative use of music-related body movements in instrumental music performance. The complexity of this subject required an interdisciplinary approach, which includes the review of multiple theoretical accounts, quantitative and qualitative analysis of data collected in motion capture laboratories, the development and implementation of technologies for the interpretation and interactive use of motion data, and the creation of short musical pieces that actively employ the movement of the performers as an expressive musical feature.

The theoretical framework is informed by embodied and enactive accounts of music cognition as well as by systematic studies of music-related movement and expressive music performance.

The assumption that the movements of a musician are part of a shared knowledge is empirically explored through an experiment aimed at analysing the motion capture data of a violinist performing a selection of short musical excerpts. A group of subjects with no prior experience playing the violin is then asked to mime a performance following the audio excerpts recorded by the violinist. Motion data is recorded, analysed, and compared with the expert's data. This is done both quantitatively through data analysis

as well as qualitatively by relating the motion data to other high-level features and structures of the musical excerpts.

Solutions to issues regarding capturing and storing movement data and its use in real-time scenarios are proposed. For the interactive use of motion-sensing technologies in music performance, various wearable sensors have been employed, along with different approaches for mapping control data to sound synthesis and signal processing parameters. In particular, novel approaches for the extraction of meaningful features from raw sensor data and the use of machine learning techniques for mapping movement to live electronics are described.

To complete the framework, an essential element of this research project is the composition and performance of études that explore the creative use of body movement in instrumental music from a Practice-as-Research perspective. This works as a test bed for the proposed concepts and techniques. Mapping concepts and technologies are challenged in a scenario constrained by the use of musical instruments, and different mapping approaches are implemented and compared. In addition, techniques for notating movement in the score, and the impact of interactive motion sensor systems in instrumental music practice from the performer's perspective are discussed. Finally, the chapter concluding the part of the thesis dedicated to practical implementations describes a novel method for mapping movement data to sound synthesis. This technique is based on the analysis of multimodal motion data collected from multiple subjects and its design draws from the theoretical, analytical, and practical works described throughout the dissertation.

Overall, the parts and the diverse approaches that constitute this thesis work in synergy, contributing to the ongoing discourses on the study of musical gestures and the design of interactive music systems from multiple angles.

Table of contents

List of figures

List of tables

Chapter 1

Prelude

> [...] the instrument glows and flashes under the master's hands. [...] he must be heard—and also seen; for if Liszt played behind the screen, a great deal of poetry would be lost.
>
> ROBERT SCHUMANN
> *On Music and Musicians*

1.1 Introduction

Do the body movements of musicians playing an instrument affect our musical experience? If so, can the gestures and body movements of instrumentalists be used as an expressive feature in composition? Can this gestural knowledge aid the design of new musical interactions? What are the conceptual and technical challenges that need to be tackled? These are some of the questions that motivate this research project. As it can be easily inferred, to address these questions it is necessary to operate in an interdisciplinary scenario. As suggested by the approach adopted in systematic musicology, music is a multifaceted phenomenon, thus *transdisciplinarity* is required to address important aspects that often go beyond the boundaries of single disciplines (Leman, 2008b). Along with systematic musicology, this dissertation will draw from the disciplinary fields of interaction design, computer music, and areas of computer science such as digital signal processing and machine learning.

1.1.1 Research Context and Core Concepts

As it will be discussed more in detail in part I, the paradigm of embodied music cognition and the study of musical gestures put forward the idea that the perception of music is closely linked to bodily experience. Consequentially, gestures have a key role in the process of musical meaning formation, and the ecological knowledge of the gestural repertoire of a traditional musical instrument contributes to the formation of multimodal embodied musical meaning.

Extensive interdisciplinary research has recently been carried out (Godøy and Leman, 2010; Gritten and King, 2006, 2011) giving rise to new paradigms for the understanding of gesture in music. In particular, insights from research on embodied music cognition (Leman, 2008a) inspired new viewpoints that required a rethink of the foundations of musical gestures. Within this theoretic framework, music perception is *embodied* (i.e. closely linked with bodily experience) and *multimodal*, in the sense that music is perceived not only through sound but additionally with the help of both visual cues and sensations of motion, effort and dynamics (Godøy, 2010). Hence, gestures become a core notion as they act as a bridge between bodily movement and meaning formation. The definition of 'gesture' and other key concepts such as 'movement' will be discussed in chapter 3.

I adopted the term *traditional musical instrument* (henceforth TMI) to define instruments that have a rich idiomatic repertoire that ranges across popular music and classical music. Examples include electric and acoustic guitar, violin, piano, electronic keyboards, etc. I deliberately included electric and electronic instruments among these examples because the term TMI refers to the use and repertoire of the instrument, not to its design or technological aspects. To mention another example, in some music scenes (e.g. hip-hop), turntables have been widely used as instruments for creating and performing new music, and the movements that characterise turntable performance are easily recognised by people familiar with the music genres in which they are employed (Godøy and Leman, 2010). Therefore, in this context, turntables *are* TMIs, regardless of the fact that they were not originally designed to be musical instruments. The choice of focusing on TMI performance is motivated by the vast knowledge that listeners have of the gestural and sound aspects of each instrument, which is learnt through experience. For instance, most people immediately know what sound to expect when they see a drummer hitting a snare drum with a stick. Similarly, if we hear the sound of a violin we can easily associate the gesture of bowing to it. This can be summarised as the *ecological knowledge* (Godøy, 2010) of an instrument; listeners have, and in some cases share, *a repertoire of sound-producing gestures*. Different TMIs will be employed in the analysis and practice parts of this dissertation. This was done in order to develop concepts and techniques that are not limited to a specific instrument but can be utilised in studies and works involving various TMIs. For instance, Periodic Quantity of Motion (PQoM) is introduced in the study involving the violin presented in chapter 5 and is also adopted in chapter 6, which describes a mapping approach for motion sensors and electric guitar. Moreover, the concepts and software tools described in chapter 8 were used with wearable sensors and saxophone for the piece '11 Degrees of Dependence' (see section 8.4) as well as for the piece 'Tuned Constraint' (involving an analogue synthesiser, see section 8.5). In addition, the movement notation system adopted for the saxophone part of '11 Degrees of Dependence' can also be used with other instruments, as will be explained in section 8.4.3.

In fact, using embodied music cognition terminology, TMIs have a rich action/gesture repertoire that the listeners (or *perceivers*) can recognise during a performance. Thus, using the instrumentalist's gestures may have a considerable expressive potential in performance as well as in composition, as composers would be able to draw from a *gestural palette* of the instrument when writing a piece.

1.2 Research Aims

The main aims of the project are:

- To determine and discuss a multidisciplinary theoretical framework useful to understand in which ways the musical gestures and body movements of a person playing a traditional musical instrument can affect the musical experience of the perceivers and contribute to the construction of musical meaning. A survey and discussion of the relevant theories is presented in part I. This will also serve as the theoretical foundations for the analysis and practice works described in the other parts of the dissertation.

- To analyse the relationship between body movements and musical features in performances with selected musical instruments using motion-capture technologies. With the aid of the framework mentioned above, experiments aimed at analysing the relationship between movement and other musical features were designed and carried out. Both quantitative and qualitative methods were adopted for analysis. This is mainly addressed in part II.

- To compose brief pieces for traditional instruments and electronics that are used as case studies to explore the role of gestural aspects of instrument playing in the formation of musical meaning. Motion-sensing technologies were employed to control electronic music parameters and mapping strategies were informed by paradigms of embodied music cognition and multimodal meaning formation. These pieces, along with the techniques employed for their composition and performance, are described in part III.

- Develop tools and techniques useful for motion data analysis and music performance involving motion-sensing devices. This is addressed both in parts II and III.

Thus, these aims lead to the formulation of the following research questions:

1.2.1 Research Questions

To what extent the movements involved in instrumental music performance are part of a shared knowledge of musical gestures, and how do they relate to other musical features?

This question will be addressed with the empirical experiments described in part II. In particular, chapter 5 will explore how the instrumental gestures associated with a traditional musical instrument (the violin) are part of an embodied knowledge shared also by people with no previous experience playing the instrument.

If gestures and body movement play a key role in how we experience and understand music, how can they be employed as expressive elements in composition and performance?

This research question will be addressed mainly in part III. In particular, musical pieces purposively composed and performed to be used as case studies will be described in chapters 7 and 8. Additionally, the pieces make use of wearable motion-sensing technologies, which allow to include in the performance live electronic elements that respond to the movement of the musicians. Gestures can be mapped to digital sound processing parameters, used to alter the timbre and other sonic features of the instrument and to control other electronic sound sources. Thus, a third research question is raised:

Can a multimodal embodied approach to musical meaning formation that takes into consideration the ecological knowledge of a traditional musical instrument be used to inform effective mapping strategies?

Considering gesture as an active constituent of embodied musical meaning implies that its role in an interactive music performance goes well beyond being a mere means of control of musical parameters. New cross-disciplinary approaches may in fact help to give rise to new engaging musical experiences that can both raise questions and provide new insights about musical expression and cognition. As noted by Godøy (2010), Western musical thought has not been well equipped for thinking of the inclusion of musical elements within the context of a gesture. Therefore, further research into gestural and embodied aspects of music might bring about unprecedented insights and possibly lead to the emergence of new aesthetic categories.

Regarding issues of parameter mapping for interactive music performance, preliminary work informed by embodied cognition and functional aspects of musical gestures is described in chapter 6. A more sophisticated system based on multimodal motion data and involving custom software tools is instead presented in chapter 9.

1.3 Research Methods and Strategies

As previously mentioned, one of the aims of the project is to define an interdisciplinary theoretical framework that allows for the analysis of musical gestures in instrument performance from multiple standpoints. Along with this multi-disciplinary framework, a combination of different methodological approaches will be adopted in order to address different aspects of gestural musical meaning formation, including:

Quantitative analysis was carried out using data collected through multimodal recordings of music performances. The studies described in chapters 4 and 5 focus on the analysis of motion capture data, while chapter 9 reports on a novel method for employing multimodal datasets to define interaction models for motion-sound interaction.

Qualitative observations integrate data analysis by situating the quantitative measurements in a broader musical context. This allows to relate the results of the analysis to other musical features and take into consideration elements of style, interpretation, and articulations found in the score.

Practice-led research completes the methodological framework. Mutual influence between research and practice in the creative arts has been widely documented and is an established approach in academic research (Smith and Dean, 2009). Chapters 7 and 8 report on compositions used as case studies whereas chapters 6 and 9 are focused on design approaches and parameter mapping for motion-sound interaction.

This three-fold methodology is supported by the multidisciplinary theoretic framework discussed in part I, which is the result of a detailed review of literature concerning embodied music cognition, enactive music cognition, ecological approaches to the study of musical meaning, and the study of musical gestures. Mixed methods have already been employed for investigating movement perception in music performance (Schacher et al., 2015). Moreover, practice has increasingly been adopted as a methodological component in the arts (Smith and Dean, 2009), in music research (Dogantan-Dack, 2015), and more specifically within the New Instruments for Musical Expression (NIME) community (Green, 2016).

Figure 1.1 in the following page and section 1.4 show how the dissertation is structured following the different components of the methodology and how these components are interrelated.

1.4 Dissertation Overview

This dissertation is structured in three parts, reflecting the three main methodological components:

Part I *Thinking Movement in Music: Theory*

The first part reviews and discusses relevant theoretical studies of music-related body movement.

Chapter 2 presents *embodied* accounts of music cognition, which underpin the idea that body movement plays a central role in the fundamental processes underlying musical experience.

Chapter 3 presents the key notions of gesture, traditional musical instrument, and affordance; framing their definition within the embodied cognition theoretical framework. The different approaches for the classifications of musical gestures are discussed, and the the idea of multiple functional components within a musical gesture is finally put forward. The definition of traditional musical instrument, gesture, and the categories of musical gestures are extensively used in the experiments described in part II.

Part II *Observing Movement in Music: Analysis*

The second part is focused on the analysis of motion data collected through empirical experiments. Methods and techniques for data analysis are also described and discussed.

Chapter 4 describes an exploratory study aimed at observing how the variation of bow strokes' length and quantity affects the body movements of a viola player. This is done through quantitive analysis of motion descriptors computed from motion capture data and qualitative observations.

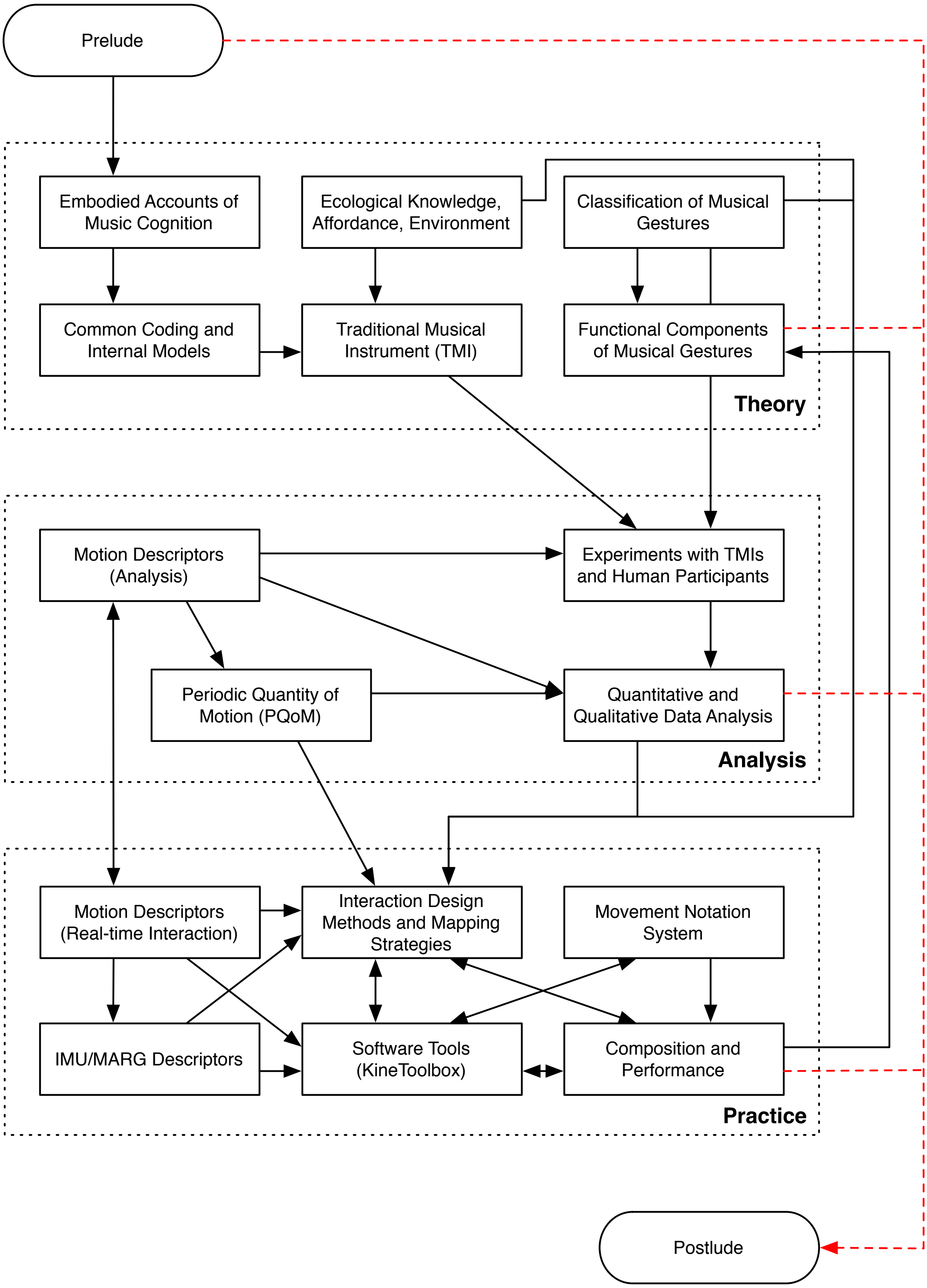


Fig. 1.1 Visual overview of the dissertation.

Chapter 5 investigates the extent in which the movement vocabulary of violin performance is part of the embodied knowledge of individuals with no experience in playing the instrument. People who cannot play the violin were asked to mime a performance along an audio excerpt recorded by an expert. They do so by using a silent violin, specifically modified to be more accessible to neophytes. Compared to the previous chapter, this study involves a higher number of participants and employs more sophisticated analysis techniques, including the analysis of a novel motion descriptors named Periodic Quantity of Motion (PQoM).

Part III *Using Movement in Music: Practice*

The third part is dedicated to the development and implementations of tools for actively employing the musicians' body movements in music composition and performance. Hardware and software solutions are described and a series of short instrumental compositions involving body movement are used as case studies. The design and development of these tools and compositions have manifold implications, as they are informed by the issues discussed in the previous parts of the dissertation but also contribute to further the discourse on this topics.

Chapter 6 describes the implementation of gestural mapping strategies for performance with a traditional musical instrument, multimodal motion sensors, and live electronics. The approach adopted is informed by concepts of embodied music cognition and functional aspects of musical gestures, which were discussed in part I.

Chapter 7 explores technical and conceptual issues related to the representation and mediation of body movement in music performance through digital technology. The chapter also reports on a case study of a musical piece where motion sensor technologies are employed to track the movements of the musicians playing their instruments. Motion data is then used to control the electronic parts of the piece in real time. In light of this case study, the chapter discusses how musical instruments can be seen as repositories of a gestural vocabulary and the score as a script that elicits an emerging choreography. The chapter includes the definitions of motion descriptors for IMU/MARG[1] sensors, some of which were inspired by the motion descriptors used in the experiments presented in part II.

Chapter 8 presents a collection of software tools for motion-sound interaction and two instrumental music pieces that make use of said tools. By describing both the technical solutions and their implementation, this chapter addresses the challenges

[1]Inertial Measurement Unit / Magnetic, Angular Rate and Gravity. See section 7.3.1 for more information.

related to the use of the body movements of the musicians as a musical feature in composition and performance. Different hardware and software solutions, and approaches to parameter mapping for motion-sound interaction are discussed, and a method for integrating movement in traditional music notation is described.

Chapter 9 presents a knowledge-based, data-driven method for creating mappings between the performance movements of a musician and sound synthesis. This is done by using a database of multimodal motion data collected from multiple subjects coupled with sound synthesis parameters. This dataset is used to build a topological representation of the performance movements of the subjects. This representation is then employed to interactively generate training data for machine learning algorithms and define complex mappings for real-time performance. This method draws from the theoretical concepts described in part I (especially ecological knowledge, see section 3.3) and the data collection and analysis techniques adopted in part II.

Finally, **chapter 10** concludes the dissertation by providing a summary of the results achieved throughout the research project in relation to the research aims and questions presented in chapter 1.

Part I

Thinking Movement in Music: Theory

Chapter 2

Embodiment and Experiencing Music Through Action

He had a far, far more accurate theory about beauty than I did. He did not tell it to me in words, but with his gestures and his eyes, with the music that he played on his flute, and with that forehead of his which emerged in the moonlight.

Yukio Mishima
The Temple of the Golden Pavilion

This chapter presents *embodied* accounts of music cognition, which underpin the idea that body movement plays a central role in the fundamental processes underlying musical experience.

2.1 Introduction

In recent years, music-related body movement has been subject to extensive interdisciplinary research. Contributions from several fields such as musicology, cognitive psychology, neuroscience, and computer science have brought about new ideas and perspectives, giving rise to new paradigms for the study of gesture and music (Godøy and Leman, 2010; Gritten and King, 2006, 2011). In particular, new insights from research on embodied music cognition (Leman, 2008a) inspired new viewpoints and required a rethink of the relationship between human body and musical experience. Within this theoretic framework, music perception is *embodied* (i.e. closely linked with bodily experience) and *multimodal*, in the sense that music is perceived not only through sound but also with the help of visual cues and feelings of motion, such as kinaesthetic sensations and kinematic images (Godøy, 2010).

2.2 Embodied Music Cognition

Embodied music cognition adopts the assumptions that stemmed from the theory of *embodiment* to look into previously unexplored aspects of musical experience. The notion of embodiment brought about a paradigm shift in cognitive science (Gibbs, 2005). By paying particular attention to the sensorimotor patterns of action and perception of a cognitive system situated in an environment, this approach provides a model to investigate cognition that goes beyond the traditional dichotomy between physiology and psychology. Traditional approaches to the study the mind often confine mental processes in brains (Rowlands, 2010). Embodied theories on the contrary support the idea that there is no real separation between mental processes and body, describing cognition in terms of dynamics between agent and environment rather than computation of passively received information (Chemero, 2009). From this perspective, perception and action become interdependent and are conjointly carried out in sensorimotor activity. Cognition therefore relies on mechanisms that occur outside of the skull; more specifically, the "[a]ctivity in the nervous system is linked to high-level cognitive processes by way of embodied interaction with culturally organized material and social worlds" (Hutchins, 2010, p. 712). Embodied accounts put the emphasis on the role of the human body as a mediator for meaning formation, and the central idea behind embodied music cognition is that "an intentional level of musical interaction is established through corporeal articulations and imitations of sensed physical information provided by the musical environment" (Leman and Maes, 2014, p. 236). In other words, at the core of the embodied music cognition paradigm there is the assumption that gesture and corporeal imitation are fundamental constituents of musical expressiveness (Leman, 2008a). In this context, "musical gestures can be described in an objective way as movement of body parts, but they have an important experiential component that is related to intentions, goals, and expressions" (Leman, 2012, p. 5). By acting as a mediator, the body will build up a repertoire of gestures and gesture/action consequences, or what Leman calls a *gesture/action-oriented ontology* (Leman, 2012, p. 5). This repertoire can be considered as a collection of movements made to achieve a particular goal (actions) linked with the experiences and sensations resulting from such actions. The coupling of actions and perceived sensations forms an engine that guides our understanding of music. Through this mechanism, the listener is able to relate physical aspects of movement in space to expressive qualities, intentions and inner feelings. Conversely, perceived patterns of musical expression recall previously learned knowledge of the corresponding body movement. This continuous two-way mirroring process allows the listener not only to attribute intentions and feelings to music but also to predict the outcomes of actions and project them onto the music (Maes et al., 2014). This is what Leman calls *action-perception coupling system*; it forms the basis of musical intentional communication and expressiveness, which then elicits several social phenomena such as empathy and social bonding (Leman, 2008a, 2010, 2012).

Theoretical assumptions brought forward by embodied music cognition are supported by empirical studies showing analogies between movements evoked in listeners and those of the performing musicians (Godøy and Leman, 2010; Leman et al., 2009a). Further studies show that the performers' movements contribute significantly to the perception of expressive intention (Broughton and Stevens, 2009; Glowinski et al., 2014b; Vuoskoski et al., 2014) and features of the melody, harmony, timbre, and rhythm are reflected in the movements of the perceivers (Burger et al., 2013a; Küssner et al., 2014; Naveda and Leman, 2010; Nymoen et al., 2013; Visi et al., 2016).

In particular, Godøy (2006) explores the relationship between sound and embodied cognition further by extending the concept of *objet sonore*[1] put forward by Schaeffer (1966). Schaffer's sonic objects are fragments of sound perceived holistically, typically in the range between 0.5 and 5 seconds (Godøy et al., 2010). Perceptually salient timbral, dynamic, and pitch-related envelopes are typically to be found within this timescale (Schaeffer et al., 1998). Godøy (2006) hypothesises that in music perception there is a continuous process of mentally tracing the shape of these highly significant features of sound objects. Therefore, resting on the idea of embodied cognition, he posits that there is a continuos mental recoding of sound into multimodal gestural-sonorous images, or *gestural-sonic objects*. Nymoen et al. (2013) studied these concepts further through experiments in which participants were asked to move their hands following perceptual features of short sound objects and motion capture data was analysed and correlated with a set of quantitative sound features.

[1]In the literature, this has been translated to English in various ways, including 'sonic object', 'sonorous object', and 'sound object'.

2.3 A Sense-giving Activity

Embodiment therefore has a central role in the cognitive and emotive processes necessary to make sense of music. It plays part in an interconnected array of functions such as affect processing, conceptualisation, tool use, and more (Leman and Maes, 2014). The empirical studies mentioned in the previous section give proof of evidence that there is a clear link between the perception of musical sound and human actions carried out through body movement. However, this should not be seen as a unidirectional phenomenon, where embodiment (i.e. being in a body) explains the fact that perceiving music makes people move. The link between perception and action works also the other way around: music-driven movement may facilitate the perception of music. Being engaged in music-related body movement is not simply the outcome of perceiving music, it is a *sense-giving activity*. In other words, music makes us move, but it is also by moving to it that we actively attribute meaning to it. Classical disembodied approaches see experiencing music as a unidirectional process, where perception and action are separated (Fig. 2.1). Action is the output of processing information obtained through perception,

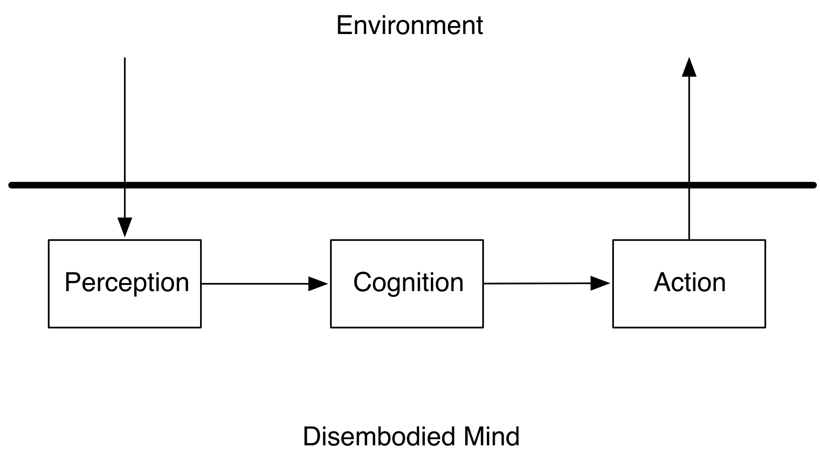


Fig. 2.1 A graphical representation of a traditional cognitivist disembodied model of cognition. Mind and environment are clearly separated and they interact through a unidirectional process. Cognition is segregated between perception and action as in the "sandwich" model described by Hurley (1998).

and the processing mechanisms are handled by cognition. In contrast, from an embodied perspective sense-giving musical activities arise from the mutual interaction between perception and action, and the sensorimotor mechanisms activated by music have a direct role in facilitating the perception of the music (Fig. 2.2).

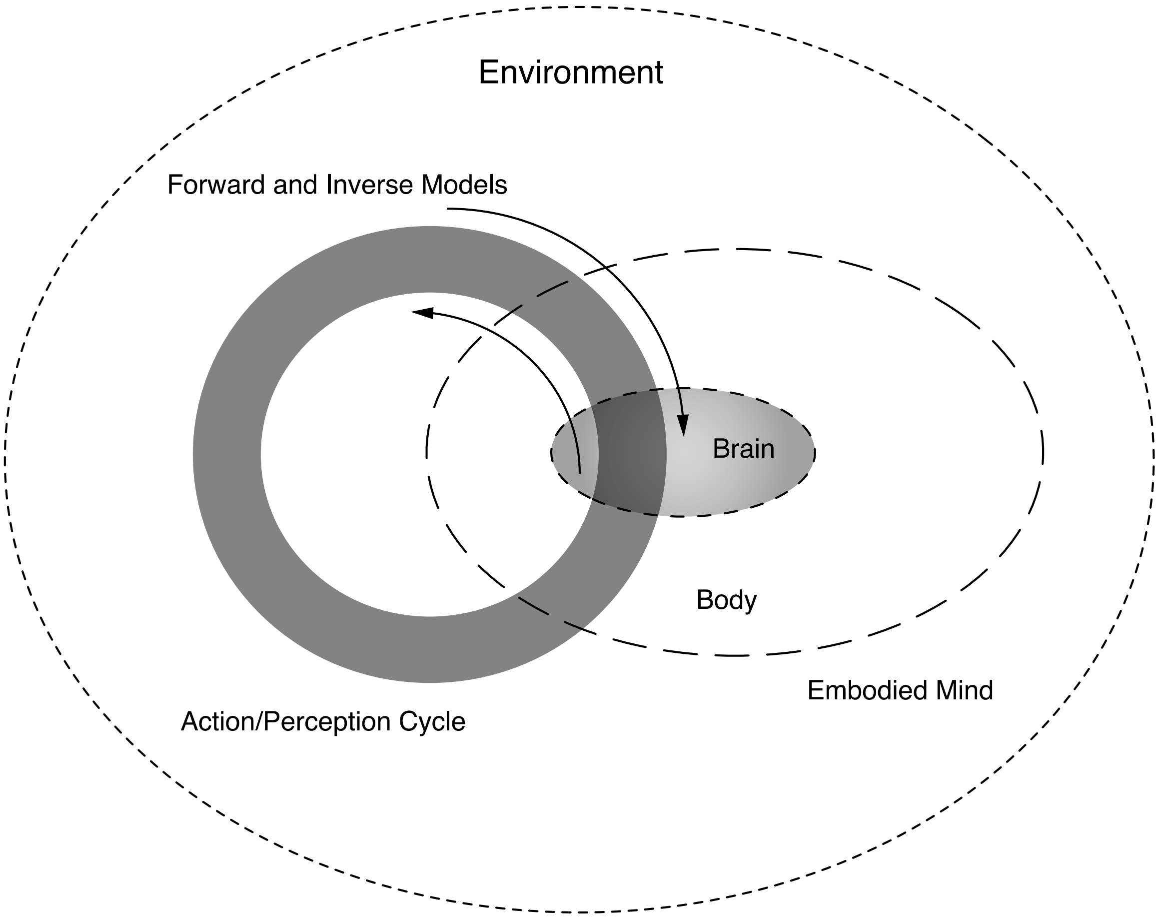


Fig. 2.2 A graphical representation of embodied cognition, freely inspired by the one proposed by Hinton (2014).

2.4 Common Coding and Internal Models

Maes et al. (2014) further develop the idea of a close coupling between perception and action by adopting the general framework provided by the common coding theory (Prinz, 1990, 1997). In a nutshell, common coding theory states that action representations (underlying planning and execution of an action) and stimulus representations (underlying the expected multi-sensory perceptual outcome of that action) are coded and stored not separately, but together in a common representational medium. In other words, their representation recruits both sensory and motor areas of the brain, substantiating the idea of close coupling between perception and action. Hommel et al. (2001) build upon common coding, proposing a more comprehensive framework denominated Theory of Event Coding (TEC) and thereby giving an overview of empirical evidence supporting this theory in domains of chief importance for music cognition, such as sensorimotor synchronisation and ideomotor action. More recently, common coding and theories have been extended in order to account for affective states (Lavender and Hommel, 2007) and address implications for action planning and control (Hommel, 2009).

The integration of motor and sensor representations leads to bidirectional internal models of the relationship between the body and the external environment, with inverse and forward components. Forward models represent an information flow from action to perception, meaning that they allow to predict the perceived outcome of an action that is being planned or executed. In other words, these models help make sense – through action – of something that is being perceived. Likely, one of the most common examples of employing forward models in music is moving to the beat to improve timing perception. In fact, empirical studies have shown that moving improves sensitivity to temporal intervals (Iordanescu et al., 2013; Manning and Schutz, 2013). Inverse models on the other hand represent information flow from perception to action, meaning they allow to predict the motor commands necessary to obtain a particular sensory state. From a musical point of view, inverse models enable us to predict the movements required to obtain a desired musical sound. This is obviously crucial for musical instruments playing and confirms the idea that listening to music is a multimodal experience that includes kinaesthetic sensations. Following these assumptions, obtaining knowledge about the sensorimotor relationships afforded by a musical instrument and the integration of these relationships into internal models clearly has an influence on the perceptual processes shaping the musical mind of both the performers and the audience. The idea that body movements may actually modulate the perception and understanding of music is supported by empirical studies, such as those carried out by Vines et al. (2006), Vuoskoski et al. (2014), and Broughton and Stevens (2009) to name but a few.

Within the framework described so far, playing a musical instrument can be seen as the result of a dynamical interaction between the sensorimotor system and the constraints and opportunities the instrument and other elements of the environment (other musicians, audience, score, social conventions, etc.) afford. Music arises from this high-dimensional, dynamical, mutual interaction between body, mind, and environment.

2.5 Enactive Music Cognition

The enactive approach in the cognitive sciences (Thompson, 2007) places emphasis on the assumption that the experience of the world is determined by the mutual interaction between the sensorimotor capacities of an organism and the environment in which it is situated. Action and perception are addressed as a unitary entity, thus proposing yet another alternative to traditional cognitivist accounts. Mainstream disembodied views represent the mind as a processor that passively receives information from a given environment and computes internal representations. From an enactive standpoint, cognitive systems actively participate in the generation of meaning, engaging in active, transformational interactions (Stewart et al., 2010).

Matyja and Schiavio (2013) review the current research in enactive music cognition and compare it to cognitivist and embodied approaches. Enactive approaches focus on a definition of music as something pertaining to the environment in which the *agents* (i.e. the organisms that experience the music through their action/perception cycle) are *embedded*. This differs from cognitivist and embodied accounts as there is no commitment to the explanatory role of mental representations. From the enactive standpoint, music is experienced *directly*, without intermediate mental representations, disclosing itself through sensorimotor patterns of action and perception. According to advocates of enactivism, this approach is able to explain the most intimate and primal levels of human musical involvement by focusing on the basic, pre-conceptual aspects of music cognition (Matyja and Schiavio, 2013).

While fully endorsing the premises of an embodied approach to musical processing with an emphasis on the role of action, Schiavio (2014) criticises certain aspects of the embodied music cognition paradigm put forward by Leman and Maes (2014). He argues that some of the experiments carried out by Leman and colleagues are still influenced by classical computationalist approaches to music psychology. Schiavio refers to the series of studies involving the *guqin* (a traditional Chinese zither) aimed at understanding expressive gesturing while listening to music (Henbing and Leman, 2007; Penttinen et al., 2006). In one of these experiments (Leman et al., 2009b), velocity patterns of a guqin player's movements were compared to those of individual listeners moving their arm along with the music. In Schiavio's view, the results obtained could still be explained by describing an *input-output* mechanism that the listeners would employ to process particular features of the music internally and eventually provide a movement output. This would not take into consideration an important aspect of a truly embodied view of musical sense-making, which is the fact that actions are directed towards a goal (i.e. they are *goal-directed*). From an embodied view, it is the recognition of a goal driving the action that enables forms of pre-linguistic understanding (Gallese, 2009). According to Schiavio (2014), research accounting for an embodied standpoint in music cognition should be focused on specific goal-directed actions rather than on a broader set of movement patterns. In addition to that, Schiavio advocates for a consistent consideration of the fundamental role of the environment in embodied theories of musical understanding, identifying it as one of the challenges that future research should try to address. As he puts it, "music cognition is always ecologically embedded, thus not reducible to structures inside the skin. [...] [T]he *system* [...] should be an integrated set of embodied actions and musical environment, without positing an explicit division between *inner* and *outer*." (Schiavio, 2014, p. 5).

2.6 Summary and Comments

This chapter presented the main tenets of interdisciplinary theories supporting an embodied approach to the study of musical understanding. An embodied model for music cognition that takes into consideration action-based effects on music perception is therefore described and compared to traditional disembodied models. In addition, arguments put forward by proponents of enactive music cognition are also reported, in order to broaden the discourse on embodied music cognition and indicate some of the challenges to address in developing the embodied paradigm further.

Within the scope of this dissertation, this theoretical framework not only supplies a conceptual apparatus for the analysis and creative use of music-related movement. It also provides a set of motivations for developing tools aimed at facilitating the analysis of body movement in music and its concrete use in composition and performance. In other words, acknowledging the central role of body movement in processes underlying musical experience both informs and drives research towards ways of employing and understanding movement in music.

Traditional cognitivist views considered body movement performed by people experiencing music merely as the byproduct of internal processes involving a system of symbolic representations. Only relatively recently, an increasing amount of studies supported by empirical evidence has begun to unfold how the human motor system and its actions can affect our experience of sound and music. Conceiving music as a multimodal medium is an essential step towards musical practice that systematically includes body movement among its expressive features.

The emphasis on *situatedness* and the central role of the environment put forward by enactivist accounts may also incentivise the adoption of a research approach that fruitfully integrates standard experimental designs in controlled laboratory environments with studies in ecological settings and actual musical practice.

Chapter 3

Gestures and Musical Instruments

> Sometimes, finally, the meaning aimed at cannot be achieved by the body's natural means; it must then build itself an instrument, and it projects thereby around itself a cultural world.
>
> Maurice Merleau-Ponty
> *Phenomenology of Perception, p. 169*

> Well, the things we do is just we play a very powerful, high-energy type of rock 'n' roll. We move around onstage a lot.
>
> Kurt Cobain
> *Interview*

This chapter presents the key notions of gesture, musical instrument, and affordance; framing their definition within the embodied cognition theoretical framework presented in chapter 2.

3.1 Movement and Gesture

Given the theoretical background introduced in chapter 2, it is useful to further unpack and analyse key terms that will be employed throughout this dissertation such as movement and gesture. Interestingly, in the Oxford thesaurus these two terms are presented as synonyms[1]. However, in the interdisciplinary field of music-related movement studies, various research communities addressed the need of a clearer distinction (Jensenius et al., 2010), especially since the word *'gesture'* has been extensively employed across multiple research contexts and with different semantic connotations (Jensenius, 2014).

Even though the thesaurus presents movement and gesture as synonyms, the individual lemmas as found in the Oxford dictionary show a particularly significant distinction. The first definition of 'movement'[2] reads "An act of moving: *'a slight movement of the body'*", whereas 'gesture'[3] is defined as "A movement of part of the body, especially a hand or the head, to express an idea or meaning: *'Alex made a gesture of apology'*". Here, the crucial distinction is that a gesture – contrary to movement – is expected to convey *meaning*. In this research context then, movement denotes physical displacement of an object in space. The term is often used interchangeably with 'motion', especially in scientific or technological settings (e.g. "motion data"). Gestures, on the other, hand also carry meaning, which can be considered as "the mental activation of an experience" (Jensenius et al., 2010, p. 13). Clearly this term becomes central when musical expression comes into play and it is a core notion within the framework of embodied music cognition. The notion of *gesture* has, in fact, the considerable advantage of working as a conceptual bridge between movement and meaning, consequently bypassing the boundary between physical world and mental experiences. This *monistic* (Leman, 2010) quality of gestures clearly makes them a key concept of the embodied music cognition paradigm, as they allow the listener to link physical aspects of movement in space to expressive qualities, intentions, and inner feelings. Acknowledging the issues such a broad and sometimes vague word might lead to, Jensenius et al. (2010) give a comprehensive look at the term and its uses in music research in order to present a clearer overview. Gesture is thereby defined from three different viewpoints: communication, control and metaphor. Within the category of *communication*, the role of gestures is to convey meaning in social interactions. The focus is on the linguistic and communicative aspects of gesture rather than on the body movement itself. From the *control* viewpoint, gestures are intended as parts of an interaction system. For example, in human-computer interaction gestures are movements of the body carrying information that can be processed and recognised by a motion capture system. Gestures are instead used as *metaphors* when they "work as concepts that project physical movement, sound or other type of perception to cultural

[1] http://www.oxforddictionaries.com/us/definition/american_english-thesaurus/movement

[2] http://www.oxforddictionaries.com/definition/english/movement

[3] http://www.oxforddictionaries.com/definition/english/gesture

topics" (Jensenius et al., 2010, p. 14). Here, gestures are mental images evoked from observable actions and/or musical sound, arising from both cognitive and social processes.

Mazzola and Andreatta (2007) argue that Hugues de Saint-Victor's medieval definition of gesture cited by Schmitt (1990) is still one of the most adequate to this day. They cite Katsman's translation from Latin, which reads:

> Gesture is the movement and figuration of the body's limbs with an aim, but also according to the measure and modality proper to the achievement of all action and attitude. (Katsman, 2006)

Even though this definition confines it to the body's limbs, it clearly states that gesture is *movement with an aim*. This is akin to the concept of goal-directed action of the embodied accounts described in chapter 2.

In the Human-Computer Interaction (HCI) context, Kurtenbach and Hulteen (1990) define gesture as "a movement of the body that contains information". In his study of non-verbal communication, Kendon (2004, p. 15-16) suggests that gesture is "a label for actions that have the features of manifest deliberate expressiveness" and "[h]ow [these actions] are interpreted, however, will depend upon context", thus acknowledging the centrality of the environment in which actions are embedded. In their definition of gesture informed by these accounts, Caramiaux et al. (2015b, p. 3) state that "a gesture is a dynamic movement of the body (or part of the body) that contains information in the sense of deliberate expression", stressing that deliberateness differentiate gesture from simple movement.

While initially difficult to pinpoint, the notion of gesture is a key concept in contemporary music research, as it provides the necessary conceptual bridge between movement and meaning, at the core of embodied accounts of music cognition. Thus, it can be stated that *gestures are a vehicle for the construction of musical meaning*. From these definitions of gesture, it also emerges that the term should be carefully framed and defined in order to avoid ambiguity. Throughout this dissertation, I will try to use the word 'gesture' when appropriate, that is when there is an explicit involvement of meaning and intentionality. This also applies to a more technical context, such as in chapter 9 when I will use the phrase 'gesture templates' to refer to couplings of motion and sound synthesis data used as training information for a machine learning model. In other cases where the link with meaning and intention is not immediately evident, I will opt for the terms 'movement' or 'motion', which are considered synomyms.

3.2 Classifications of Musical Gestures

3.2.1 Instrumental Gesture

The role of body movements in instrumental music performance has been analysed further by looking at functional aspects of specific musical gestures. In their seminal work, Cadoz and Wanderley (2000) review the analysis of the playing technique of the pianist Glenn Gould carried out by Delalande (1988). In that work, Delalande proposed three levels of gesture classification based on functional characteristics:

- **Effective gesture**, which is necessary for the production of sound.
- **Accompanist gesture**, that supports the performance of effective gestures. Cadoz and Wanderley (2000, p. 77) note that Delalande suggests that "its function is as related to imagination as to the effective production of the sound", thus implying that they contribute to the musical experience. Another term often used interchangeably to designate these gestures is *ancillary gestures* (Wanderley et al., 2005, p. 97)
- **Figurative gesture**, linked to a more metaphorical, symbolic definition of gesture and thus without a direct correspondence to physical movement.

Using Delalande's classification as a starting point, Cadoz and Wanderley (2000, p. 78) define *instrumental gesture* as a subgroup of effective gesture, consisting of "the actual instrument manipulation and playing technique" produced by what they call the "gestural channel". That is a channel of human communication acting as a means of action on the physical world and – *at the same time* – a means of communication of information. Action and perception are therefore inextricably coupled, similarly to the embodied model of cognition described in section 2.3.

To further clarify the definition of instrumental gesture, Cadoz and Wanderley (2000, p. 79) state that these gestures are characterised by *physical interaction with material objects*, the ways this interaction evolves over time *can be mastered by the subject*, and the instrumental gestures themselves may *support communication and engender the production of a material action*. In addition, instrumental gestures are distinguished from empty-handed gestures (e.g. conductor gestures), which are purely semiotic and do not involve energy transfer.

In an earlier publication, Cadoz (1988) further classified instrumental gesture in three functional subclasses:

- **excitation gesture**, which provides the energy required to produce the sound and may be either instantaneous (e.g. plucking a string, hitting a drum) or continuous (e.g. bowing a string);
- **modification gesture**, involved in the modification of the instrument properties and thus altering the sound produced by excitation gesture and may be either parametric (e.g. fretting a stringed instrument) or structural (when the structure of the instrument is altered, e.g. by installing a mute in a trumpet);
- **selection gesture**, involved in the selection of a subset of similar elements of an instrument, such as the keys of a piano. They may be sequential (e.g. fingering an arpeggio on a keyboard) or parallel (e.g. fingering a chord on a keyboard).

3.2.2 Functional Categories of Gestures

More recently and to better understand the function of musical gestures in performance, Jensenius et al. (2010) expanded on the work of Delalande (1988) and Cadoz and Wanderley (2000) identifying four functional categories; *sound-producing* gestures, *sound-facilitating* gestures, *sound-accompanying* gestures and *communicative* gestures.

Sound-producing gestures are closely involved in the production of sound and can be further subdivided into *excitation* gestures (e.g. plucking or bowing a string, hitting a key) and *modification* gestures (not producing sound by themselves but affecting its quality, e.g. operating pedals in a piano).

Sound-facilitating gestures are not directly involved in the production of sound but provide support for sound-producing gestures in different ways, therefore influencing the resulting sound. For example, strumming a chord on a guitar involves movement of various parts of the arm and the joints. With the body being a complex system, many *support* gestures (even if very subtle) are often necessary for a sound-producing gesture to occur. *Phrasing* gestures are another type of sound-facilitating gesture, they are closely connected to the phrases being played in the music and are an integral part of the performance, both helping the musician to perform the piece and improving the experience of the listeners. This is true also for *entrained* gestures, like tapping a foot or moving in synchronicity with the music. These gestures help the musician to keep track of the tempo, therefore facilitating the performance, and they may in addition support the interplay with other musicians and enrich the experience of the listeners.

Sound-accompanying gestures are distinguished from the previous two categories as they have no role, whether direct or indirect, in the production of sound but they follow features of the music. The most obvious example is indeed dancing, other examples are tracing the contour of sonic elements and mimicking sound-producing gestures in the air, like in so-called air instrument performances (Godøy et al., 2006a).

Communicative gestures, on the other hand, are primarily intended for communicating between performers and listeners. Examples are facial expressions or lifting the hand theatrically before strumming a chord on a guitar. While it could be said that all gestures are to some extent communicative gestures, this category is specifically intended for gestures whose function is chiefly communicative. In fact, it is important to underline that gestures have usually multiple functions and do not fall exclusively into one category.

A case study employing these concepts to devise mapping strategies for a performance involving wearable sensors and electric guitar is described in chapter 6.

3.2.3 Blurred Boundaries

As clearly specified by both Cadoz and Wanderley (2000, p. 79) and Jensenius et al. (2010, p. 25) musical gesture categories are neither exclusive nor independent and their boundaries are usually blurry, since in actual musical performances gestures tend to have multiple functions. The rationale behind these categorisations is not to provide a comprehensive taxonomy of gesture in music performance. Rather, these concepts are a useful starting point for making sense of the role a gesture may have in a performance.

This section described a set of useful conceptual tools for addressing the main subject of this thesis, that is body movement in instrumental music performance. In their background investigation on mixed methodologies for studying movement perception in music, Schacher et al. (2015) sketch out a map that includes about forty gestural classes and subclasses proposed by key authors in the field. That is useful as a reference and also as a snapshot showing the complexity of musical gesture analysis.

3.3 Musical Instruments, Ecological Knowledge, Affordances, Constraints

3.3.1 Musical Instruments and Instrumentality

Another key concept for this dissertation is that of *musical instrument*. As noted by Kvifte (2008), there are several definitions of musical instrument, since the concept is quite difficult to define in a precise way and different research interests call for different definitions. In fact, the entry ‘Instruments, classification of’ on Grove Music Online avoids an explicit definition:

> 'Musical instrument' is a self-explanatory term for an observer in his own society; it is less easy to apply on a worldwide scale because the notion of music itself in such a wide context escapes definition. (Wachsmann et al., 2017)

Even though implicit, this formulation highlights an important aspect of musical instruments: their "cultural embeddedness". An object is a musical instrument because it has undergone a process of culturalisation that makes it recognisable as such for the member of a society. This, as implied above, ties it with definition of music itself, which is constantly being renegotiated across times and cultures. Kvifte (2008) argues that a more precise and explicit definition of 'musical instrument' is not necessarily more useful for research purposes:

> Concepts are tools for grasping the world around us, and their utility in research is measured by their ability to let us make new and relevant questions. If a traditional and relatively precise definition of 'instrument' excludes large areas of contemporary musical practice from our field of study, we might be better off with less precise alternatives. (Kvifte, 2008, p. 55)

Hardjowirogo (2017) revisits Hornbostel's often cited definition, which states that "[f]or purposes of research everything must count as a musical instrument with which sound can be produced intentionally" (Hornbostel, 1933, p. 129). While Hornbostel's statement highlights sound production and intentionality as defining qualities of a musical instrument, Hardjowirogo (2017) points out that musical instruments are not the only things used to produce sound and, at the same time, are more than only sound-producing devices. She then suggests that the specificity that distinguishes a musical instrument from other sound-producing devices is expressed by the concept of *instrumentality*:

> "[. . .] instrumentality, or simply being a musical instrument must not be understood as a property an object as such has or has not. Rather, it seems to result from using something in a particular way which we think of as instrumental. Consequently, an object is not per se a musical instrument (ontological definition) but it becomes a musical instrument by using it as such (utilitarian definition)."(Hardjowirogo, 2017, p. 11)

Thus, instrumentality is a dynamic concept that is not tied to an object by design, it is rather the result of cultural negotiation. The "degree of instrumentality" is a dynamic quality of an object, that changes depending on context and processes of culturalisation. An object may be more or less instrumental according to various characteristics associated with instrumentality. Hardjowirogo (2017) compiles a preliminary list of the criteria that presumably play a major role in the construction of instrumentality, which includes sound production, intention, learnability, playability, cultural embeddedness, liveness and others.

Therefore, in the context of this dissertation, a musical instrument is any object that is used intentionally to perform music within a specific cultural environment, regardless of whether it was originally designed to be a musical instrument or not. Different instruments may have different degrees of instrumentality, and what is considered a musical instrument in a specific cultural environment may not be recognised as such in another context. Some subcategories of musical instruments relevant for this dissertation will be discussed in section 3.3.3.

3.3.2 The Instrument as Extension of the Body

The study of musical gestures and embodied music cognition also brought about a new understanding of the relationship between musician and musical instrument. From embodied perspectives, the musical instrument is *embodied in the body of the performer* (Hirose, 2002) and becomes *a natural extension of the musician* (Nijs et al., 2009). It is therefore part of the mediation together with the body, thus allowing a spontaneous corporeal articulation of the music, contributing to the formation and conveyance of embodied musical meaning. According to Godøy (2003), people continuously re-enact mental simulations of musical gestures when listening attentively to music, adding a motor-mimetic element in music perception and cognition. Additionally, Cox formulates a similar hypothesis: "we normally imagine (most often unconsciously) what it is like to make the sounds we are hearing" (Cox, 2001). Moreover, there is empirical evidence that the gestures of the instrumentalist can alter the sound perception of the listener (Schutz and Lipscomb, 2007).

Clark (2008) developed a theory of extended cognition, suggesting that artefacts serve as scaffoldings onto which cognitive processes are offloaded. This way, the mind extends beyond the body to include the tools, symbols, and other parts of the environment that we deploy to make sense of and engage with the world. Magnusson (2009) claims that musical instruments act as cognitive extensions, pointing out that "[t]echnological objects are [...] never neutral, they contain scripts that we subscribe to or reject according to our ideological constitution." In other words, a musical instrument is a conveyor of knowledge and our extended mind thinks through it.

Maes et al. (2014, p. 10) effectively delineate musical instrument playing as a dynamic system with multiple components having different weights but no causal priority, drawing a parallel with music listening:

> In the case of musical instrument playing, music can be considered as the result of a dynamical interaction between the musicians' motor and sensory system, the constraints and opportunities of the pre-composed musical notation, the musical instruments and the social environment, and the musicians' intentions, personality, mental states, etc. [. . .] Similarly, music listening can be considered as a dynamical process, in which the experience, the perception, and the understanding of music is guided and shaped by the intrinsic dynamics of the body, the mind, and the external environment.

With this theoretical framework in mind, it is clear that instrumentalists' gestures have considerable expressive potential. Gesture has been employed as an expressive element in musical practice across different genres and styles and has also inspired the development of several digital musical instruments (DMI) (Jensenius and Lyons, 2016; Miranda and Wanderley, 2006). To mention some applications, the composer Roberto Doati has written a series of pieces for guitar that make use of the gestures of the fretting hand of the performer to control parameters of live electronics (Doati, 2004). Maes et al. (2011) use the theory of embodied music cognition to inform a different approach to parameter mapping and develop a human-computer interface that facilitates gestural control over real-time digital signal processing of the singing voice. Camurri et al. (2001) instead employ a similar theoretical framework to implement interactive artistic applications and understand expressiveness in gestures using computational modelling.

3.3.3 Traditional Musical Instruments

I adopted the term *traditional musical instrument* (henceforth TMI) to define instruments that have a rich idiomatic (Huron and Berec, 2009) repertoire and an established set of playing techniques[4]. Examples include electric guitar, violin, electronic pianos, etc. As also mentioned in chapter 1, this term refers to the use and repertoire of the instrument rather than to its technological characteristics, and depends on cultural context. The term 'electronic' is used of instruments in which the sound is generated by means of analogue or digital circuitry, while 'electric' usually refers to instruments in which sound is generated mechanically and then amplified electrically (Davies, 2017). Finally, purely

[4]This does not rule out the possibility of extended and unconventional techniques employed by a minority of players.

acoustic musical instruments do not incorporate any device for electric or electronic sound manipulation (Libin, 2017). TMIs can be found in all of the above categories. The term does not simply refer to the physical object. Rather, it refers to the physical object *and* the established gestural repertoire related to it, the known constraints and opportunities it affords. For example – according to the definition of TMI used in this dissertation – a turntable as it is employed by DJs to perform scratching and other common techniques of 'turntablism' (Snapper, 2004) *is* a traditional musical instrument. This is because it is consistently employed in established musical practices and there is a set of techniques shared among instrumentalists (Hansen, 2002; Sonnenfeld and Hansen, 2016), regardless of the fact that the device itself was not originally designed to be a musical instrument. In other words, this term takes into consideration (and it is relative to) the environment in which the instrument is embedded.

The choice of focusing on TMI performance is motivated by the vast knowledge that listeners have of the gestural and sound aspects of each instrument, which is learnt through experience. For instance, most people immediately know what sound to expect when we see a person hitting a snare drum with a drumstick. Similarly, if we hear the sound of a violin we can easily associate the gesture of bowing to it. This is because there is a shared *ecological knowledge* (Godøy, 2010) of the instrument's repertoire of sound-producing gestures. TMIs in fact have fairly explicit and known *affordances* (Gibson, 1977) that delineate action relationships between the instrument and the musician, inform expectations in the listeners and can be used to devise mapping strategies for controlling electronic aspects of music performance (see chapter 6). Using embodied music cognition terminology, instruments have a rich action/gesture repertoire that the listeners can recognise during the performance. In fact, Nijs et al. (Nijs et al., 2009) note that expert musicians have an extensive toolbox of movement schemes that they can unconsciously select and perform in response to the challenges provided by the musical environment.

Using the concepts discussed in section 3.3.1, Traditional Musical Instruments are a category of musical instruments that have (at a certain time and within a certain cultural context) a "high degree of instrumentality", mainly thanks their "cultural embeddedness", to cite one of the criteria mentioned by Hardjowirogo (2017). Hence, similarly to the concepts of 'music' and 'musical instrument', TMIs constitute a dynamic category that changes over time and across cultures.

3.3.4 Environment, Affordances, Constraints

The concept of *affordance* was introduced by J.J. Gibson and further codeveloped by his wife Eleanor (Gibson, 2000).

> The *affordances* of the environment are what it *offers* the animal, what it *provides* or *furnishes*, either for good or ill. The verb to *afford* is found in the dictionary, but the noun *affordance* is not. I have made it up. I mean by it something that refers to both the environment and the animal in a way that no existing term does. It implies the complementarity of the animal and the environment. (Gibson, 1986)

Affordances can therefore be understood as properties of environmental structures that provide opportunities for action to organisms. The *complementarity* Gibson mentions refers to the fact that the physical capabilities of the organism are as important as the properties of the object in defining an affordance. For example, a guitar affords certain actions (e.g. fingerpicking) to humans that it does not afford to elephants. Thus, an affordance is defined by the relationship between certain properties of an object in the environment and the actions an agent can potentially perform.

In the past three decades, the concept of affordance has become hugely popular in several academic environments such as cognitive psychology, design, and human-computer interaction; thus undergoing several developments, criticisms, and refinements (Jenkins, 2008; Mcgrenere and Ho, 2000). One of the most notable developments – which adds notions particularly useful for this dissertation – was introduced by Norman (1988, 1999). Norman's *perceived affordances* differ from the notion introduced by Gibson. They can be *dependent on the knowledge or culture of the agent* whereas Gibson's affordances exist regardlessly of the ability of the actor to perceive them (Mcgrenere and Ho, 2000). This implies that the actions afforded by an artefact are culturally conditioned, so a violin might afford bowing to a member of one culture and might not afford the same action to a member of another. In other words, there can be an ecological knowledge of musical gestures related to an instrument that is shared by a certain group of people.

In music research, Clarke (2005) presents an approach to the study of music perception based on principles derived from James Gibsons ecological perceptual theory. Clarke stresses the importance of the environment and its social component in determining how affordances are perceived: "A violin, for example, affords burning, but social factors ensure that this is a rather remote affordance—which might only be realized in extreme circumstances or by an individual who had no regard for (or even deliberately disdained) the musical context which regulates its affordances." (Clarke, 2005, p. 38). Will Gibson presents a study of jazz performance (Gibson, 2006), showing the relevance of social action, embodied skills, and musical instrument affordance in the understanding of musical practice.

Magnusson (2010) employs the concept of affordance and pairs it with the one of *constraint*, claiming that while the two terms are complementary, the latter is a more productive analytical tool for the the design of DMIs compared to the common usage of affordance in HCI.

In part III (particularly in chapters 6 and 9), I will use constraints and affordances of various TMIs to develop mapping strategies for DMIs. These mappings will be used in compositions involving both TMIs and DMIs (see chapters 7 and 8). This integration of TMIs and DMIs could be seen as a form of instrument augmentation (Miranda and Wanderley, 2006; Newton and Marshall, 2011). However, if further developed and integrated into regular practice, instrument augmentations could hypothetically become part of the extended set of techniques of the instrument, similarly to how effect pedals are now a consolidated part of the electric guitar idiomatic style. Alternatively, instrument augmentation might also go towards building other instrumental identities, as in the case of the magnetic resonator piano (Mcpherson and Kim, 2012). As discussed in section 3.3.1, several factors are involved in the construction of instrumental identities, many of which require the involvement of wider communities and the unfolding of cultural processes. Thus, predicting how the concept of musical instrument – and therefore music itself – will evolve as the result of contemporary practice is very challenging, but knowing more about these cultural processes can certainly aid the design of new instruments and inspire the composition of new music.

3.4 Summary and Comments

This chapter presented the working definitions of some concepts pivotal to this dissertation. The notions of gesture and movement have been unpacked and analysed, showing the centrality of a clear definition of gesture that provides the necessary conceptual bridge between movement and meaning.

A set of categories useful for the classification of musical gestures have been presented, with particular attention to those related to musical instrument playing. Musical instruments are examined from embodied and extended cognition perspectives, pointing out an intimate relationship between the instrument and the body of the performer and a set of movement schemes that take part in the production of musical meaning. In addition, the working definition of *traditional musical instrument* has been presented, motivating the choice of the term and stressing the importance of context and environment in establishing what a traditional musical instrument is. Finally, the role of the environment is discussed further, exploring the notion of affordance from different stances and suggesting that the opportunities and constraints offered by a musical instrument have a distinct in role in our ways of making sense of music.

The theoretical framework and the supporting empirical studies presented so far substantiate the idea that body movement is not an incidental byproduct of instrumental music practice. Rather, movement plays a role in communicating musical expression to an audience (Broughton and Stevens, 2009) and therefore should be considered *a musical feature*. This motivates the development and implementation of technologies for the use of body movement in instrumental music performance and other musical contexts.

A system of categories is a useful tool for studying the properties of musical gestures. However, it is worth noticing that pursuing an exhaustive taxonomy of musical gesture is likely an impossible task. In addition, the proliferation of ambiguous terminologies might hinder further research instead of facilitate it. As mentioned in section 3.2.3, key authors behind the study of musical gestures specify that categories of musical gestures should not be considered neither absolute nor exclusive, but should rather serve as pedagogical tools and as an aid for the design of musical interactions (Cadoz and Wanderley, 2000, p. 79). The porous quality of this categorisation becomes even more evident in performances involving both traditional and digital musical instruments, as shown by the practical work discussed in chapter 8. Therefore, rather than defining additional categories, future research might go towards understanding how multiple gestural categories co-exist and interact in concrete musical gestures.

From the experience gained working on the practical and analytical parts of this dissertation, it is felt that a potentially fruitful way of dealing with the functional aspects of musical gesture is to think not exclusively in terms of *categories* but also in terms of *components*[5]. In other words, rather then placing gestures in categorical containers, functional components are attributed to gestures. As noted in section 3.2.3, key authors who proposed concepts for the classification of musical gestures stressed the fact that, in actual performance, musical gestures often fall into multiple categories (Cadoz and Wanderley, 2000; Jensenius et al., 2010). Thinking also in terms of components would have the advantage of making the multifaceted functional identity of musical gesture immediately clear, thus avoiding the risk of seeing categories as independent or exclusive containers. This way of operating favours a system where hierarchies are inverted. Here,

[5]The definition of 'component' used here is in some ways analogous to the one employed when dealing with vectors such as velocities and forces (themselves familiar concepts in the field of music-related movement research): constitutive parts of a whole with individual magnitudes and directions.

gestures are not seen as elements belonging to one ore more sets (e.g. functional categories), but are considered as entities constituted by a number of (functional) components with different qualities and magnitudes that can vary over time. For example, consider a cello vibrato gesture. The iterative movements of the fretting hand on the bowed string that modulate the pitch can certainly be considered a *sound-producing/modification* gesture (see section 3.2.2). Now, imagine the dynamics and the depth of the vibrato progressively increasing to obtain a dramatic effect. A *modification component* will still be part of the gesture. In addition to that, a second *communicative* or *theatrical* (Jensenius et al., 2010, p. 24) component might increasingly become a constitutive part of that same vibrato gesture, depending on the intentions and style of the performer and also on other environmental factors that may affect the expectations of the perceivers.

This approach should not be seen as substitutive of the classification systems mentioned so far, but rather as an additional tool for analysis that afford a fluid conceptualisation of the functional roles of musical gestures. This might facilitate analytical models where functional aspects of music-related actions vary over time, possibly providing concepts useful for studying dynamic phenomena such as coarticulation and chunking (Godøy et al., 2010). Moreover, this interpretation of the functional qualities of music-related movements could also be helpful for the design of motion-based musical interfaces and for composition and performance of music that deliberately employ movement as an expressive feature.

The definition of Traditional Musical Instrument adopted in this dissertation is closely related to the concept of affordance, as they both depend upon factors related to the environment. The constraints and opportunities afforded by TMIs are consistently (en)acted upon in musical practice, thus contributing to the creation of the idiomatic style of the instrument. Perceived affordances offer a cognitive grasp for the musician and other music perceivers (see chapter 5 for an empirical analysis), influencing the expectations and understanding of what is happening musically.

To conclude the theoretical part of this dissertation with a catchphrase that gives the gist of some of the topics discussed so far, we could say that *"musical instruments are weapons loaded with culture"*.

Part II

Observing Movement in Music: Analysis

Chapter 4

Effects of Different Bow Stroke Styles on the Movements of a Viola Player

Anton Giulio Bragaglia
Uomo che suona il contrabbasso (1911)

This chapter[1] describes an exploratory study of different gestures and body movements of a viola player resulting from the variation of bow strokes length and quantity. This work also served as a pilot study for the larger-scale study described in chapter 5.

4.1 An exploratory experiment: viola bow strokes

The aim of this exploratory study is to observe how the variation of a musical feature within the piece affects the body movements of the performer. A viola player is asked to

[1]This chapter is based on Visi et al. (2014a). The full peer-reviewed article can be retrieved online at `http://www.federicovisi.com/publications/`.

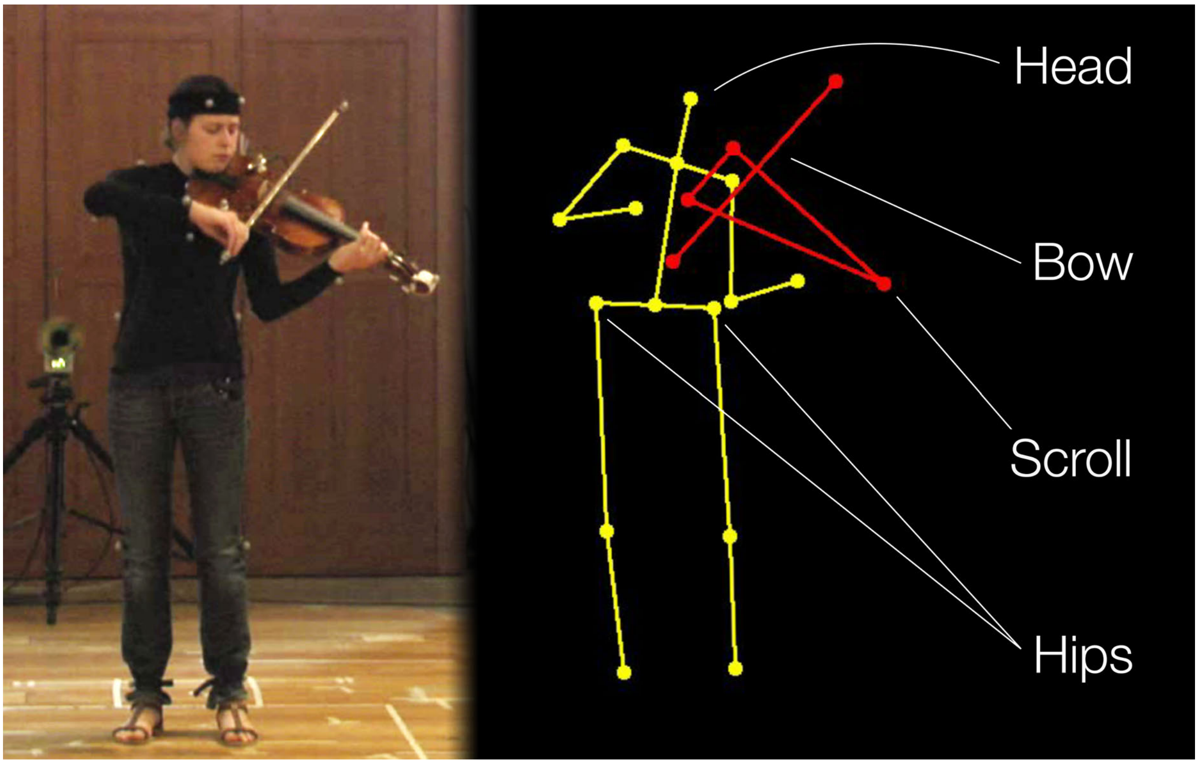


Fig. 4.1 Still from the video stream (left) and skeleton generated from the motion capture data with labels of the four joints used in analysis (right).

perform two short pieces of music four times each, every time with different directions regarding length and quantity of bow strokes. The performances were recorded using a multimodal recording platform that included audio, video and motion capture (MoCap) data obtained from high-speed tracking of reflective markers placed on the body of the performer and on the instrument. The intent is to observe how the variation of a musical feature also affects the movement of the performer and, secondly, if there are correlations in the way sound-producing gestures and ancillary gestures vary according to the different bow stroke styles. Quantitative analysis focused on motion capture data, while audio and video were used as reference. We extracted measurements of quantity of motion and velocity of different parts of the body, the bow and the viola. Results indicate that an increased activity in sound-producing and instrumental gestures does not always resonate proportionally in the rest of the body and the outcome in terms of ancillary gestures may vary across upper body and lower body. Past studies have observed gestures and movements of string instrument players, focusing on motion features of different bow strokes (Rasamimanana et al., 2008), the physical interaction between the player and the instrument (Schoonderwaldt and Chen, 2009) and expressivity and interaction in ensemble playing (Glowinski et al., 2013a). Similar studies have been carried out for other musical instruments, such as piano (Thompson and Luck, 2011), harp (Chadefaux et al., 2012) and clarinet (Desmet et al., 2012a).

4.1.1 Pieces and bow stroke variations

Two excerpts of two different musical pieces were chosen: a sarabande from *Pièces de violes, Livre I* (1686) by Marin Marais (Fig. 4.2) and a passage from Tchaikovsky's *Barcarolle*, from *The Seasons* (1876, Fig. 4.3). These pieces were chosen to allow comparison of body movements between two different styles (baroque and romantic respectively).

The viola player was asked to perform each piece in four different versions:

- as she would normally interpret it according to the score (this variation was labelled **'Natural'** in graphs for short);
- using the full length of the bow, from tip to frog, during each bow stroke (labelled **'Long'**);
- using only the central part of the bow (about one third of the total length, labelled **'Short'**);
- by performing a bow stroke for every note, therefore increasing the total amount of bow strokes necessary to perform the piece (labelled **'Many'**).

Fig. 4.2 Baroque tune: excerpt from *Pièces de violes, Livre I* (1686) by M. Marais.

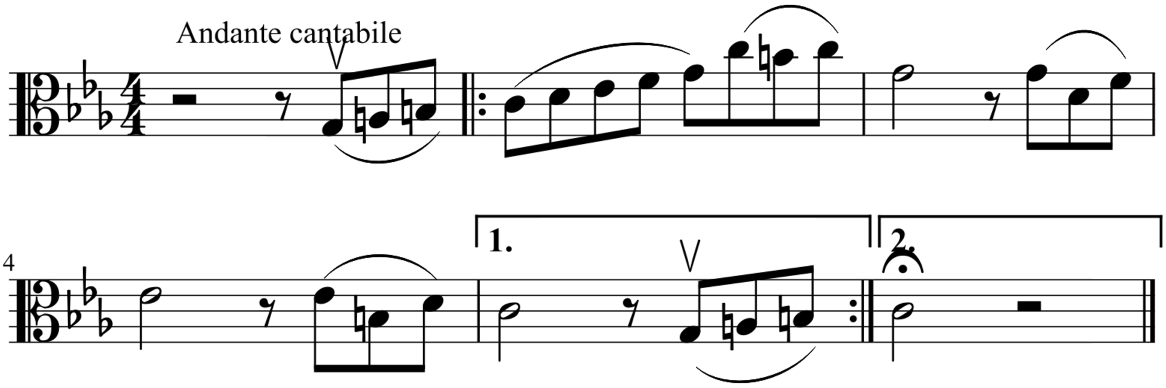


Fig. 4.3 Romantic tune: excerpt from *Barcarolle*, from *The Seasons* (1876) by P. I. Tchaikovsky.

4.1.2 Equipment and setup

The recording took place in an auditorium/concert hall, suitable for experiments in an ecological setting. The musician performed on a stage where MoCap data was recorded using Qualisys Oqus cameras. The viola player wore a suit equipped with 19 reflective markers: 3 on the head, 4 on the shoulders, 1 on the back, 1 on the sternum, 2 on the elbows, 2 on the wrists, 2 on the hips, 2 on the knees and 2 on the ankles. Additionally, 3 reflective markers were placed on the viola, 2 on the body and 1 on the scroll. Markers were also placed on the frog and the tip of the bow. Overall, 24 markers were used. Along with the MoCap data, video and audio were recorded by means of a digital videocamera and a piezoelectric microphone placed on the viola. The multimodal stream of data was recorded and synchronised using EyesWeb XMI[2].

4.2 Data analysis and results

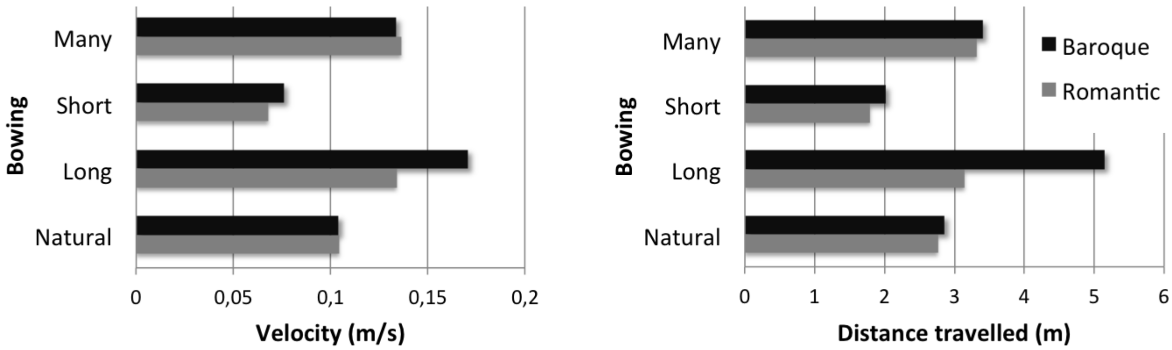
4.2.1 Movement feature extraction

MoCap Toolbox for MATLAB (Burger and Toiviainen, 2013) was used to extract various kinematic features. First, the data was trimmed to the duration of each performance. To simplify the movement analysis, the MoCap data was restructured. This was done using joints, also called secondary markers, obtained by averaging the locations of a subset of markers. Of the initial 24 markers, 4 joints (head, hips, scroll and bow) were taken into account. This particular choice allows for comparison between instrumental sound-producing gestures (bow) and ancillary sound-facilitating gestures (head, hips, scroll). The joint of the scroll consisted of only one marker. The head joint was calculated from the three head markers, the hips from the two markers on the left and right hip and the bow from the markers at the tip and the frog (Fig. 4.1). Subsequently, two movement features were extracted from the joint location data:

1. Velocity for head, hips, scroll and bow was calculated in order to measure the activity of the different body parts. The instantaneous velocity was averaged for each joint, in order to obtain a general value for the eight different performances.

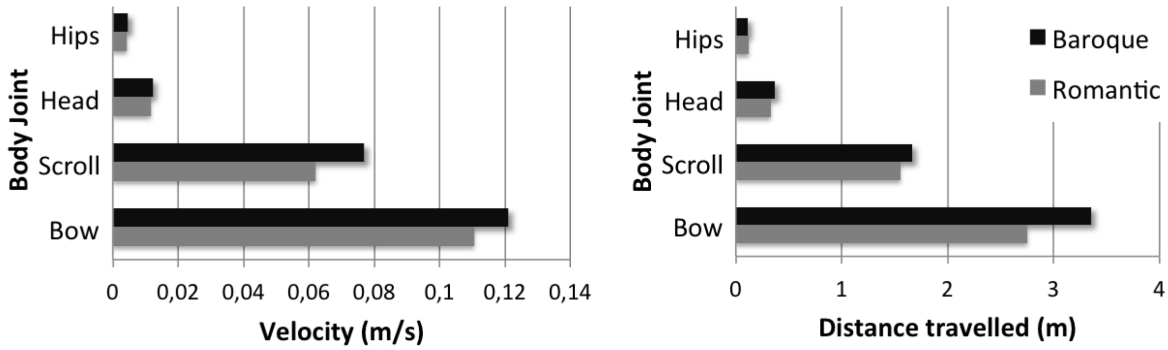
2. The cumulative distance travelled by each joint was taken into account to measure the quantity of motion (QoM). This gives a good indication of the total amount of movement of each body part over the whole performance (Burger et al., 2013b).

[2] http://www.infomus.org/eyesweb_eng.php



(a) Average velocity of the bow joint. (b) Quantity of Motion of the bow joint.

Fig. 4.4 Comparison of the movement features of the bow joint between the two musical excerpts.



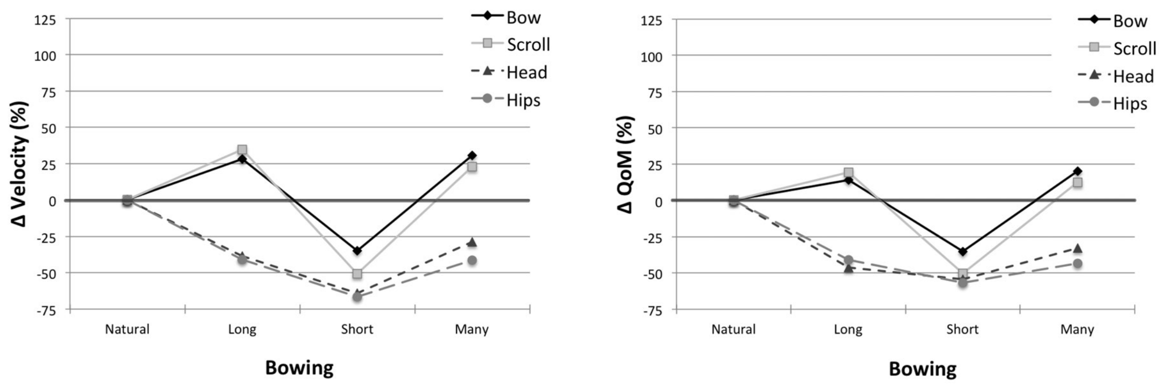
(a) General velocity. (b) General Quantity of Motion.

Fig. 4.5 Comparison of the general movement features of the four joints in analysis between the two musical excerpts.

4.2.2 Results

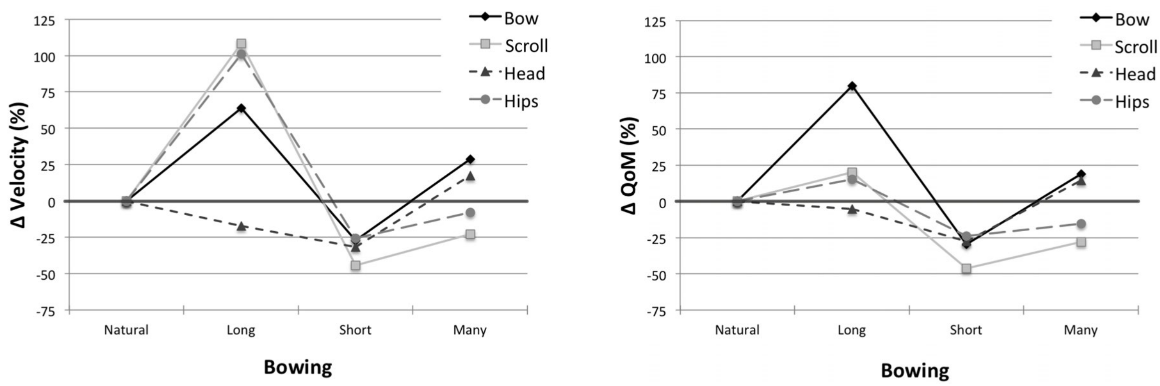
Velocity and quantity of motion of the bow joint indicate the most immediate outcome that the bow stroke's variations had for both pieces (Fig. 4.4a, 4.4b). In the 'Romantic' piece performance, bow velocity and QoM were much lower for the short bowing condition. In the 'Baroque' piece, the long bowing condition stands out more. In general, the bow is the most active of the four body joints for each piece in each performance, followed by the scroll of the viola (Fig. 4.5a, 4.5b).

Since the variations only involved instructions about bowing, changes of velocity and quantity of motion in other body parts are not directly induced by the task. For each joint, the velocity and QoM of the 'Natural' bowing performance were taken as a reference (0%) to compare against the values obtained in the other variations.



(a) Differences of velocity. (b) Differences of Quantity of Motion.

Fig. 4.6 Comparison of the movement features of the 'Romantic' piece.



(a) Differences of velocity. (b) Differences of Quantity of Motion.

Fig. 4.7 Comparison of the movement features of the 'Baroque' piece.

The velocity and QoM graphs of the 'Romantic' piece performance (Fig. 4.6a, 4.6b) appear similar, showing analogous ratios among the four different performances. The movement and activity induced in the scroll are very similar to that in the bow, and give even larger extremes in the long and short bowing performance. A different trend is observed for head and hips. First of all, their overall velocity and QoM in general is much lower than that of scroll and bow (Fig. 4.5a, 4.5b). As opposed to the scroll, the head and hips do not increase in QoM and velocity when longer bow movements are used. For short bow movements, the head and hips are more consistent with the scroll as their QoM is reduced by a half and their velocity decreases even more. Overall, head and hips are active the most in the 'Natural' performance variation.

The outcomes for the 'Baroque' piece differ to a certain extent from the 'Romantic' ones. Here, velocity and QoM do not change equally across the different variations

(Fig. 4.7a, 4.7b). The QoM for the scroll increases when longer bow strokes are used, but not as much as the value for the bow joint. On the contrary, the velocity of the scroll more than doubles in the 'Long' bowing condition, as compared to the 'Natural' condition. When short bow strokes are used, both QoM and velocity of the scroll decrease, but not as much as in the performance of the 'Romantic' piece. The head joint in the 'Baroque' performances follows a similar trend as in the 'Romantic' piece: its velocity and QoM do not increase with longer bow strokes and decrease even more in the 'Short' bowing condition. On the contrary, the hips do not follow the head movement this time. Similarly to the scroll, its QoM increases with long bow strokes and its velocity doubles, while in the 'Short' bowing condition it decreases again.

When many bow strokes are used, another difference between the 'Romantic' and 'Baroque' piece can be observed. There is an increased effect on the head in the latter whilst the scroll is more affected in the former. Moreover, many bow strokes induce almost as much movement and velocity as longer bow strokes in the performance of the 'Romantic' piece, which is not the case for the 'Baroque' piece, where longer bow strokes induce much more movement in other body parts as well. In contrast, the velocity and QoM of head and hips in the 'Romantic' piece are reduced to less than a half in the 'Short' bowing variation.

In general, short bow strokes induce the least movement and activity in all the body parts and long bow strokes induce the most QoM and velocity in bow and scroll. When many bow strokes are used, only the activity and movement of the bow is consistently increased in both pieces, compared to the natural performance.

4.3 Summary and Comments

This chapter described an exploratory study of different gestures and body movements of a viola player resulting from the variation of bow strokes length and quantity. The aim was to observe how the variation of a musical feature within the piece affects the body movements of the performer.

The movement data shows that ancillary and instrumental gestures may shift in analogous ways across the different bow stroke variations, but may also diverge. Similar effects of different bow strokes are found both in the 'Romantic' and the 'Baroque' pieces. Nevertheless, some variations occur, especially in the 'Long' and 'Short' bowing conditions. This is partially due to the musical structure and conventional musical style of both pieces. A sarabande is a Baroque dance, which is usually performed with shorter and lighter bow strokes to give the piece a dancing character. In the 'Romantic' piece, the performance guideline *andante cantabile* requires more bowing and vibrato to create the intended sound effect. Moreover, when comparing the original scores of both pieces, we see that in the 'Romantic' piece there are more notes per slur than in the Baroque sarabande, which implies the use of more bow in the former, and less in the latter. This explains why using short bow strokes in the 'Romantic' piece has more effect on scroll and bow than using long ones and the other way around for the 'Baroque' piece.

What happens with the head and the hips is more ambiguous. As the QoM and velocity increase in the 'Long' and 'Many' bowing condition, the movements made by head and hips are reduced in comparison with the natural performance. A possible cause for this effect could be the constraints posed by the task. By adding supplementary directions regarding bow strokes, the performer focuses on the additional movements required in order to accomplish the task, which may reduce spontaneous movement in other body parts. However, a different effect is observed in the performances of the 'Baroque' piece. Here, the velocity and QoM of the hips increase when the performer uses long bow strokes, and the same can be observed in the head joint when many bow strokes are used. Again, the difference in musical style and structure could partially explain these contradictions. The long bowing condition implies more changes in the movements required to perform the 'Baroque' piece as compared to the 'Romantic' piece. Even with the task constraints in mind, these changes could affect the movements of the hips too. Still, this does not explain why this effect does not occur in the movements of the head when long bow strokes are used. Moreover, there is increased head movement in the 'Baroque' performance, but only when the performer uses many bow strokes. In this condition however, the hips seem unaffected. A study by Glowinski et al. (2014a) shows similar results. Three violinists performed a piece in metronomic, emphatic and concert-like styles and movements of head, torso, forearms, hips and violin were measured. Here, the movement amplitude of the hips was significantly different from the other body parts. The differences between upper and lower body parts were interpreted as part of a compensation process in which the lower body is seen as an anchoring point to enhance stability and compensate for the higher movement activity of the upper body.

Overall, it is worth noting that increasing QoM and velocity of instrumental gestures resonate into ancillary gestures of the rest of the body in different ways and that this resonance may be hindered if the difficulty of the task increases. Different musical styles may also have an effect on how movement changes across the different bow stroke variations. It is important to note that variating the bow strokes alters the musical outcome in terms of timing, timbre and loudness of the notes. However the main goal of the experiment is to observe the results in terms of body movements and underline that variating bow articulations in the score alters not only the sound but also the corporeal expressivity of the performer, therefore affecting the experience of the performer and the audience on multiple levels.

In a broader perspective, the purpose behind this study was to approach possible ways in which movement and gesture can be employed as an expressive musical feature, whether directly determined in the score (an example in this direction is shown in section 8.4.2) or indirectly induced by other musical features (as in the musical piece described in section 7.4). It is not clear yet how gesture can be fully integrated with other expressive features in composition and performance, but further research-led practice may lead to new insight. Moreover, movement in music performance is highly idiosyncratic; it depends on anatomical differences between players (Dahl et al., 2010) and their different approaches to the instrument (Chadefaux et al., 2012). This preliminary work involved only one performer, so different playing styles among different performers could not be compared and statistical testing is clearly beyond the scope of this exploratory study. However, consistency with other studies (Glowinski et al., 2014a) could be observed and the adopted methodology and the focus on the relation between pre-determined variations of musical features and resulting variations of body movement has inspired the larger-scale analysis work described in chapter 5 and has informed the practice-led works described in chapters 7 and 8.

Gestural idiosyncrasies may constitute interesting expressive challenges for composers, leading them to work closely with performers in order to examine relationships between scored musical features and body movement, and explore the expressive possibilities of writing *gesture-aware* music.

Chapter 5

Analysis of Mimed Violin Performance Movements of Neophytes

What sculptors do is represent the essence of gesture. What is important in mime is attitude.

MARCEL MARCEAU

I've been imitated so well I've heard people copy my mistakes.

JIMI HENDRIX

5.1 Introduction and Motivation

The study presented in this chapter[1] investigates the extent to which the movement vocabulary of violin performance is part of the embodied knowledge of individuals with no experience in playing the instrument. People who cannot play the violin were asked to mime a performance along an audio excerpt recorded by an expert. They do so by using a silent violin, specifically modified to be more accessible to neophytes. Motion data analyses suggest that, despite the individuality of each performance, there is a certain

[1]This chapter is based on Visi et al. (2016). The full peer-reviewed article can be retrieved online at http://www.federicovisi.com/publications/.

consistency among participants in terms of overall rhythmic resonance with the music and movement in response to melodic phrasing. Individualities and commonalities are then analysed using Functional Principal Component Analysis.

While the study in chapter 4 observed how variation in bow stroke styles affected the movements of the performer, the experiment in this chapter aims at empirically exploring the the notions of shared and ecological knowledge of the gestural repertoire of a musical instrument (see section 3.3.3). Albeit having a somewhat different focus, this study builds upon the methodology adopted in chapter 4. Motion data is once again collected using marker-based motion capture, and a similar procedure based on a series of short trials is adopted. This time though, the analysis method is more sophisticated as it involves a higher number of participants, a novel motion descriptor that takes into account the music tempo (Periodic Quantity of Motion), and Functional Principal Component Analysis. The traditional musical instrument involved this time is a violin, member of the same instrument family of the viola used in the previous study, with which it shares several playing techniques and movement patterns. Similarly to the experiment in the previous chapter, this study focuses on areas directly involved in sound-producing movements (bow, right wrist) and ancillary movements (hip, head). This is motivated by the results described in chapter 4, which suggested that there are non-obvious relationships between the kinematic features of sound-producing and ancillary movements that may lead to new insights into the relationship between the instrument, the body of the performer, and the music being played.

As pointed out in chapter 3, musical instruments have a *a repertoire of sound-producing gestures* that contribute to build the *ecological knowledge* associated to that instrument. Hence, this shared knowledge affects one's musical experience, by creating expectations and guiding musical understanding. In fact, by adopting an ecological approach, musical perception is seen as an active experience influenced by a highly-structured environment rather than a passive, disembodied phenomenon. From this perspective *"exposure to the environment shape perceptual capacities of an individual"* and *"perception and actions are inextricably bound together"* (Clarke, 2005).

The goal of this study is to empirically explore the shared knowledge of the gestural repertoire of a well-known musical instrument among people who have no previous experience in playing that particular instrument. This is done by analysing the motion data gathered during an experiment where neophytes are asked to mime a violin performance. The analysis focuses on several body parts and movement features, in relation to the music and in comparison to the actual performance of an experienced violinist.

This experiment draws its motivation from the assumption that the musician encodes gestures in sound and the listener can decode particular aspects of them through corporeal imitation. As Leman notes, the listener is capable of grasping music as intended moving form and perception and understanding of musical expressiveness is based on corporeal resonance behaviour: *"Obviously the movements of the listener are not [...] the same as the movements of the player. What is more or less the same [...] is the motor system that encodes and decodes sonic forms."* (Leman, 2008a). Therefore, a more detailed analysis of the extent of the gestural vocabulary of an instrument also among non-experts can contribute to the understanding of musical perception and expression.

A relevant aspect of the design of this experiment is the use of an actual violin, specifically modified to not emit any sound when bowed and to be more accessible to people who have never used one before. Previous studies have analysed so-called "air performances" of experts and beginners mimicking the use of various instruments (Dahl, 2014; Godøy et al., 2005). Here, the choice of using an actual instrument is motivated by the adoption of an ecological approach, assuming that the relationship with the object (indeed part of the aforementioned environment) and its affordances (Gibson, 2000, 1977) may have a significant impact on the movements of the subjects. In addition, experience using tools has also been the subject of embodied music cognition research (Leman et al., 2010) and the concept of affordance has seen renewed interest in multidisciplinary music research (Altavilla et al., 2013; Menin and Schiavio, 2012).

The analysis of the motion data gathered during the experiment focuses prevalently on intermediate and high-level movement descriptors. This is motivated by ecological perceptual theories suggesting that, when processing information, people seem to be aware of high-level features more directly than lower-level features (Clarke, 2005). Therefore, high-level movement features are expected to be more readily identified and shared by the participants. Moreover, body movement and entrainment in response to music are complex and dynamic phenomenons. Therefore, movement analysis should try to address complex patterns from multidimensional motion data, rather than single values that capture a particular feature of a movement segment. Amelynck et al. (2014) proposed a new method that avoids this segmentation and takes into account the complete movement dynamics. They analysed the spontaneous bodily responses of people to a musical stimulus and tried to model expressiveness in terms of commonalities and individualities using Functional Principal Component Analysis (FPCA) (Ramsay, 2006).

5.2 The Experiment: Material and Methods

5.2.1 Participants

A total of thirteen participants took part in the study. This includes twelve neophytes (7 male, 5 female, average age: 33.4, SD of age: 9.8) and one experienced violinist (male, aged 23), who performed and recorded the stimuli for the experiment. All participants gave their informed consent and were free to take breaks or abandon the experiment at any point. Ethical approval was granted by the Arts and Humanities Research Ethics Sub-committee at the Faculty of Arts and Humanities, Plymouth University. Participants were also asked to fill out a brief anonymous questionnaire with basic personal data and information about their musical background.

5.2.2 Stimuli

Participants were asked to mime a violin performance using the modified violin along 5 randomly-ordered musical stimuli, which consisted of brief solo violin excerpts recorded by the experienced violinist. Stimuli were between 8.5 and 34 seconds long and were chosen to cover a variety of different styles and instrumental techniques.

List of Stimuli

- Antonio Vivaldi *"Violin Concerto in A minor, Op 3, No 6, RV 356"* (1711)
- Kaija Saariaho, *"Nocturne for solo violin"* (1994)
- Camille Saint-Saëns *"Le Carnaval des Animaux - 10. Volière"* (1886)
- Niccolò Paganini *"Caprice No. 1 'The Arpeggio' in E major: Andante"* (1819)
- Sergei Prokofiev *"Five Melodies for Violin and Piano, Op. 35bis"* (1925)

This study focuses on the data collected using the first and second stimuli. The first stimulus consists of bars 1–12 of the first movement (Allegro) of Vivaldi's Violin Concerto in A minor (Fig. 5.1), whereas the second one includes bars 45–48 of the Nocturne for solo violin by Kaija Saariaho (Fig. 5.2).

5.2.3 Apparatus

Data collection was carried out at the Interdisciplinary Centre for Computer Music Research (ICCMR), Plymouth University, United Kingdom and at fourMs - Music, Mind,

Fig. 5.1 Excerpt of the violin part of Vivaldi's Violin Concerto in A minor. The audio recording of the first twelve bars was used as stimulus for the experiment.

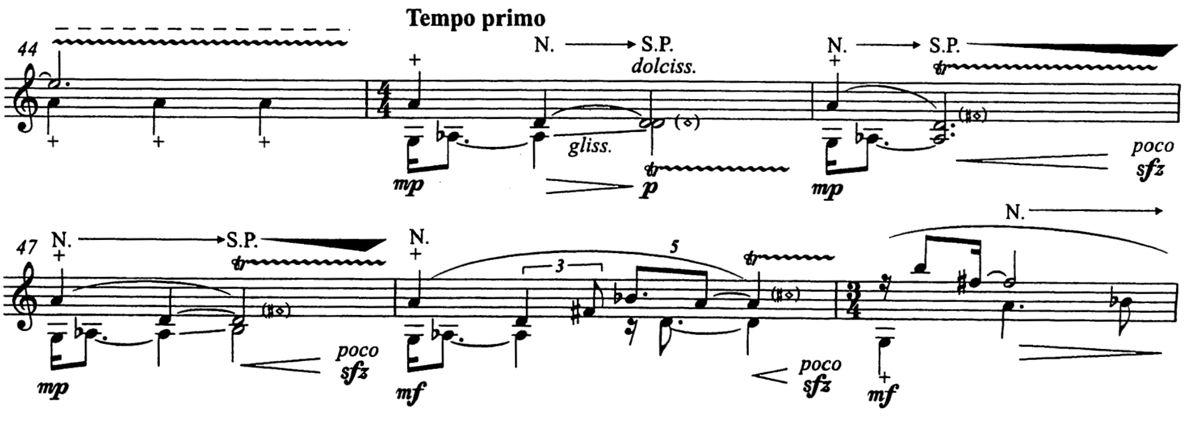


Fig. 5.2 Excerpt of Saariaho's Nocturne for solo violin. The audio recording of bars 45–48 was used as stimulus for the experiment.

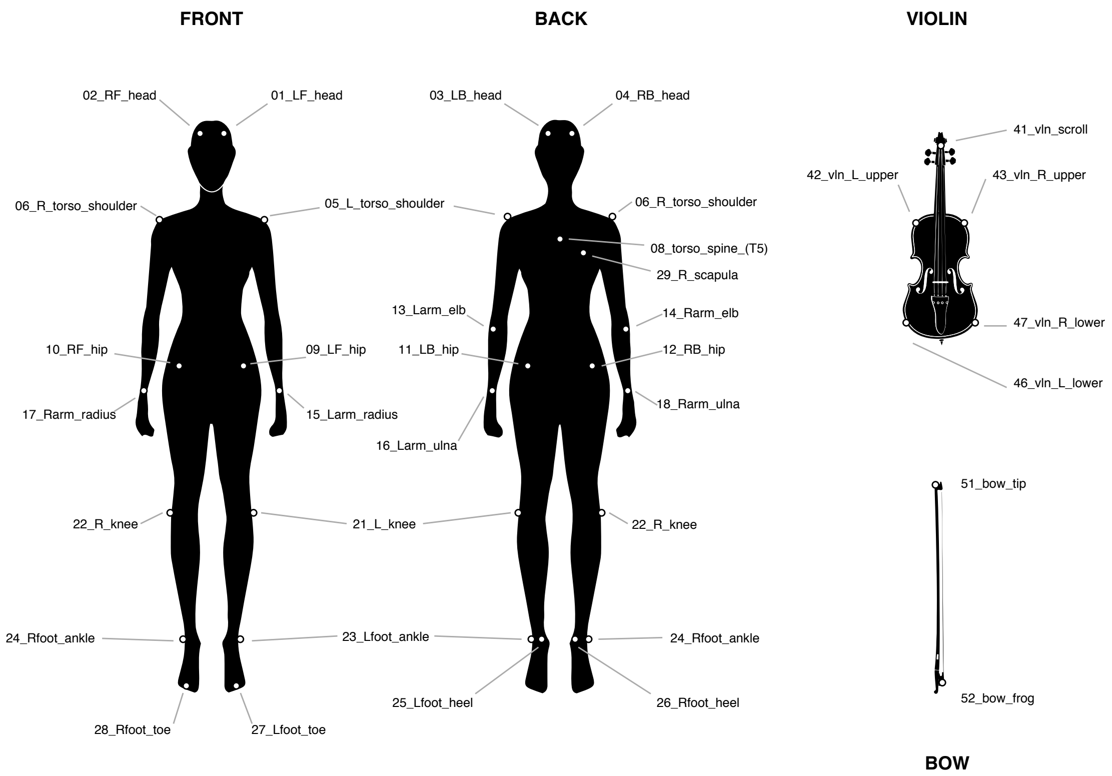


Fig. 5.3 Marker locations and labels.

Motion, Machines, University of Oslo, Norway. In Plymouth, participants' movements were recorded using a six-camera marker-based optical motion capture system (Natural Point Optitrack Flex 3[2]) tracking at a frame rate of 100 Hz. A total of 33 reflective markers were attached to each participant and to the instrument and were located as follows[3]: LF head, RF head, LB head, RB head, L shoulder, R shoulder, spine (T5), LF hip, RF hip, LB hip, RB hip, L elbow, R elbow, L wrist (radius), L wrist (ulna), R wrist (radius), R wrist (ulna), L knee, R knee, L ankle, R ankle, L heel, R heel, L toe, R toe, R scapula[4], violin scroll, violin L upper bout, violin R upper bout, violin L lower bout, violin R lower bout, bow tip, bow frog (see Fig. 5.3).

In Plymouth, an additional marker located on the sternum of the participants was used. However, the data associated to that marker was eventually discarded as it contained too many dropouts due to the frequent occlusion caused by the right arm during bowing

[2] http://www.optitrack.com

[3] L=Left; R=Right; F=Front; B=Back. A similar configuration can be found in (Burger et al., 2014).

[4] Used to obtain an asymmetrical marker set, useful for marker identification and tracking. Not used for analysis.

movements. That marker was therefore not used in the subsequent recording sessions in Oslo. The stimuli were played back through a pair of Genelec 8020C loudspeakers using a DAW[5]. The audio interface also generated the SMPTE signal used for synchronising audio, video and motion capture sources. The audio in the room was recorded by a pair of condenser microphones placed in a XY stereo configuration as well as by a video camera used to film the sessions.

In Oslo, the performances were recorded using a nine-camera marker-based optical motion capture system (Qualisys Oqus 300[6]) using the same frame rate (100 Hz) and marker configuration (except for the sternum marker) used in Plymouth. The feed from a digital video camera was recorded within the Qualisys Track Manager software alongside the motion tracking data. The stimuli were played back using the same model of loudspeakers and the same DAW software while recording and playback of the various sources was synchronised using a custom Max[7] patch.

The participants were asked to simulate the performance using a modified violin designed specifically for the experiment. This violin was fitted with a support system that allowed the instrument to be safely strapped to the shoulder of the participant. This was done in order to allow the participants – who, in most cases, had never held a violin before – to move with more confidence without being afraid to drop the instrument. Two thin metal plates soldered to a metal strip that follows the profile of the bridge were mounted on the violin body above the strings (see Fig. 5.4). This add-on had a dual purpose—it helped novices to quickly overcome the initial difficulties of holding the bow in a correct standard playing position and it prevented contact between the strings and the bow hair, hence making the violin silent.

5.2.4 Procedure

The expert violinist was recorded first. He performed all the selected excerpts, which provided both the audio stimuli for the neophytes and video and motion data to use as a benchmark for the analysis of the participant's movements.

Each neophyte was recorded individually. For each stimulus, the participant was asked to first listen to the audio once in order to familiarise themselves with the music and then use the modified violin to mime a performance along the played back audio twice. Audio, video and motion data were recorded during each trial.

[5] http://www.reaper.fm
[6] http://www.qualisys.com
[7] https://cycling74.com

Fig. 5.4 The modified violin used for the experiment.

5.3 Analysis of periodicity and phrasing using Mocapgrams and Periodic Quantity of Motion

5.3.1 Movement Data Preprocessing

The motion data was first preprocessed, labeled and exported to C3D files using Optitrack Motive and Qualisys Track Manager. The C3D files were then loaded in MATLAB using MoCap Toolbox (Burger and Toiviainen, 2013). The 33 markers described above were then transformed into a set of 23 secondary markers, which in the MoCap Toolbox framework are referred to as 'joints' (Burger et al., 2014; Burger and Toiviainen, 2013). The locations of these joints are represented in Fig. 5.5. Seven joints are obtained by calculating the centroid of two or more markers: joint 1 (head) is the midpoint of the four head markers; joint 2 (manubrium) is the midpoint of the shoulder markers, joint 5 (left wrist) of the left ulna and radius markers, joint 8 (right wrist) of the right ulna and radius markers, joint 9 (mid torso) of the spine and the four hip markers, joint 10 (root) of the four hip markers, joint 11 (left hip) of the the two left hip markers, joint 15 (right hip) of the right hip markers. On the violin, joint 20 (violin left bout) is the midpoint of the two violin left markers whereas joint 21 (violin right bout) is the midpoint of the two

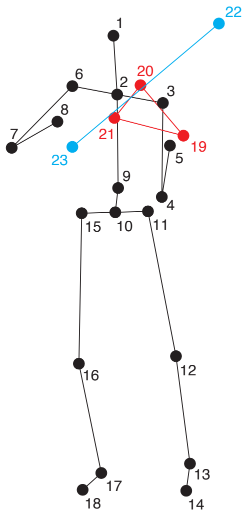


Fig. 5.5 Joints configuration: 1=head; 2=manubrium; 3=left shoulder; 4=left elbow; 5=left wrist; 6=right shoulder; 7=right elbow; 8=right wrist; 9=mid torso; 10=root; 11=left hip; 12=left knee; 13=left ankle; 14=right toe; 15=right hip; 16=right knee; 17=right ankle; 18=right toe; 19=violin scroll; 20=violin left bout; 21=violin right bout; 22=bow tip; 23=bow frog.

right ones. The locations of the remaining joints (3, 4, 6, 7, 12, 13, 14, 16, 17, 18, 19, 22, 23) are identical to the location of the respective markers (see Fig. 5.5).

5.3.2 Comparative movement data analysis using full Mocapgrams

By plotting Mocapgrams (Jensenius et al., 2009) (a graph in which position coordinates of each marker are normalised and projected onto an RGB colorspace) it was possible to do a preliminary analysis and observe recurring patterns and periodicities in the motion data. Fig. 5.6 shows full Mocapgrams for the performances of the expert violinist and of one of the neophytes (top left and top right graphs respectively) of Vivaldi's Violin Concerto excerpt (henceforth 'first stimulus'). Regular colour patterns in the horizontal rows corresponding to each marker suggest periodicity in certain parts of the body and the instrument. As an example, the thinnest pattern can be observed in the right elbow and wrist joints (labeled '7_R_elbow_J' and '8_R_wrist_J' respectively), which is consistent with the pattern visible in the bow markers ('22_bow_tip_J' and '23_bow_frog_J'). This shows, expectably, a certain coherence in the movement of the bow and the arm that holds it, as well as high frequency periodicity caused by the repetitive bowing movements. Similarly, it is straightforward to notice that the left toe of the expert ('14_L_toe_J') changes position only three times throughout the whole take.

For the purpose of this study, Mocapgrams are useful not only to observe general periodicity in the movement of certain parts of the body during the performance; by providing an overall view of all the motion data, they also allow to locate movements that affect the whole body, which are visualised by vertical stripes that go across all the marker rows. In the first stimulus, the most evident perturbation in the motion data of the expert can be clearly seen between sec. 23 and 25. The waveform aligned to the graphs shows that this general shift coincides with the peak the melody reaches at the beginning of bar 9, before concluding the phrase on the minim at the end of the same bar. A similar, albeit slightly delayed[8], general perturbation in the motion data can be observed in the neophyte around sec. 25. This is consistent with the data of the other participants. Fig. 5.7 shows the magnitude of the mean velocities of the upper body joints (labelled 1 to 8 in Fig. 5.5) of all the neophytes. Right hand and elbow are plotted separately, since they are involved in the main instrumental movement and therefore show the highest magnitudes. As it can be seen in the graph of the first stimulus, the mean velocities of all joints drop near sec. 25, and so do the values of the standard deviation. This confirms what observed above for the expert and one of the neophytes and suggests a general tendency to parse evident melodic phrases with overt body movements. As it can be noticed from the full Mocapgrams, this phenomenon occurs repeatedly during the neophyte's performance and the same trend is visible in the data of the other participants. This is consistent with findings in previous studies on air-performance showing that beginners tend to move more than experts (Godøy et al., 2005).

The second stimulus used is an excerpt from Kaija Saariaho's Nocturne for solo violin (Fig. 5.2). Compared to the first stimulus, the pulse of the piece is less steady and distinct. Each bar begins with a left hand pizzicato ('+' sign in the score) and continues with a glissando (or with two extra-metric groupings in the case of bar 48), which then leads to a tremolo. The downbeat of each bar is clearly punctuated by the pizzicato while the intermediate beats are more indistinct. In the mean velocities for the second stimulus depicted in Fig. 5.7, there are sharper peaks and valleys around the bar lines. Those are the points were the tremolo reaches its peak leading to the onset of the pizzicato gesture. The individual Mocapgrams of the data (shown in Fig. 5.8 for the expert and one of the neophytes) also show general perturbations around those key points, especially around

[8]The delay is plausibly due to the fact that the neophytes follow the audio recorded during the expert's performance, therefore their movements slightly lag behind the ones of the expert.

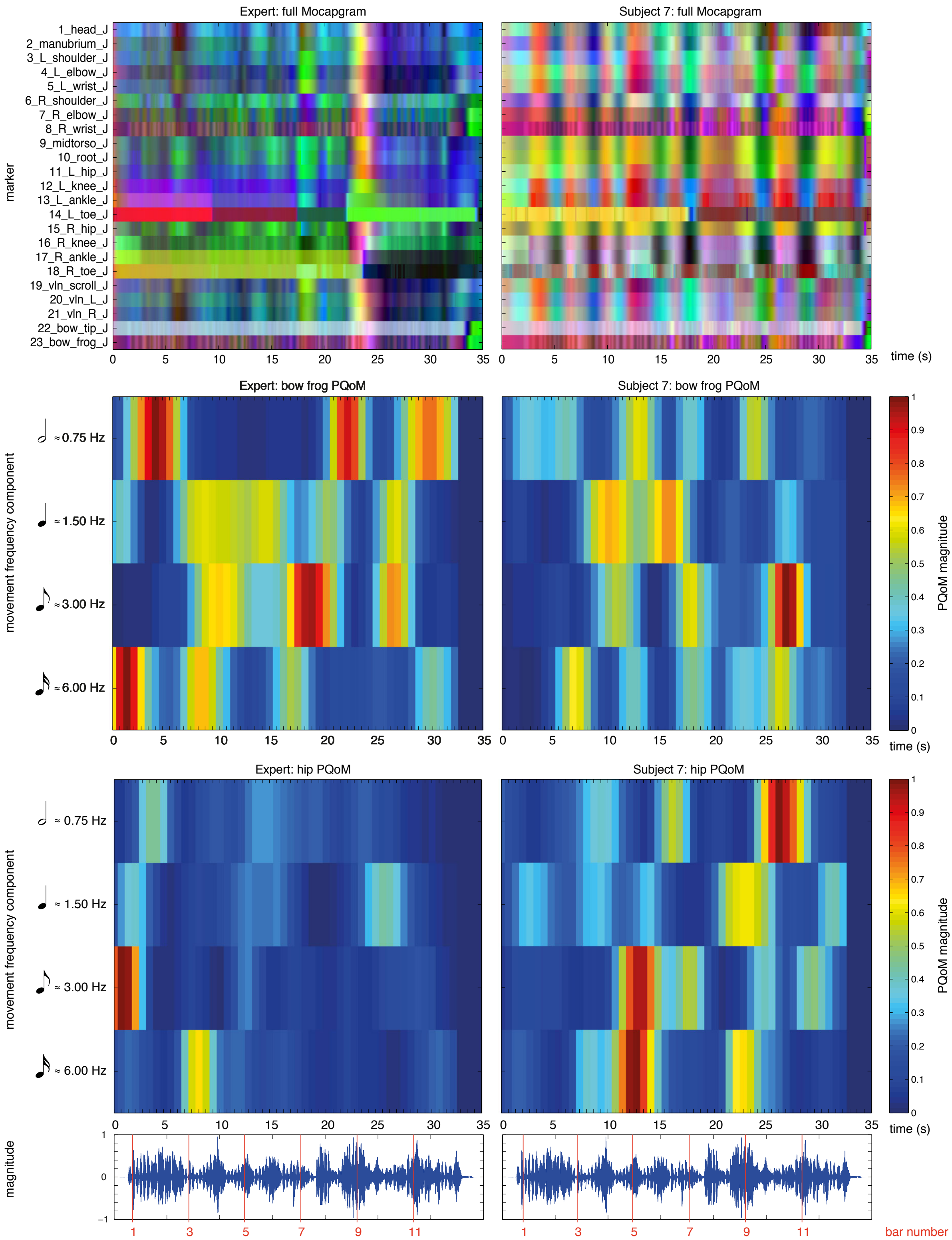


Fig. 5.6 First stimulus (Vivaldi, score in Fig. 5.1). Full Mocapgrams and Periodic Quantity of Motion (PQoM) estimates of the bow frog marker and the left hip joint for the expert violinist (left) and one of the neophytes (subject 7, right) aligned to the waveform of the audio.

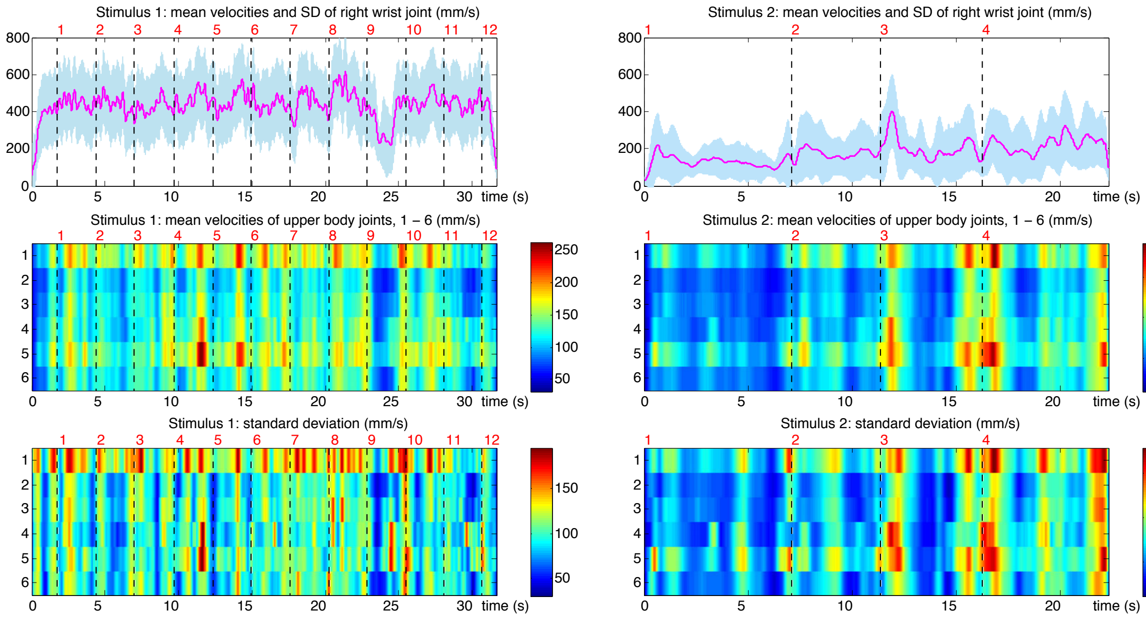


Fig. 5.7 Mean velocities of right wrist and upper body joints (labelled 1 to 6, see Fig. 5.5 for the location). The graphs on the left show the values for the first stimulus (Vivaldi, Fig. 5.1) while those on the right show the values for the second stimulus (Saariaho, Fig. 5.2). The red numbers and the vertical dashed lines indicate the bar number and downbeat location.

the downbeat of bar 3 and 4. Overall, there is ostensibly a clear intention among all the participants to interpret the musical gesture of the pizzicato with a sharp movement. Even though the location and extent of the movement may differ from subject to subject, those musical events are consistently mapped to similar gestural reactions.

5.3.3 Analysis of movement periodicity using Periodic Quantity of Motion

Another useful descriptor used for analysing movement periodicity is Periodic Quantity of Motion (PQoM). First introduced in (Visi et al., 2014b), this index gives an estimate of the resonance of the movement periodicity with different rhythmic subdivisions in the music. Inspired by the widely known Quantity of Motion (QoM) (Camurri et al., 2004b; Camurri and Volpe, 2011), PQoM is a motion descriptor useful to observe how movement relates to rhythmic aspects of the music. PQoM is calculated by subdividing the magnitude vector of the 3D motion data into frequency components by using filter banks (Müller, 2007). The frequencies of the filters correspond to multiples and subdivisions of the

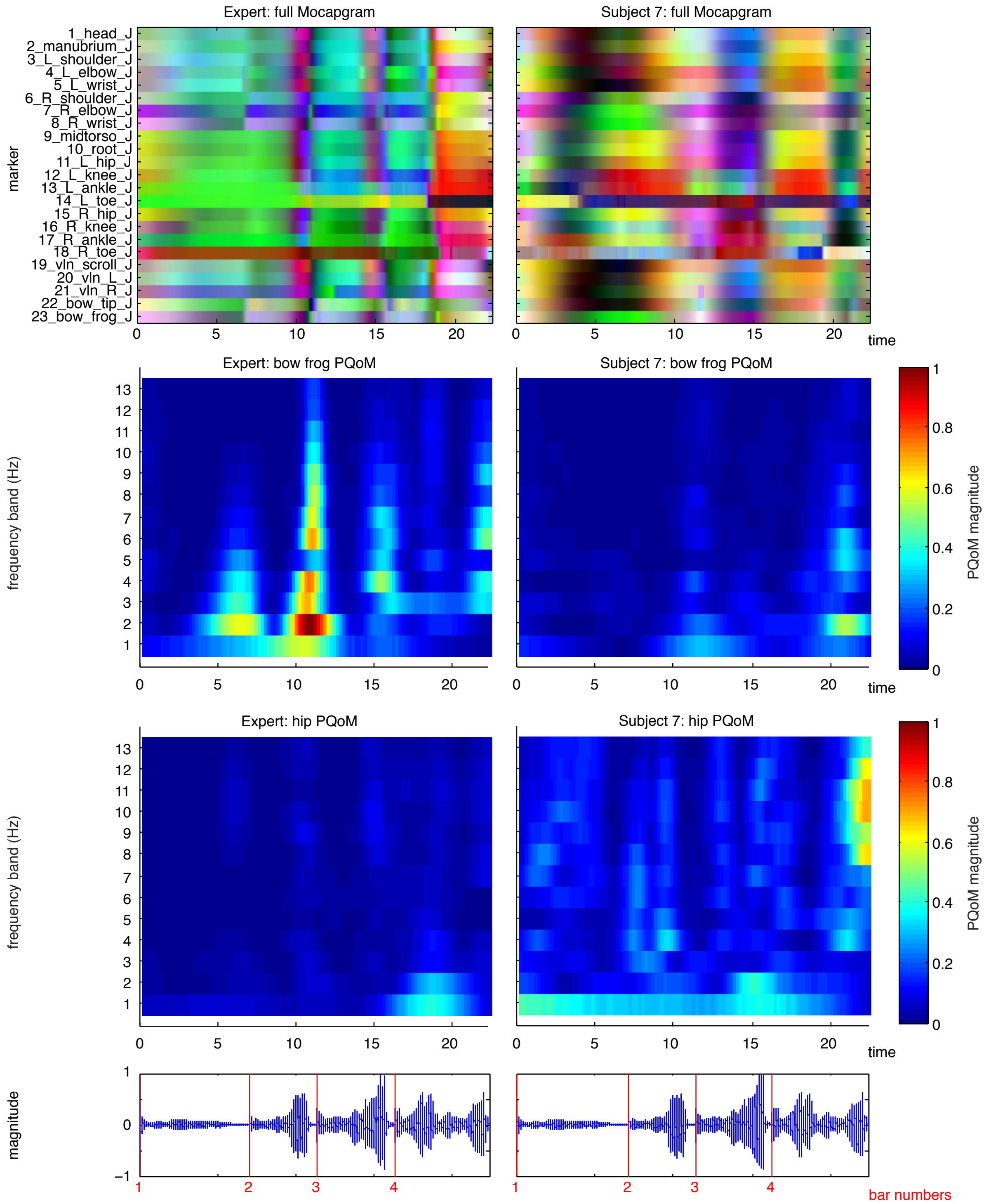


Fig. 5.8 Second stimulus (Saariaho, score in Fig. 5.2). Full Mocapgrams and Periodic Quantity of Motion (PQoM) estimates of the bow frog marker and left hip joint for the expert violinist (left) and one of the neophytes (subject 7, right) aligned to the waveform of the audio.

musical rhythm of the piece. The algorithm for computing PQoM is available as an extension of MoCap Toolbox[9].

The PQoM function uses three main input parameters: window size, tempo in BPM, and the centre frequency of each band-pass filter. These parameters are associated with the musical content. The window size parameter is used to define the amount of time that will be integrated to compute the quantity of motion for each frequency band. Once tempo is defined in BPM, the centre frequencies of each band pass filter can be specified in note values, as 𝅝, 𝅗𝅥, ♩, ♩. , ♪, . . . , following the rhythmic subdivision of the music. The PQoM function converts these note values into hertz and each frequency f_c is then used to compute the transfer function coefficients of a fourth-order bandpass digital Butterworth filter. Then, the input signal is filtered by a zero-phase forward and reverse digital IIR filtering algorithm (non-real-time implementation). Finally, similarly to QoM, the filtered values for each sample are summed over a time window of length N (window size) samples.

For optical motion capture data, each tracked point is normalised using the origin $[0, 0, 0]$ of the coordinate system as reference[10], and a weight vector $\boldsymbol{w}$ containing the weight coefficients for each tracked point. These coefficients determine the influence each tracked point has on the final PQoM value. It is worth noting that it is possible to ignore specific markers by setting the related coefficient in $\boldsymbol{w}$ to zero. Thus, PQoM can be calculated as follows:

$$PQoM[t, f_{c_k}] = \sum_{n=t-N}^{t} H_{f_{c_k}}\{(\boldsymbol{w}(|\boldsymbol{x}[n] - \boldsymbol{x}[n-1]|))\}, \tag{5.1}$$

where the $\boldsymbol{x}$ vector contains the Euclidean norms for each tracked point, and $H_{f_{c_k}}$ is the k^{th} bandpass filter operator.

In the case of the first stimulus (Fig. 5.1), a frequency of 1.5 Hz corresponds approximately to a steady crotchet beat, while 0.75 Hz correspond to a minim beat, 3 Hz to a quaver beat and 6 Hz to a semiquaver beat. The PQoM at a certain rhythmic subdivision is the magnitude of the corresponding frequency component in the movement normalised between 0 and 1. PQoM was estimated for four components until the end of the audio stimulus (sec. 33). By using PQoM, it is possible to relate the motion periodicities initially observed in the Mocapgrams to the actual rhythmic features of the musical stimulus. In Fig. 5.6, PQoM graphs of the bow frog marker and one of the hip markers are aligned to the Mocapgrams. These two markers were chosen as the former is a good

[9]MoCap Toolbox and the PQoM extension are freely available at https://www.jyu.fi/hum/laitokset/musiikki/en/research/coe/materials/mocaptoolbox

[10]In fact, it is possible to define any point as reference, [0,0,0] is the default option.

indicator of the main instrumental gesture (bowing), while the latter traces ancillary movements occurring in the lower half of the body during the performance. As shown in the study in chapter 4, movements in this area in some cases do not resonate evenly with the instrumental movements in the upper body. This is noticeable here as well after a first glance at the PQoM graphs in Fig. 5.6. Expectably, the bow frog PQoM of the expert is generally higher. However, expert and neophyte seem to follow similar patterns throughout their performances, with PQoM peaking around sec. 4 and 25 in the 0.75 Hz component, between sec. 6 and 16 in the 1.5 Hz component and at sec. 26 in the 3 Hz one. However, the two hip PQoM graphs look very different from each other, sharing only a relative peak in the 1.5 Hz component around the minim that closes the phrase in bar 9. In fact, the neophyte's hip PQoM graph shows that entrainment is remarkably more frequent and intense in that area than in the expert's. Moreover, by aligning PQoM graphs with Mocapgrams, it is possible to add further details to previous observations. In correspondence with the end of the phrase described previously (bar 9, sec. 23–25), there is a sudden shift of the PQoM index from a peak in the minim beat frequency to a peak in the quaver beat frequency after sec. 25. This salient turning point in the melody is therefore consistently reflected in the movement of both subjects, denoting a shared, embodied knowledge of the expressive qualities of the music, which they express through their instrumental gestures regardless of their expertise with the instrument.

As noted in the previous section, compared to the first stimulus, the second stimulus (Fig. 5.2) lacks of the steady and evident *allegro* pulse. The excerpt used has a much slower pace with expressive variations (the tempo indication reads "*Sempre espressivo, calmo* (♩ = c.54)"). This is evident when looking at the red bar lines plotted on the waveforms in Fig. 5.8, which denote bars of slightly different durations. Therefore, the PQoM for several different frequency components (from 1 Hz to 13 Hz with a step of 1 Hz) was computed. The results show, expectably, an overall less pronounced movement periodicity compared to the first stimulus. However, there are notable peaks in the PQoM of the expert's bow frog, especially in the 2 Hz and 4 Hz frequency components. These peaks occur right before the downbeat of each bar, in correspondence with the tremolo notes that lead to the pizzicato notes. Notice the alignment of these peaks of periodicity with the vertical stripes denoting movement across the whole body in the expert's Mocapgram at sec. 10 and 15. These peaks in PQoM values are ostensibly due to the faster, repetitive bow strokes necessary to perform the tremolos. A peak of periodicity, although weaker, can be observed also in the neophyte's bow frog marker, towards the end of bars 2 and 4 (approximately between sec. 10 and 12 and after sec. 20). Similarly to the previous stimulus, the bow frog PQoM of the expert is generally higher. However, there are relative peaks around the same areas, in correspondence with the tremolos.

The expert's hip marker shows much weaker periodicity. Apart from some barely noticeable peaks near in the areas leading to the downbeats of each bar (which may denote some modest resonance with the periodicity of the bow frog during the tremolo gestures), the PQoM values slightly increase in the low frequency component in bar 4. On the other hand, the PQoM values of the neophyte's hip marker are noticeably higher throughout, confirming the tendency observed on the first stimulus. There is, again, more entrainment in the ancillary movements of the neophyte. This supports the assumptions made in section 5.3.2, and is yet again consistent with the findings of previous studies (Godøy et al., 2005).

5.3.4 Periodic Quantity of Motion Correlations

In order to have a detailed overview of the significant correlations between the PQoM values of the expert and those of the neophyte, the Pearson correlation coefficients for all the PQoM data were computed.

As expected considering the observations made in section 5.3.3, there is little significant correlation between the PQoM values of the second stimulus (Saariaho, score in Fig. 5.2). This is likely due to the lack of a steady rhythmic pulse in the second musical excerpt. However, the PQoM values of the first stimulus (Vivaldi, score in Fig. 5.1) – which has a steady *allegro* pulse – show much higher correlation coefficients. The correlation matrices in Fig. 5.9 display the Pearson correlation coefficients for the four PQoM frequency components for the bow frog marker of all the participants, including the expert (row/column #1). The graphs clearly suggest that there is a very strong correlation between the PQoM values of the quarter note frequency component (1.50 Hz, top right matrix). Row and column 1 are of particular interest, since they display high correlation coefficients between the expert's PQoM and those of all the neophytes. To further investigate the hypothesis of high correlation between the expert and the neophytes' values, I tested the statistical significance of the correlation results using a significance threshold of $p = 0.05$. Table 5.1 reports the values of the correlation coefficients with the significant ones ($p < 0.05$) marked in bold. The values in the column for the 1.50 Hz frequency component (which correspond to a quarter note rhythm) are all positive and with the relative p-values below the threshold of significance.

Regarding the left hip marker, the correlation matrices in Fig. 5.10 show much less consistent correlation among the participants. This is compatible with the observations made in the previous sections, which showed noticeable differences between the expert and the neophyte. This is confirmed by the values shown in Table 5.2: the correlation coefficients are both positive and negative with several respective p-values above the threshold of significance.

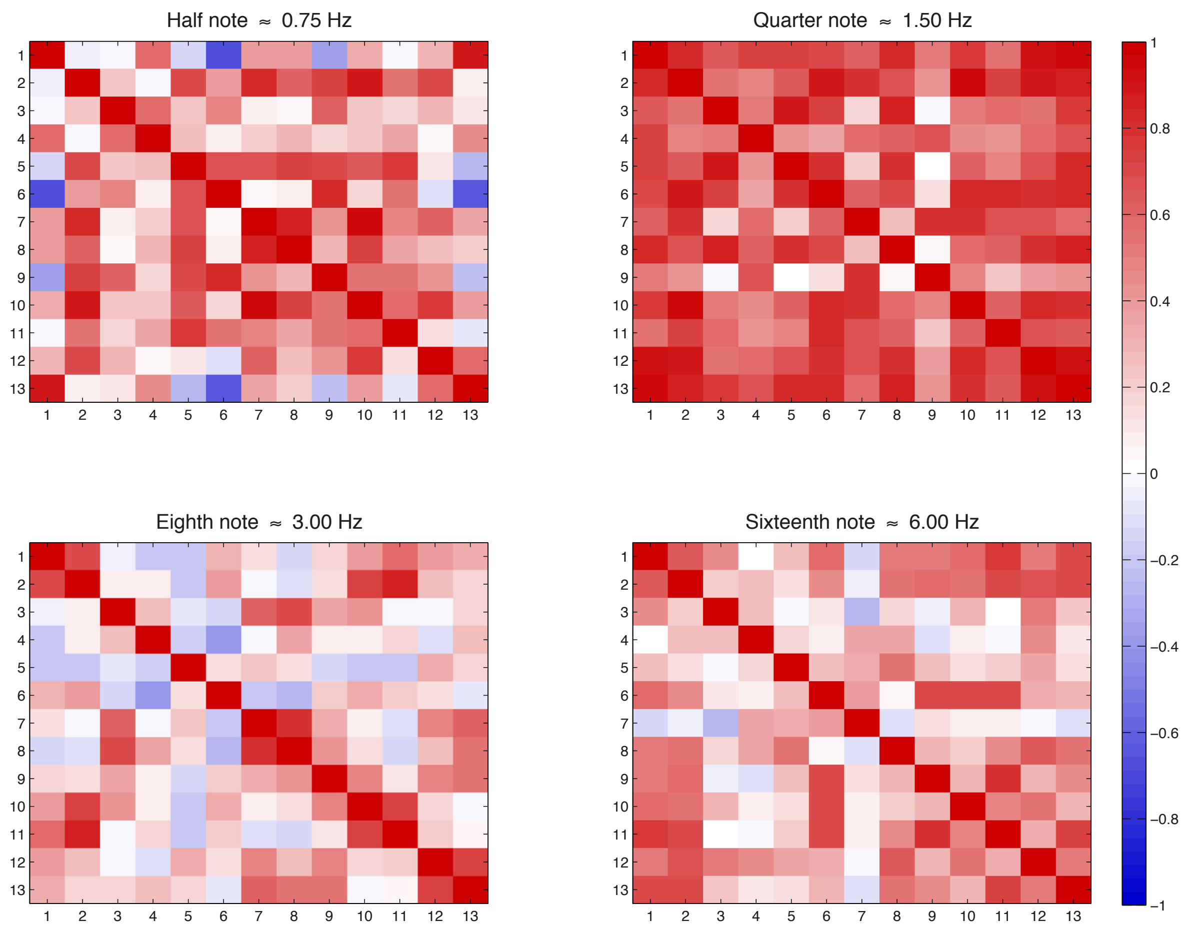


Fig. 5.9 First stimulus (Vivaldi, score in Fig. 5.1), **bow frog marker**. PQoM correlation matrix for each frequency component. Row/column 1 display the correlation coefficients for the expert violinist, 2 through 13 for the neophytes.

5.3.5 Results

The preliminary comparative analysis of the motion data of the first stimulus suggests that high-level, structural features of the music are expressed through instrumental movements in similar ways by the subjects, regardless of their ability to play violin. In particular, the turning point of the melodic phrase at bar 9 seems to be something that is *'felt'* by the subjects also in a strongly embodied way, as it impacts the movement of the whole body and the periodicity of the instrumental movements, which shifts sharply

Table 5.1 First stimulus (Vivaldi, score in Fig. 5.1), **violin frog joint**. Results of the correlation between the PQoM of each neophyte and the PQoM of the expert (subject #1 in the correlation matrices in 5.9). Significant correlations ($p < 0.05$) are marked in bold.

Subject #	0.75 Hz	1.50 Hz	3.00 Hz	6.00 Hz
2	-0.04	**0.81**	**0.71**	**0.63**
3	-0.01	**0.63**	-0.05	**0.46**
4	**0.58**	**0.73**	**-0.19**	0.01
5	-0.14	**0.72**	**-0.19**	**0.28**
6	**-0.66**	**0.71**	**0.31**	**0.57**
7	**0.38**	**0.60**	0.14	-0.13
8	**0.39**	**0.81**	-0.13	**0.52**
9	**-0.36**	**0.52**	0.16	**0.51**
10	**0.33**	**0.77**	**0.38**	**0.57**
11	-0.02	**0.55**	**0.57**	**0.77**
12	**0.31**	**0.91**	**0.40**	**0.51**
13	**0.90**	**0.94**	**0.32**	**0.71**

Table 5.2 First stimulus (Vivaldi, score in Fig. 5.1), **left hip joint**. Results of the correlation between the PQoM of each neophyte and the PQoM of the expert (subject #1 in the correlation matrices in 5.9). Significant correlations ($p < 0.05$) are marked in bold.

Subject #	0.75 Hz	1.50 Hz	3.00 Hz	6.00 Hz
2	**0.58**	**0.42**	**-0.50**	-0.15
3	**0.66**	**0.55**	**0.51**	**-0.24**
4	0.05	**-0.76**	**0.48**	**-0.37**
5	**0.50**	0.02	0.17	**0.57**
6	**0.64**	0.14	-0.16	**-0.34**
7	0.03	**-0.35**	0.09	-0.16
8	0.17	**0.78**	**0.26**	**0.41**
9	-0.01	**0.49**	**0.19**	0.14
10	**0.84**	**0.57**	**0.33**	**0.52**
11	**-0.22**	**0.68**	**0.57**	-0.02
12	**0.60**	-0.04	**0.53**	**0.23**
13	**0.27**	**-0.21**	**0.31**	**0.70**

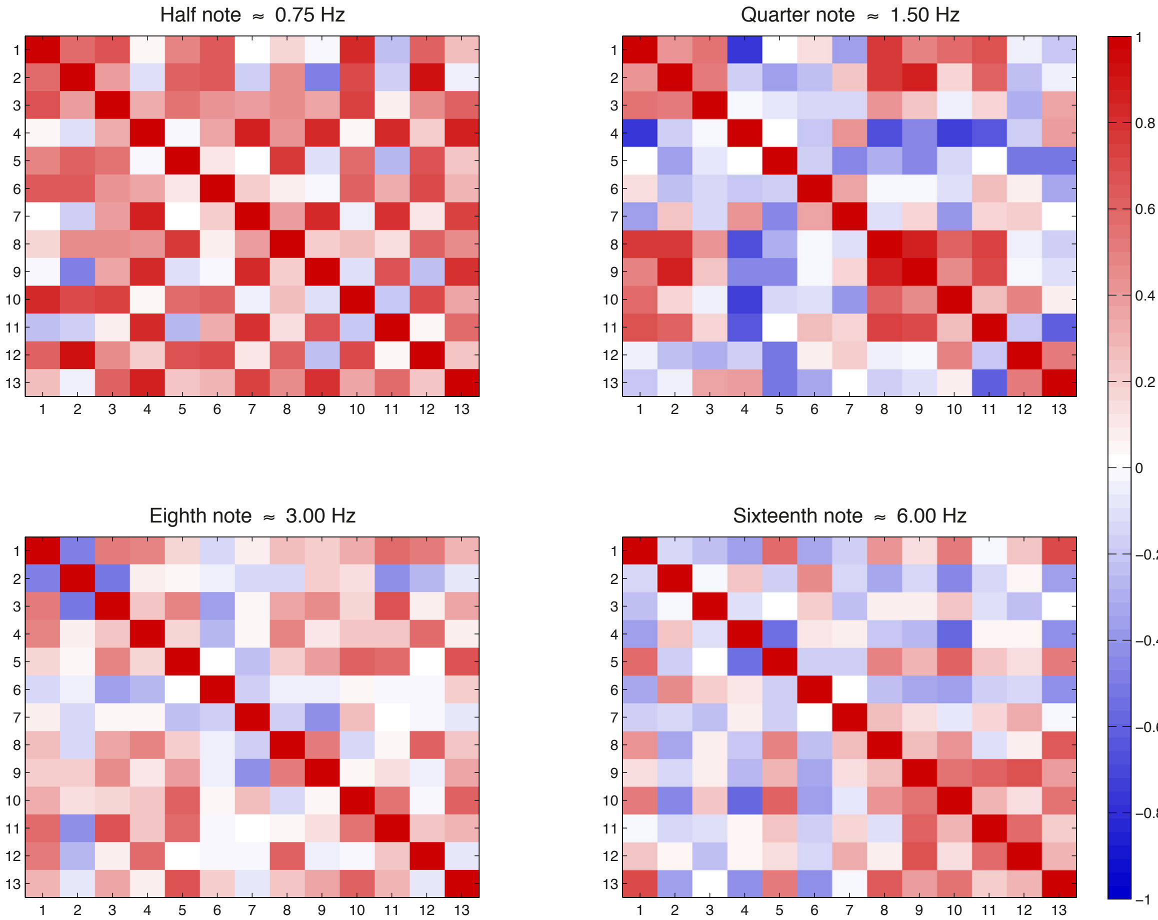


Fig. 5.10 First stimulus (Vivaldi, score in Fig. 5.1), **left hip joint**. PQoM correlation matrix for each frequency component. Row/column 1 display the correlation coefficients for the expert violinist, 2 through 13 for the neophytes.

from a frequency to the other. In addition to that, after the minim that closes the phrase there is a peak in the 3 Hz PQoM of the expert and an even higher one in the neophyte. This may suggest that the suspension created by a longer note ending a phrase creates a stronger expectation for the following melodic part, with which the neophyte engages also through ancillary movements, as shown by the hip PQoM graph. In fact, all the neophytes seem to have a more pronounced full-body periodicity. This can be hypothesised by simply looking at the vertical stripes in the Mocapgrams. However, PQoM gives a much more precise estimate of the periodicity in relation to the musical rhythm. Nearly all the neophytes seem to have a generally higher resonance with the periodicity of the music at the hip compared to the expert, whose PQoM is instead higher at the bow frog. This can be observed in both stimuli, and may lead to hypothesise that neophytes tend to follow the pulse and the features of the music with ancillary movements also to compensate for the lack of expressivity of their silent instrument. Doing so, they express the musical content of the stimulus using their full bodies.

Compatibly, the data of the second stimulus (even though musically very different from the first one) also suggest the major structural features of the musical excerpt are expressed through instrumental movements in similar ways. The second stimulus shows expectably less entrainment due to its slower and more implicit rhythmic features. However, the bar subdivision delineated by the rising tremolos and the pizzicatos are clearly expressed by the neophytes through instrumental and body movements.

More general results from the correlation and statistical tests show strong correlation between the periodicity of instrumental movements related to the musical rhythm of the first stimulus. Particularly relevant is the significant correlation between the expert and each neophyte for the quarter note PQoM of the bow frog (as shown in the second column of Table 5.1). This suggests that this music-related motion feature is part of the embodied knowledge shared by the participants.

On the other hand, the ancillary movements of the hips appear more individual and idiosyncratic, showing positive correlation between some of the participants, especially in the half note frequency component. This could be due to periodic weight shifting, which has been previously observed in performances with other musical instruments (Wanderley et al., 2005). However, this hypothesis would have to be further verified in a dedicated study that also takes in consideration the motion data of other parts of the body.

5.4 Analysis of Individualities and Commonalities

Our interaction with music engages the whole body, but not all body parts show the same behaviour (Burger et al., 2013c; Visi et al., 2014a). This analysis is focused on the movements of the head and the right wrist. In fact, previous research has shown that string players communicate expressive qualities of the music through head movements (Coorevits et al., 2014; Glowinski et al., 2014b). In addition, the movements made by the right wrist are also addressed, since it is a body part that is directly involved in sound-producing gestures.

In this analysis, the performances of all the 12 neophytes are taken into consideration. The first derivative (velocity) was calculated from the motion data of each subject using a Savitsky-Golay smoothing filter with a regression window of 7 frames and the resulting signals were set equal to the norm of the derivatives. Secondly, the speed envelope was calculated using a moving average filter of 100 frames in the case of the Vivaldi and 150 frames in case of the Saariaho to make sure the beat of the music (1.5 Hz for the

Vivaldi, 0.7 Hz for the Saariaho) was covered by the window and, at the same time, avoid losing too many of the nuances of the movement. The speeds of the body movements are then compared, as this feature is closely related to kinetic energy (Dempster and Gaughran, 1967). To check if the data was normally distributed, a Weibull function was fitted to the distribution of the speed values across subjects at all moments in time. The mean speed signal of both head and wrist at each timestamp over participants was approximately normally distributed, corresponding to a shape parameter of the fitted Weibull distribution between 2.7 and 3.2 for both pieces.

5.4.1 Modelling head and wrist movements

The method for the analysis of expressiveness proposed by Amelynck et al. (Amelynck et al., 2014) is based on Functional Principal Component Analysis (FPCA). FPCA allows to describe a signal as the sum of an average signal $\bar{f}(t)$ with a linear combination of a set of eigenfunctions $\xi_k(t)$ (commonality). Each subject can then be represented by one score (α_i) per eigenfunction (individuality):

$$f_i(t) = \bar{f}(t) + \sum_{k=1}^{K} \alpha_{ik}\ \xi_k(t)\ . \tag{5.2}$$

This way, the dimensionality of the problem is reduced and as much variance as possible is covered by only a small set of eigenfunctions. According to this method, the set of eigenfunctions should explain at least 70% of the variance. For our modelling, a correlation matrix based on the speed envelope of all subjects over time $C(t1, t2)$ is used as an input. An additional assumption for using FPCA is that there is a relationship between values in C that are only a few samples apart. Therefore, the data is decomposed in a set of Cubic B-spline basis functions. To determine a reliable number of basis functions, the Mean Squared Error between model and signal was calculated. For both the head and wrist, and in both stimuli, the number of basis functions could be set to 60. A set of eigenfunctions could then be calculated by means of FPCA, using a least square algorithm. As the human body shows complex behaviour, varimax rotation of the functional principal component axes was applied to calculate a basis of eigenfunctions that most economically represent each individual by a linear combination of only a few basis functions. FPCA was performed using Ramsay's FDA toolbox for MATLAB. His approach (Ramsay, 2006) was followed throughout the procedure.

5.4.2 Results

To cover more than 70% of the variability of the head in the performance of the Vivaldi excerpt, up to three eigenfunctions are needed, totalling 87% of the variability and accounting for 40%, 28% and 19% respectively (Fig. 5.11). An equal amount of eigenfunctions is needed for the wrist as 84% of the variability is covered with 40%, 15% and 29%. This means that, with only three eigenfunctions, we can model more than 80% of the commonalities in the head and wrist movements of the neophytes miming a violin performance following the musical stimulus. The individuality of each subject's performance was obtained by calculating the Functional Principal Component Score for the three eigenvalues. The individual performance can hence be modelled by three values indicating a positive or negative score for each eigenfunction. As few individualities were required for the model, this suggests that music was embodied in similar ways among the subjects.

For the head, the first eigenfunction has a positive deviation from the group's mean for almost the entire stimulus. This means that subjects with a positive factor on this function will perform with higher speed than the average, nearly throughout the whole recording. In more detail, this eigenfunction reveals something about the periodic movement and phase of the head. Subjects scoring low on this eigenfunction will have low velocity in the beginning of the bar, and higher velocity in the middle (bars 1, 3, 4, 7, 9, 10, 11, 12), while subjects scoring high will have their velocity peak in the beginning of each bar. In the middle of bar 7, this is reversed and bar 8 and 9 have an opposite velocity profile. Note that this is the moment where the repeated note sequences end and new musical material starts. In the beginning of bar 6, the eigenfunction values are close to the mean. The second eigenfunction has a major positive deviation from the mean in bars 5, 6, and 7, the second half of bar 8 and bar 9, and the last two bars. This is complementary to the first eigenfunction. The third eigenfunction has a major negative deviation, especially in the first 4 bars and bar 8.

The first eigenfunction of the right wrist has a major positive deviation in bars 2-3, 6-7 and 10-12 and the second eigenfunction accounts for a positive offset in bars 1 and 8 in particular. Again, the third eigenfunction has a negative deviation from the group's mean, especially covering the variability in bars 3-6, and 8-10. Fig. 5.12 shows the individualities for the head and wrist, clustered using k-means clustering. The number of clusters was set to 5 for the head and 4 for the wrist, after considering the optimal number with k-fold cross validation. These three variables are the principal component scores, or weights for the eigenscores, which represent the performance of the individual subject.

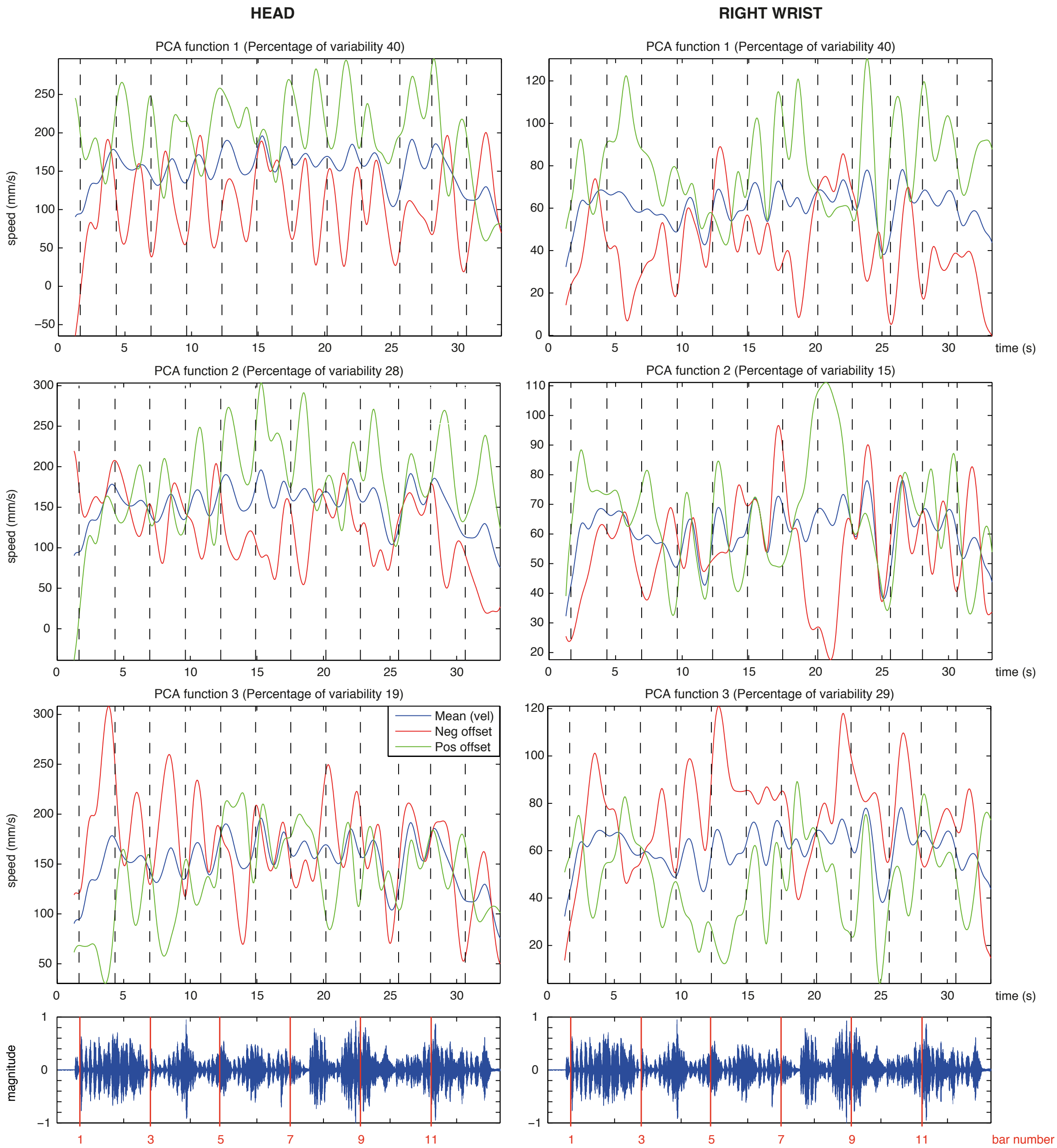


Fig. 5.11 First stimulus (Vivaldi, score in Fig. 5.1). Eigenfunctions for the speed envelope of head and right wrist movements after varimax rotation. The green line indicates a positive offset from the group's mean, the red line a negative offset.

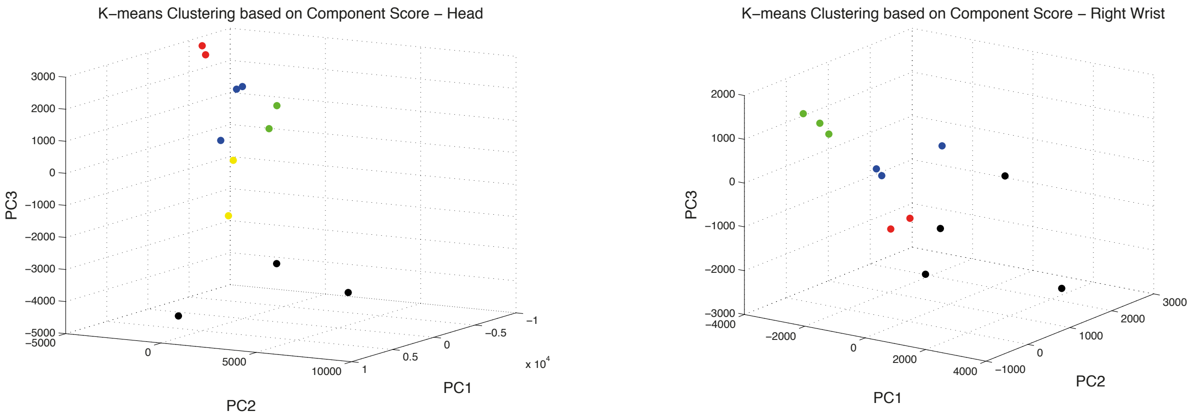


Fig. 5.12 First stimulus (Vivaldi, score in Fig. 5.1). Individualities for head and right wrist movements, clustered using k-means clustering.

Some intervals of coherence (i.e. time intervals of equal signs) could be derived from these results. When an eigenfunction has multiple intervals of coherence it can be considered consistent. For the head, the first eigenfunction shows this behaviour from the middle of bar 4 until the end of bar 5, as well as from bar 7 until the end of bar 10. The second eigenfunction covers this for bars 5-7, 9, 11 and 12. The third eigenfunction does not show long intervals of equal signs, except for the first bar. Coherence for the right wrist movement is found in the first eigenfunction from bars 2-4, the middle of bars 5-7 and 11-12. The second eigenfunction reveals coherence in the first bar and from the middle of bar 7 until the end of bar 8. From bar 3 until the beginning of bar 7 and bars 8-10 are coherent in the third eigenfunction. Thus, each eigenfunction dominates specific time intervals in the musical structure and they are mostly complementary to each other. The third eigenfunction of the wrist, for example, nicely reflects the repeated notes in the music (bars 3-7) and the new musical material introduced in bars 8-9 and 10. A similar effect can be seen in the second eigenfunction of the head. The last two bars of the musical stimulus (bars 11-12) are also represented in two eigenfunctions (the second eigenfunction of the head and the first of the wrist).

For the second stimulus, two eigenfunctions are sufficient to explain 70% of the variability of the head, and with three eigenfunctions, even 88% of the variability is covered. For the wrist, three eigenfunctions (47%, 18%, and 21% resp.) are sufficient as well. Again, this means that we can model the behaviour of the neophytes with a limited number of eigenfunctions, pointing at similar embodied behaviour (Fig. 5.13).

The first two eigenfunctions considered in the head are very consistent, with the first eigenfunction showing increasing positive and decreasing negative offsets towards the end of the excerpt. The second interval shows very distinct intervals of coherence that correspond to the intervals of the bar. Participants with a positive eigenvalue for this function in general will have a lower velocity within each bar and an increase in velocity towards the next bar, while the participants with a negative eigenvalue show opposite behaviour. The points of convergence at the bar transitions show that the second eigenfunction mostly accounts for individual behaviour within each bar.

The third eigenfunction reveals different timing behaviour of the participants, with a negative offset meaning an earlier peak than a positive offset. This was validated using cross-correlation with the maximal correlation at -0.296 seconds, showing an interesting parallel with the typical reaction time of people in tapping tasks, which generally lies between 200 - 300 ms (Bååth, 2015). It could be that the difference between participants in how well they predict what will happen in the stimulus, is expressed with this eigenfunction.

The first eigenfunction of the wrist movement in the Saariaho excerpt shows a general positive offset. Moreover, it can be clearly observed that the difference in musical material in the fourth bar is reflected in the movement of the wrist: there is a point of convergence in the middle of the fourth bar, while this does not appear in the other three bars where pitch changes appear mostly in the beginning of the bar. The second and third eigenfunction of the wrist are very complementary in the case of the Saariaho piece, where the second eigenfunction accounts for the variability in the beginning of the first bar and the transitions between bars 2-3 and 3-4 and the last part of the fourth bar, while the third eigenfunction accounts for the variability in the middle of these bars. This last eigenfunction is very similar to the second eigenfunction of the head, though less pronounced.

In general, it could be observed that the head is a body part that shows very clear intervals of coherence that match well with the timing structure of the piece (whole phrase, bars and reaction times of participants).

5.5 Summary and Comments

This chapter presented a study aimed at empirically exploring some of the theoretical assumptions presented in chapters 2 and 3. A group of people with no experience playing the violin were asked to mime a performance following the playback of some short musical excerpts recorded by an experienced violinist. To do so they used a silent violin specifically designed for the experiment. Motion capture data was recorded during every session.

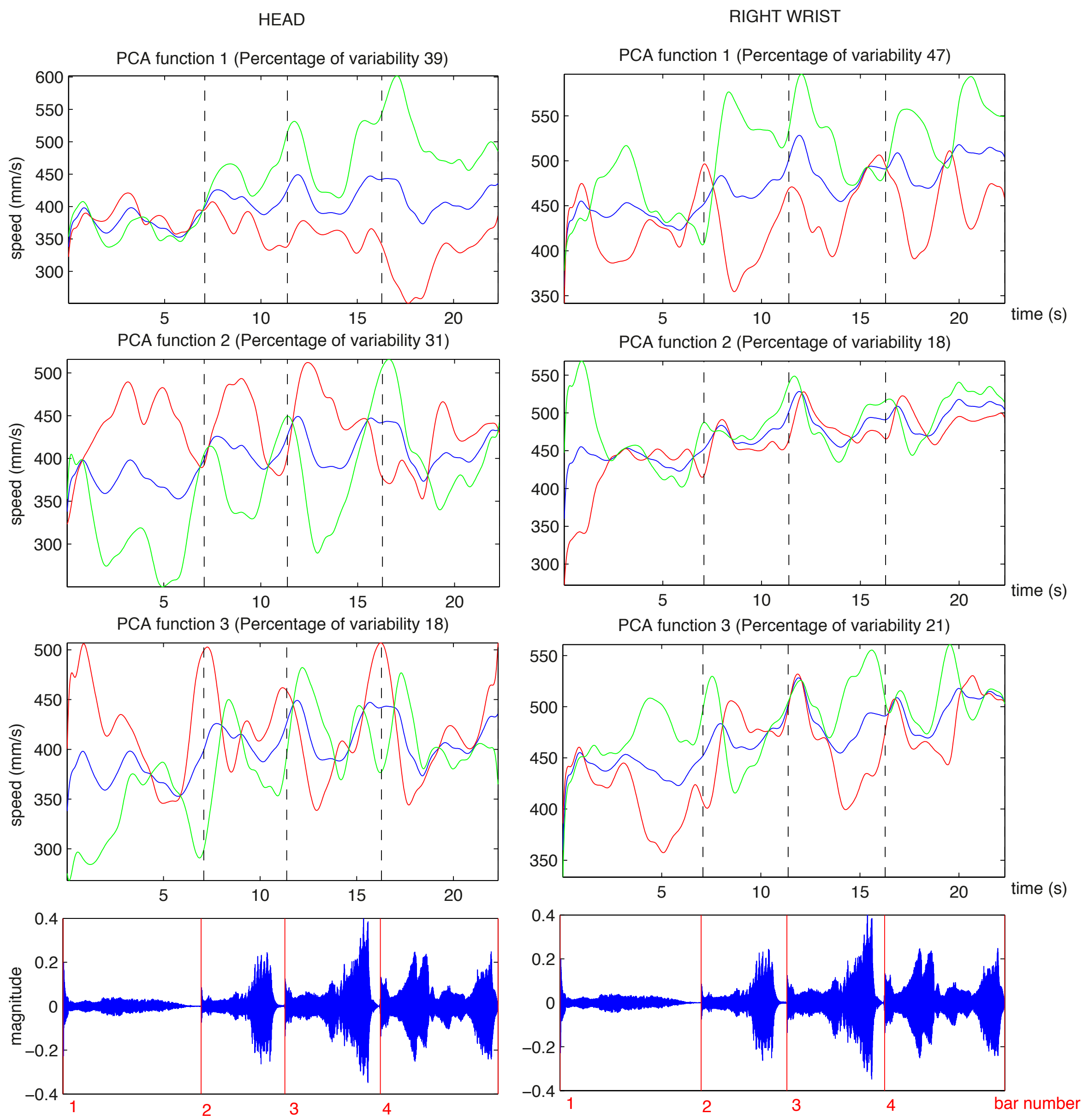


Fig. 5.13 Second stimulus (Saariaho, score in Fig. 5.2). Eigenfunctions for the speed envelope of head and right wrist movements after varimax rotation. The green line indicates a positive offset from the group's mean, the red line a negative offset.

Even though low-level features of movement appear to vary considerably in some of the subjects, there is a certain degree of consistency among participants, especially in response to melodic, rhythmic, and timbral features of the music. This suggests a shared knowledge of a vocabulary of instrumental movements, which is then combined with the idiosyncrasies of each subject. The analysis of commonalities and individualities confirms this, and other studies (MacRitchie et al., 2013) support the idea that musical structure is communicated also through body movements, and idiosyncrasies contribute to express musical meaning.

New approaches to movement analysis are in continuous development and there is an increasing need for tools that can aid the retrieval of meaningful features in complex, multidimensional motion data. Therefore, other approaches – like Topological Gesture Analysis (TGA) (Naveda and Leman, 2010) – can possibly be employed in future analysis works alongside the methods presented here. New techniques for motion data analysis could be inspired by concepts suggested by theories of music perception and cognition, therefore making the analysis more akin to how humans perceive and move to music. This is indeed a challenging task since retrieving meaningful, articulated information from motion data requires complex algorithms and technologies.

Motion data analysis has provided great detail for understanding the role of body movement in musical expression and cognition. However, it is felt that integrating quantitative data analysis with qualitative analysis and practice-based research may broaden the scope of the research, allowing to test the assumptions made through the analysis in musical contexts, outside of the sterile environment of the laboratory.

Part III

Using Movement in Music: Practice

Chapter 6

Augmenting Traditional Musical Instruments: a Multimodal Embodied Approach

> At a certain point, there was an attempt to take this biological fact, these ways of mapping the process within the body, as a way of describing a larger aesthetic system.
>
> MATTHEW BARNEY
> *Drawing Restraint Vol I 1987 – 2002*

6.1 Introduction

This chapter[1] describes the implementation of gestural mapping strategies for performance with a traditional musical instrument, multimodal motion sensors, and live electronics. The approach adopted is informed by concepts of embodied music cognition and functional aspects of musical gestures (see chapters 2 and 3). Within this framework, gestures are not seen as means of control subordinated to the resulting musical sounds but rather as significant elements contributing to the formation of musical meaning similar to auditory features. Moreover, the ecological knowledge of the gestural repertoire of the instrument is taken into account as it defines the action-sound relationships between the

[1] This chapter is based on Visi et al. (2014b). The full peer-reviewed article can be retrieved online at http://www.federicovisi.com/publications/.

instrument and the performer, and contributes to form expectations in the perceivers. Subsequently, mapping strategies from a case study involving electric guitar will be illustrated, describing what motivated the choice of a multimodal motion capture system and how different solutions have been adopted considering gestural meaning formation as well as technical constraints.

6.1.1 Mapping

In the past two decades motion sensing technology have become more accessible and have been increasingly employed in academic research and artistic practice. More recently, such technologies have become pervasive, as many everyday electronic devices use motion data for different purposes and new developments made the recognition of motion features more precise and robust.

In musical contexts, there is a long and prolific tradition of electronic interfaces that exploit gestures and body motion as means of control of musical parameters (Jensenius and Lyons, 2016; Miranda and Wanderley, 2006). Adopting an effective strategy when designing the mappings between control signals and sound parameters is crucial for the expressiveness of the interface, being the relationship between motion features and musical parameters often far from obvious. Mapping has in fact received increasing academic interest (Wanderley and Malloch, 2014) and it is recognised as a critical element in instrument and interaction design (Hunt et al., 2003). Several mapping approaches have been adopted over the years and insights from artistic practice show that mapping is not solely an issue of interface and control, but also a part of the compositional process (Di Scipio, 2003; Murray-Browne et al., 2011).

6.2 The Electric Guitar as a Case Study

Given the background scenario described in chapters 2 and 3, it is clear that gestures have a significant influence on how music is experienced, and traditional musical instruments are a rich repository of shared gestural information. Therefore, the theoretical apparatus of embodied music cognition (EMC) could be employed to devise effective mapping strategies that may give a substantial contribution to both the expressiveness and the liveness of a performance involving TMIs and live electronics. In recent years, there have been applications of EMC within interactive multimedia environments (Camurri et al., 2001) and singing performance (Maes et al., 2011).

In this chapter, I describe the implementation of some gesture mapping strategies for a performance with electric guitar and electronics. There are other documented approaches to electric guitar augmentation using the performer's gesture (Lähdeoja et al., 2009) and gestural control of digital audio effects (Verfaille et al., 2006). Apart from aspects related to control of musical parameters, this approach emphasises the fact the gestures contribute to the formation of musical meaning, therefore the function of gestures and their relationship with musical features are taken into consideration throughout the implementation process.

The multimodal motion capture system adopted features sensors worn by the performer and a Microsoft Kinect. Flex sensors and accelerometers located on hands and wrists of the guitarist are employed to obtain accurate data of hand movements. In fact, sound-producing gestures of guitar playing – as with several other instruments – usually involve hands and arms, and such gestures are the most readily noticed by an observer (Dahl et al., 2010). Thus, obtaining more stable and detailed signals is useful to capture the subtle movements of hands and wrists. Conjointly, the Kinect data – even at relatively low frame rates (30 hz) or with jittery joint tracking – allows to track full-body movements, enabling the extraction of higher-level expressive feature.

6.3 Use of wearable sensors

Flex sensors and 3-axis accelerometers are mounted on custom wristbands together with a custom Arduino-based board equipped with an XBee wireless chip[2]. This sends the sensors' signal to a computer for signal processing and parameter mapping. The mapping strategies for these sensors are informed by the functional categories described by Jensenius et al. (2010). For example, the flex sensor on the left wrist is used to monitor the activity of this articulation. In guitar playing, movements of the left wrist act in support of the fingers operating on the fretboard. These movements are therefore defined as *sound-facilitating support gestures*. Stressing the wrist articulation may cause discomfort and alter the tone of the notes being played (Costalonga, 2009). To underline this, the flexion sensor is mapped to a bit reduction DSP algorithm that deteriorates the audio signal of the guitar, reflecting the uncomfortable stretching occurring on the left-hand (Fig. 6.1). The analogue signal of the sensor is converted to OSC data, which is then rescaled in a software dedicated to mapping[3] according to how the guitarist bends the wrist when playing comfortably or when stressing the joint. The OSC data is then

[2] http://www.sensestage.eu
[3] http://steim.org/product/junxion/

Fig. 6.1 Use of wearable sensors: flex sensors placed on the wrist activated by a sound-facilitating support gesture.

converted to MIDI and sent to a DAW[4], which hosts several DSP units that process the audio signal of the guitar. In this example, the MIDI data obtained from the flexion sensor controls the downsample resolution of a bit reduction DSP unit. (Fig. 6.1).

The accelerometers are instead used to follow the *communicative expressive gestures* (Jensenius et al., 2010) of the right arm that immediately follow strumming. These movements can also be considered suffixes of the strumming gesture and are important for its performance and perception (Godøy et al., 2010). To reinforce the meaning of the gesture, the accelerometer is mapped to a Max[5] patch that affects the timbre and decay of the strummed chord, following the intensity of the movement. The dynamic range of the accelerometer is relatively high, therefore the OSC data obtained from the sensor needs to be compressed and filtered through the same mapping software used above. (Fig. 6.2).

[4] `http://www.ableton.com`
[5] `http://cycling74.com`

Fig. 6.2 Use of wearable sensors: 3-axis accelerometer placed on the wrist activated by a communicative expressive gesture.

6.4 Use of full-body motion analysis

Whilst wearable sensors are used to follow the subtle gestures of the upper limbs, full-body motion analysis is employed to extract features from complex movements occurring over longer time intervals. One measure widely used by other authors (Camurri et al., 2005; Fenza et al., 2005) is Quantity of Motion (QoM). QoM is proportional to translational movement and it is extracted from a global set of features evaluated over time. It gives high values when the body is moving fast and low values when it is more stationary. Camurri et al. (2004b), for example, implemented this feature in the EyesWeb processing library. QoM is also useful for extracting contextual syntactic structures from musical performance (Lesaffre et al., 2003). QoM can be estimated from the skeleton joints tracked by a motion capture system (Fenza et al., 2005). In this case, the user can measure QoM for specific combinations of skeleton joints.

Initially, the QoM definition presented by Camurri et al. (2005) and also by Fenza et al. (2005) was adopted. Motion bells can be estimated in a simple but effective way by applying a low pass filter on the QoM estimates over time. After the filtering process, it is possible to identify the phases of the movement by analysing the shape of the curves. This can be seen in Fig. 6.3, which illustrates the motion bell generated by a sudden movement of the performer.

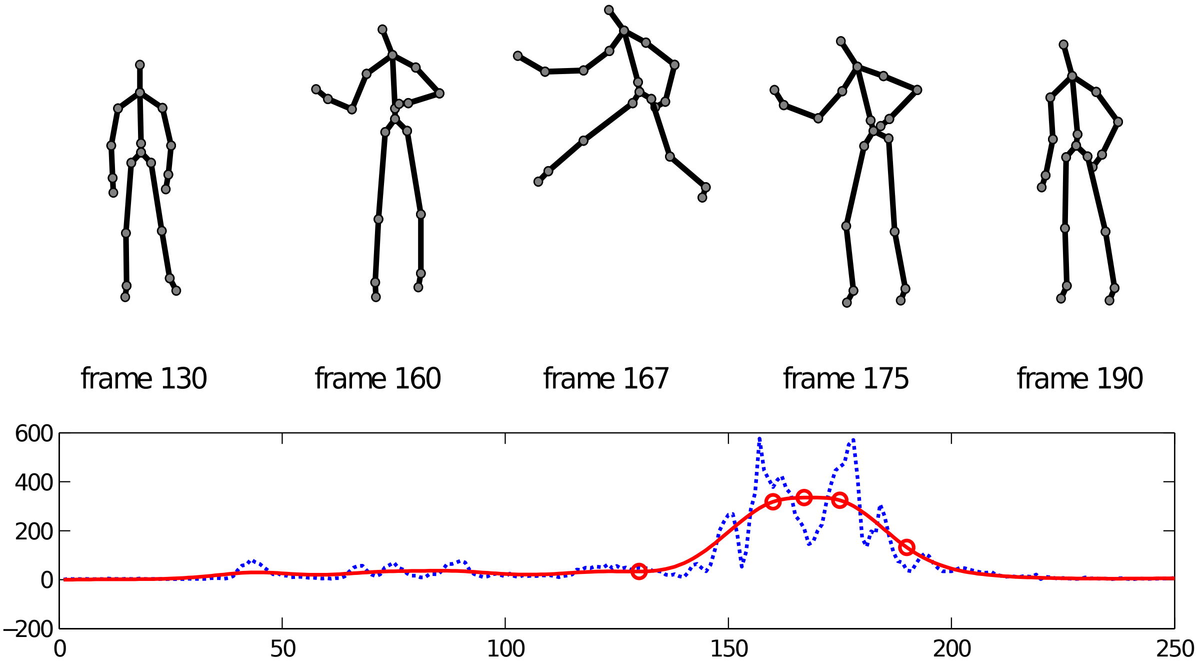


Fig. 6.3 The dashed line shows QoM values on a movement sequence. The solid line is the final measure after low pass filtering. The skeletons (on top) illustrate the skeleton position at each time stamp (circle marks).

The information obtained through this simple procedure is useful to segment the different motion phases. In fact, "gesture spotting" (Elmezain et al., 2009) is crucial for classification accuracy of complex gesture recognition algorithms based on Hidden Markov Models (Fink, 2008) or Dynamic Time Warping (Müller, 2007). Furthermore, offset and onset points can be used to control musical events associated with the respective gesture.

To find new ways of using QoM in musical contexts, we[6] devised a feature named *Periodic Quantity of Motion* (PQoM, see section 5.3.3) that describes the quantity of motion in relation to the periodicity of the movement. There is often sensorimotor synchronisation between the rhythmic structure of the piece and the periodic motion of the body (Repp and Su, 2013). PQoM allows to measure the resonance between rhythmic subdivisions of the music and body movement.

In our case study on electric guitar, the movement of the strumming hand often plays a periodic rhythm. For example, during a crescendo the periodic hand movement might be performed with greater amplitude and different body parts might also be engaged in periodic movements with the same frequency. The values describing this corporeal resonance with the music can then be mapped to sound parameters that alter the timbre

[6] Periodic Quantity of Motion was developed in collaboration with Rodrigo Schramm.

of the guitar, (such as gain, saturation, etc.). This idea can also be used with other instruments, such as bowed strings and percussion instruments.

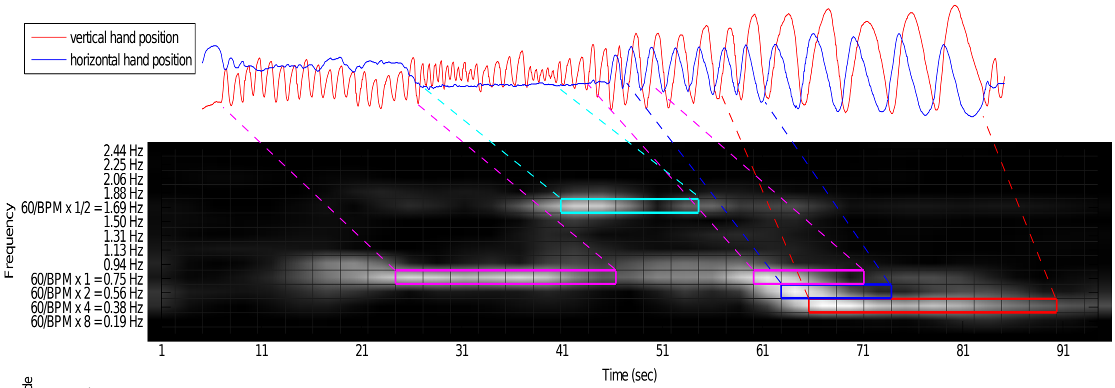
The PQoM estimate is obtained by decomposing the motion capture signal into frequency components by using filter banks (Müller, 2007). For each frequency, the process is similar to the computation of QoM, but it uses only the amplitude of that specific component. For example, the resonance of body movements with a certain rhythm might be used to add sampled percussive elements to the music, reinforcing the relationship between auditory and kinematic elements of the performance. Fig. 6.4 illustrates the extraction of such features from the body motion captured by the Kinect. In this example, the guitarist performs a periodic movement with the strumming hand. The hand oscillates with a frequency equal to multiples or fractions of the music tempo in BPM. In Fig. 6.4a, the lighter regions show where the movement resonates more with the rhythm. In the same figure, the bounding boxes represent the individual movements with distinct frequencies, detected using the threshold scheme described for the QoM approach. The light blue, pink, dark blue, red and green colours indicate movements corresponding to rhythms made of quavers, crotchets, minims, semibreves, and breves respectively. Finally, Fig. 6.4b shows the motions bells for each one of the expected frequencies, and the segments delimiting the individual movements.

The system can be further customised to follow other periodic features of the music and to recognise specific motion patterns. It can also be used to attune the music to the gestures of the performer, allowing for a two-way feedback between music and movement. This allows to explore interactions based upon corporeal resonance and entrainment, which are spontaneous phenomena that can be observed in listeners and performers (Dahl et al., 2010; Leman, 2008a).

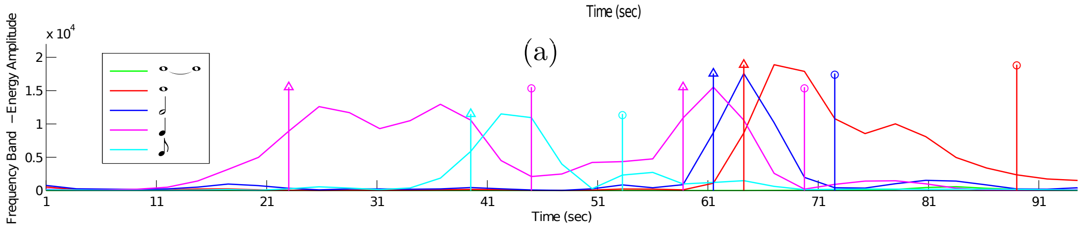
6.5 Summary and Comments

This chapter described an approach to the use of body motion features to augment traditional musical instrument performance. This approach is informed by studies of embodied music cognition and musical gestures and adopts a multimodal motion sensing system to extract various movement features. A novel feature named Periodic Quantity of Motion used to measure the resonance of body movement with musical rhythmic subdivisions is introduced.

Looking at music and gesture within the framework of embodied music cognition can radically influence the development of new expressive interaction tools. Gestures very often appear both as body movement schemes and mental representations, bridging



(a)



(b)

Fig. 6.4 (a) The lighter regions show where the movement resonates more with the rhythm. The light blue, pink, dark blue, red and green bounding boxes indicate movements corresponding to rhythms made of quavers, crotchets, minims, semibreves, and breves respectively. (b) The plotted lines (light blue, pink, dark blue, red and green) show the QoM related to each frequency response. The vertical stems show the starting point (triangle) and the ending point (circle) of each periodic motion.

body and mind (Leman, 2010). Considering gesture as an active constituent of embodied musical meaning implies that its role in an interactive music performance goes well beyond being a mere means of control of musical parameters. This interdisciplinary approach can inform different mapping strategies and technical solutions. First, by considering the function of different gestures in a performance to electronically modify the sound of the instrument played using sensors, and then by measuring the quantity of body motion of the performer in relation to musical rhythmic features. Working on different layers of gestural complexity not only allows for the development of more advanced systems but also reflects the multi-level nature of gesture within the mechanism of musical meaning formation. Gestures are in fact experienced as elements of a nested hierarchical structure (Leman, 2010) and taking this aspect into consideration can aid the design of expressive musical systems. Achieving the illusion of a total, fully conscious control is not the goal of this approach. Lähdeoja et al. (2009) explored non-direct control using semi-conscious gestures showing how these can be used to control subtle aspects of the music performed. This approach is also directed towards gestural aspects of music and aims nonetheless at exploring the tight relationship between music and body, on both a conscious and a sub-conscious level.

Focusing on a traditional musical instrument allowed to draw gestures from an existing, well-established gestural vocabulary. However, this does not mean that the approach cannot be extended to new digital instruments or employed to develop new interfaces. From this perspective, acoustic instruments are seen as rich repositories of gesture-sound couplings unmediated by mapping, which may be a useful resource to understand gestural aspects of music and, at the same time, find new ways of musical expression.

Embodied music cognition is the subject of an ongoing interdisciplinary research and new contributions to the understanding of important elements of its workings have been recently published (Kilner and Lemon, 2013; Maes et al., 2014; Matyja and Schiavio, 2013). Practice may lead to new intuitions, as it did in other contexts (Smith and Dean, 2009). New cross-disciplinary approaches may in fact help to “move beyond designing technical systems” (Waisvisz, 2006) and give rise to new engaging musical experiences that can both raise questions and provide new insights about musical expression and cognition.

Chapter 7

Instruments, Bodies, and Data: Music as an Emergent Multimodal Choreography

> RULE EIGHT: Don't try to create and analyze at the same time. They're different processes.
>
> Sister Corita Kent, popularised by Merce Cunningham and John Cage

7.1 Overview

This chapter[1] explores technical and conceptual issues related to the representation and mediation of body movement in music performance through digital technology. In particular, it focuses on IMU/MARG[2] wearable sensors. IMU/MARG data is compared to optical motion capture data and dedicated computable motion descriptors are proposed. The chapter then describes an implementation of machine learning algorithms for the use of IMU/MARG sensors in interactive music applications, reporting on how concepts related to the topology of data informed the mapping approach. The chapter also reports on a case study of a music performance where motion sensor technologies are employed

[1] This chapter is based on Visi et al. (2017). The full peer-reviewed article can be retrieved online at http://www.federicovisi.com/publications/.

[2] Inertial Measurement Unit / Magnetic, Angular Rate and Gravity. IMU/MARG sensors are sometimes marketed as 9DoF (9 Degrees of Freedom) sensors. More information about this technology and its naming in section 7.3.1.

to track the movements of the musicians while they play their instruments. Motion data is used to control the electronic parts of the piece in real time. In light of this case study, the chapter discusses how musical instruments can be seen as repositories of a gestural vocabulary and the score as a script that elicits an emerging choreography. Signification occurs through sound, physical movement, and digital movement data, while the body of the performer becomes the medium at the core of a layered process of multimodal meaning formation. Finally, it is suggested that computable motion descriptors and machine learning techniques are useful tools for interpreting motion data in a meaningful manner. However, qualitative insights on how human body movement is understood and experienced are necessary to inform further development of motion capture technologies for expressive purposes. Thus, music performance can be an effective test bed for new modalities of human-computer interaction.

7.2 Background Scenario

In the interdisciplinary field of musicological research, the idea that music is a multimodal phenomenon that engages body movement has given rise to a wide range of methodologies for the study of musical gestures. On the one hand, the advent of new technologies, such as infrared motion capture, has allowed researchers to observe human movement in detail, extracting precise three-dimensional data and kinematic features of bodily movement. This brought about a corpus of studies where motion analysis is based on the computation of several low-level descriptors – or movement features – that could be linked with musical expression (Godøy and Leman, 2010). Acceleration and velocity profiles have shown to be useful in the study of musical timing (Burger et al., 2014; Glowinski et al., 2013b; Goebl and Palmer, 2009; Luck and Sloboda, 2009). Quantity of Motion (QoM) has been related to expressiveness (Thompson, 2012) and features of the bass (Van Dyck et al., 2013), while contraction/expansion of the body can be used to estimate expressivity and emotional states (Camurri et al., 2003). More advanced statistical methods, such as Functional Principal Component Analysis and physical modelling have led to mid-level descriptors, including Topological Gesture Analysis (Naveda and Leman, 2010), curvature and shape (Desmet et al., 2012b; Maes et al., 2012), and commonalities and individualities in performance (Amelynck et al., 2014) (also described in chapter 5).

Gestures in music performance can also be described by means of high-level descriptors. Verbal descriptions, subjective experiences, and the musician's intentions play an important role in our daily interaction with music. This is the way performers and audiences naturally communicate about music. Leman (2008a) refers to these descriptions as "first person" perspectives on music experience, resulting in intention-based symbolic/linguistic expressions. In the analysis of music performance, this qualitative approach has been explored profoundly in the studies of, among others, Davidson (2007, 2012); King (2006); Williamon and Davidson (2002). Here, musical gestures are accessed by means of verbal descriptors, or directly perceivable movements that appear to be expressive. In that sense, the concept of musical gesture can be useful to bridge the gap between mental and subjective experiences of the performer/listener and the direct observable physical world. Recent studies have made an attempt to close the gap between these two perspectives in music performance research, by applying a performer informed analysis (Coorevits et al., 2015; Desmet et al., 2012b). In trying to understand the relationship between the physical aspects of movement in space and expressive qualities, intentions and feelings, the study of musical gestures has resulted in new understandings of the relationship between musician and musical instrument as well. The instrument then becomes a natural extension of the musician (Nijs et al., 2013), and hence a part the mediation process of communicating musical meaning. The development of human-computer interfaces also exploits the expressive potential of musical gestures to enhance the interaction between digital and human environments and to create meaningful applications for musical practice (Camurri et al., 2004c). Recently, artistic practice has also been increasingly adopted as a complementary research method for the arts and humanities, leading to mixed, interdisciplinary methodologies (Smith and Dean, 2009).

7.3 Capturing, Storing, and Mediating Movement

Human movement can be digitally captured and stored via different means, for purposes of analysis, description and notation. In the context of musicological studies, movement has been recorded throughout the years using visual media such as photography (Ortmann, 1929) and video (Davidson, 1993). More recently, motion capture has become widely adopted as the medium of choice for quantitative studies of human motion. Even though new technologies are emerging, marker-based optical motion capture is still regarded as the most reliable solution for precise, high-speed tracking. Data obtained from these systems is usually in the form of three-dimensional vectors referring to a global coordinate system. Each sample in the data returns three-dimensional information regarding the position of a point (marker) in space in relation to the origin of the Cartesian axes. The origin is defined during the calibration procedure and is usually set in an arbitrary place on the floor within the capture area.

As Salazar Sutil (2015) points out, the term motion capture (sometimes shortened to “MoCap”) indicates not only a technological setup but it is also used to refer to a “technologized language of movement, involving the formalized description of movement coordinates and movement data for its subsequent computational analysis and [...] processing.”. Compared to photography and film, MoCap is definitely a younger medium. This has obvious technological implications as MoCap technologies are still being developed and only recently have become more widely accessible to researchers and practitioners. However, compared to visual media, it is perhaps the nature of body movement itself that makes its mediation somehow still conceptually challenging. Salazar Sutil (2015) points out that the conceptualisation of corporeal movement is often optically biased, whereas sensations that are independent from sight are often neglected. The ubiquity of visual record is certainly a factor in this process. Still, movement cannot be entirely represented and therefore fully understood exclusively by means of visual media. In fact, interpreting human movement objectively as displacement of body parts in a three-dimensional space would result in a limited interpretation. Merleau-Ponty (2002) famously points this out giving the example of typing:

> The subject knows where the letters are on the typewriter as we know where one of our limbs is, through a knowledge bred of familiarity which does not give us a position in objective space. The movement of her fingers is not presented to the typist as a path through space which can be described, but merely as a certain adjustment of motility, physiognomically distinguishable from any other. (Merleau-Ponty, 2002, p. 166)

This is possibly one of the reasons why the use of absolute position in a Cartesian coordinate system imposes some constraints and challenges to high-level analysis of motion data and its use for expressive applications.

In previous works, I used MoCap to carry out experiments aimed at analysing relationships between body movements and other musical features in instrumental music performance (Visi et al., 2014a, 2015b). For real time music applications, the use of various wearable sensors was preferred, as they are easier to transport and use in performance situations, whereas optical MoCap is definitely more demanding in terms of portability and setup time. As it will be described more in detail, the raw data returned by wearable sensors is intrinsically different from that of MoCap, and this has some implications for how the data is eventually interpreted and used. Understanding how to extract meaningful descriptors from such sensors is useful beyond the domain of musical practice, since similar technologies are becoming ubiquitous and are already employed in everyday objects such as mobile devices and game controllers.

Previous research (Freedman and Grand, 1977; McNeill, 1996) has pointed out that upper body movements are of particular interest when observing expressive behaviour. In instrumental music performance, the upper limbs have a central role, since they are often involved in the main sound-producing gestures. Moreover, in most cases, hands and arms are the main points of contact between the body of the performer and the instrument. Therefore, in the studies described in this chapter the sensor bands were placed on the forearms of the performers. However, as shown in the following sections, processing data from Inertial Measurement Units (IMU) using motion descriptors and Machine Learning models allows us to obtain information related to full body movements, which can be used to extract expressive movement features.

In the work described in chapter 6, I made use of fingerless gloves in order to interfere less with the musical instrument manipulation. I progressively moved away from gloves in order to obtain an even less obtrusive configuration. I first placed the sensor on the wrists and eventually moved further away from the hands of the performer, on to the upper forearm. Doing so did not reduce the amount of information about hand movements I was able to retrieve. On the contrary, by using specifically designed descriptors and exploiting the constraints imposed by the structure of the limbs and the interdependence of its parts, I was able to estimate various measures describing the movement of both hands.

During the initial stages of the research I mostly employed IMU sensors, while later I sought to include a form of muscle sensing. This was done in order to address and estimate body movement components beyond those strictly related to displacement in space, such as proprioception and effort qualities.

To address some of the challenges in storing, sharing, and displaying multimodal and motion capture data I have written a MATLAB function that allows to convert MoCap Toolbox (Burger and Toiviainen, 2013) data structures into a format that can be uploaded and correctly visualised in the repoVizz online repository (Mayor et al., 2013). RepoVizz[3] is a system for remote storage, browsing, annotation, and exchange of multimodal data developed by researchers at Universitat Pompeu Fabra, Spain. The MATLAB function, named *mcrepovizz*, is included in the MoCap Toolbox[4] for MATLAB, which is maintained by researchers the University of Jyväskylä, Finland.

[3] http://repovizz.upf.edu/

[4] https://www.jyu.fi/hum/laitokset/musiikki/en/research/coe/materials/mocaptoolbox

7.3.1 IMU/MARG Sensors and Degrees of Freedom

Inertial Measurement Units (IMU) are small, low cost, highly portable devices that incorporate accelerometers and gyroscopes. When these devices are paired with magnetometers, the resulting arrays are also known as Magnetic, Angular Rate and Gravity (MARG) sensors. These sensor arrays allow the tracking of acceleration, rotational velocity and orientation relative to the earth's magnetic field of whatever they are attached to. They are used extensively in aviation, robotics and Human-Computer Interaction (HCI). Their increasing affordability and small size have made them a very common feature of mobile and wearable devices and other consumer electronics. Recently, sensors featuring 3D accelerometers, 3D gyroscopes, and 3D magnetometers have become the most widely used type of IMU/MARG. They enable to estimate various motion features including optimised three-dimensional orientation obtained by fusing together the data from the different types of sensors. These devices are often marketed as 9DoF (9 Degrees of Freedom) sensors, since they consist of three tri-axis sensors and thus have a total of nine sensitive axes. However, in the context of motion tracking the label '9DoF' might be misleading. In fact, a rigid body in a tridimensional space has a total of 6 degrees of freedom, divided in three basic translations and three basic rotations[5]. Thus, any possible movement of a rigid body, no matter how complex, can be expressed as a combination of the basic 6 degrees of freedom. Therefore, in order to avoid confusion I will use the acronym IMU/MARG to refer to sensors comprising accelerometer, gyroscopes, and magnetometers. I might use the acronym IMU if only the inertial sensors (accelerometers and gyroscopes) are taken into consideration. The acronym '6DoF' (6 Degrees of Freedom) will instead be used when motion data actually describes both translation and rotation movement, as in the study presented in chapter 9. Another way of referring to IMU/MARG sensors found in biomechanics and gait analysis literature is MIMU (Magnetic and Inertial Measurement Unit) (Bergamini et al., 2014; Trojaniello et al., 2014). However, the acronym was previously used to indicate other sensors (Shang et al., 2002) and therefore might result ambiguous.

Whereas the raw data obtained using marker-based optical motion capture consists of samples of position based on a 3D Cartesian coordinate system[6], the data returned by IMU/MARG sensors is usually in the form of three three-dimensional vectors, each one expressing acceleration, rotational velocity, and orientation respectively. The sensor

[5]To learn more: `https://en.wikipedia.org/wiki/Six_degrees_of_freedom`

[6]Most marker-based systems also allow to capture 6DoF data (consisting of three-dimensional position and Euler angles) by defining rigid bodies. However, this is achieved by processing positional data of the single markers grouped into a rigid body.

band adopted more recently[7] returns acceleration in units of G, rotational velocity in degrees per second, and orientation angles in radians. Orientation is estimated also using a quaternion representation, which – unlike Euler angles – is not subject to problematic singularities such as gimbal lock (Brunner et al., 2015). In addition to that, the sensor band returns 8-channel electromyogram (EMG) data, which was used to compute descriptors of muscular effort and estimate the movements of wrists and fingers.

Calculating absolute position from IMU/MARG data in real time is technically very difficult if not unfeasible, as the operation would require double integration of acceleration data. This would result in a considerable amount of residual error since drift would accumulate quadratically. Moreover, it would also be relatively expensive in terms of computation. Madgwick et al. (2011) designed computationally efficient algorithms for compensating the residual error. These were implemented for estimating position from IMU data recorded in situations where specific constraints could be exploited, such as gait analysis and cyclic motion.

The data obtained from IMU/MARG sensors is therefore morphologically very different from positional data returned by optical MoCap. The differences in the way movement is tracked and represented by the two different technologies has implications on how movement data is eventually interpreted and used, particularly in the context of expressive movement tracking. High-level movement descriptors are often used to extract features from the raw motion data that can help describing the *meaning* that the movements of the subject convey. This is no trivial task, and different interdisciplinary approaches have been adopted in the past two decades. The following section will look at some of the motion descriptors most widely used with positional data and discuss how they can be adapted and used with IMU data.

7.3.2 Movement Descriptors and Wearable Sensors: Understanding Digitised Movement Qualities

Computable descriptors of human motion are used across several disciplinary fields for various applications ranging from kinesiology and gait analysis to HCI and gaming. Even though human motion data analysis has become an increasingly active field, there is still little consensus regarding which descriptors and methodologies yield meaningful representations of human body motion.

The MoCap Toolbox (Burger and Toiviainen, 2013) provides a wide range of MATLAB scripts for offline kinematic analysis and visualisation, whereas expressive feature

[7]Myo armband, produced by Thalmic Labs: `https://www.myo.com`

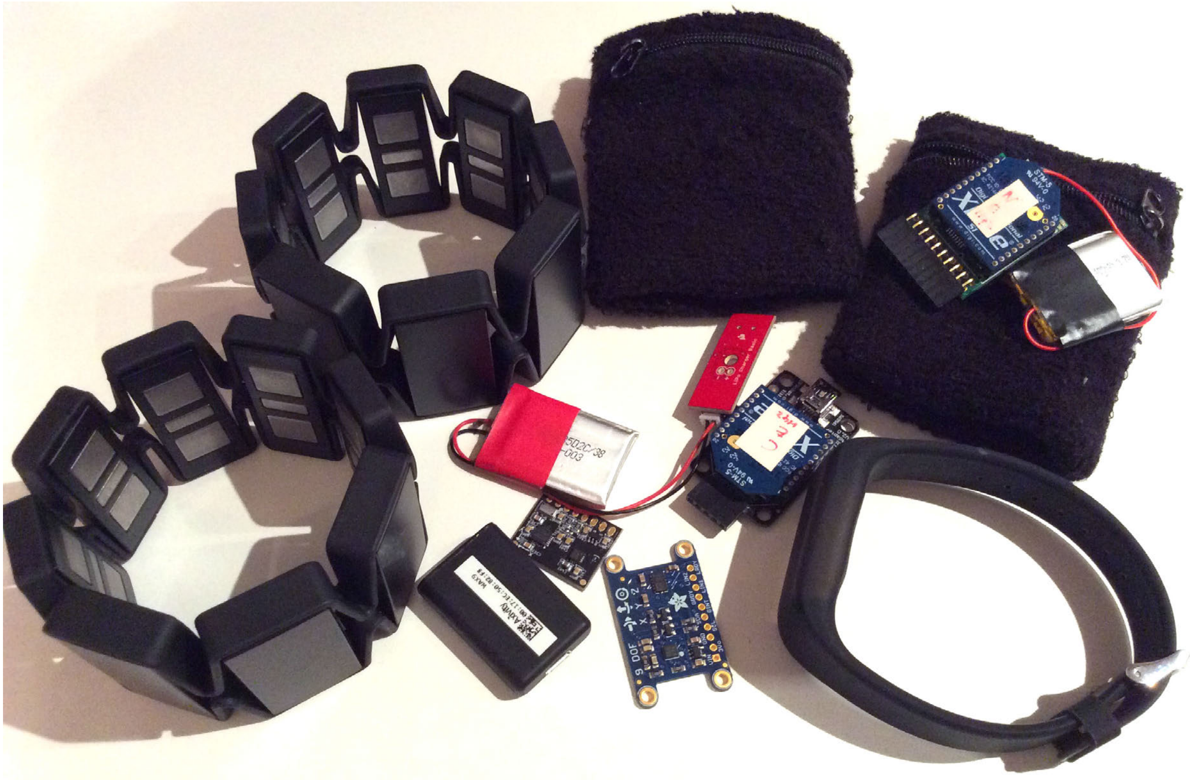


Fig. 7.1 Some of the wearable devices used. Clockwise from the bottom-left corner: Myo armbands, Sense/Stage controllers with wristbands, Axivity WAX9 with silicone wristband, Adafruit 9-DOF IMU Breakout, FreeIMU v0.4

extraction and real time interaction are prominent features of the Eyesweb platform (Camurri et al., 2007).

Going beyond traditional low level kinematic features has proven to be a challenge especially when dealing with expressiveness, emotions, affective states, and meaning. Larboulette and Gibet (2015) recently attempted a thorough review of computable descriptors of human motion. This is indeed a useful endeavour, however it shows that there is still segmentation and many procedures are either ill-defined or redundant, as very similar concepts appear in other research literature under different names.

Most of the descriptors mentioned below were conceived using positional data. However, the principles behind their design are nonetheless useful for describing certain movement qualities, therefore I attempted to adapt them to the data obtained form IMU/MARG sensors. A series of abstractions were implemented using Max[8], which was chosen over other programming environments as it allowed to rapidly prototype and test algorithms for real-time interaction and easily integrate them with other music applications. These abstractions – along with example patches and other tools – are included in KineToolbox[9] (see section 8.3.1).

Fluidity and Jerkiness

In kinematic analysis, "jerk" is the name given to the third order derivative of movement position, namely the variation of acceleration over time. The definition of "jerk index" as the magnitude of the jerk averaged over the entire movement (Flash and Hogan, 1985) was used by Pollick et al. (2001) alongside other descriptors to correlate arm movement to basic affective states. This relates to the fluidity or smoothness of a movement – as fluid movement tend to have even velocity therefore low values of higher-order derivatives – and can be used to detect emotionally relevant information in movement data (Glowinski et al., 2011). In fact, roughly speaking jerkiness could be seen as the inverse of fluidity. Piana et al. (2016) define the fluidity index as a local kinematic feature equal to $\frac{1}{\int (j_i+1)dt}$, where j_i is the jerk of the *joint*[10] i. This means that higher values of jerk corresponds to lower fluidity.

To estimate jerkiness using IMU data instead of positional data, I averaged the derivatives of longitudinal acceleration returned by the accelerometer $(\dot{ax}, \dot{ay}, \dot{az})$ into a single jerk index. In order to calculate jerk values also from rotational movements, I averaged the second order derivatives of the angular velocities returned by the gyroscope

[8] https://cycling74.com/products/max/

[9] https://github.com/federicoVisi/KineToolbox

[10] A point belonging to a three-dimensional representation of a body, usually defined by positional coordinates.

$(\ddot{gx}, \ddot{gy}, \ddot{gz})$. The values obtained from both sensors are then combined and summed over a time window of length N samples, as shown in equation 7.1:

$$IMUJerkiness(t) = \sum_{k=0}^{N-1} \alpha_1 \frac{\mid \dot{ax}_{t-k} \mid + \mid \dot{ay}_{t-k} \mid + \mid \dot{az}_{t-k} \mid}{3} + \\ + \alpha_2 \frac{\mid \ddot{gx}_{t-k} \mid + \mid \ddot{gy}_{t-k} \mid + \mid \ddot{gz}_{t-k} \mid}{3}. \tag{7.1}$$

Coefficients α_1 and α_2 are weights that balance the data magnitudes obtained from the accelerometer and gyroscope sensors. These coefficients can be useful to adjust the influence of rotational and translational movement on the Jerkiness estimate. It is worth mentioning that in real world implementations derivatives are very sensitive to signal noise, therefore sensor data may require low-pass filtering before Jerkiness can be computed.

From the conceptual framework of Laban Effort Elements/Qualities (Laban and Lawrence, 1947) Jerkiness (and its counterpart Fluidity) are useful descriptors that can aid the computational analysis of expressive movements. Laban defines four basic Effort Factors (Flow, Weight, Time, and Space), each Factor is a continuum between polarities described by Effort Element/Qualities. In particular, Flow is related to the continuity and control of the movement. Its polar qualities (Free Flow and Bound Flow) have been previously associated also to aspects of fluidity of the movement (Hackney, 2002). A movement characterised by Free Flow Effort Quality is "fluid", "liquid", and "outpouring". On the other hand, the Bound Flow quality indicates containment, restrain, and control. In addition to Flow, Jerkiness can be related also to the Time Effort Elements. A movement characterised by Sustained Effort qualities is expected to have low level of Jerkiness. On the other hand, a movement with Sudden Effort qualities ("urgent", "quick", "staccato") will have most likely a higher rate of change of acceleration and therefore higher levels of Jerkiness.

Jerkiness and Fluidity can then contribute to the analysis and recognition of expressive movement qualities, in particular in multimodal frameworks that involve multiple descriptors and sensing modalities (Camurri and Volpe, 2011; Caramiaux et al., 2015a). However, it is important to restate that Laban Effort Elements are qualitative "inner attitudes" of the mover towards the Effort Factor. Using computable descriptors should not be seen as an attempt to quantitatively measure Effort Qualities but rather as a means to aid the design of computational models capable of discerning and recognising different expressive movement behaviours.

Quantity of Motion and Quantity of Rotation

Fenza et al. define Quantity of Motion (QoM) as the sum of Euclidean distances between successive points in a time window (Fenza et al., 2005) and Glowinski et al. included a similar measure in their expressive feature set defined "overall motion energy" (Glowinski et al., 2011). To compute an analogous feature using IMU/MARG sensor data, I aggregated the magnitude of the variations of the norm of the orientation quaternion ($\|q\|$) and of the average acceleration over the three axes (a). The values for each frame are once again summed over a time window of length N samples, as shown in Equation 7.2.

$$IMUQoM(t) = \sum_{k=0}^{N-1} \beta_1 \mid \|q_{t-k}\| - \|q_{t-k-1}\| \mid + \beta_2 \mid a_{t-k} - a_{t-k-1} \mid . \tag{7.2}$$

Similarly to Equation 7.1, β_1 and β_2 are weights to balance individual contributions from distinct sensors. As with *IMUJerkiness*, the value of these coefficients can be changed to adjust how rotational and translational movement affect the *IMUQoM* estimate. If the value of β_1 is set to zero, the accelerometer data is removed from the equation. In this specific case, only rotational motion data is used for the computation of the descriptor. Thus, a new motion descriptor named Quantity of Rotation can be defined as follows:

$$QoR(t) = \sum_{k=0}^{N-1} \mid \|q_{t-k}\| - \|q_{t-k-1}\| \mid . \tag{7.3}$$

This descriptor can be employed to measure the amount of rotational motion of an object or area of the body independently from translational motion. It can be used with any rotational motion data as long as it is expressed in quaternions, therefore its use is not limited to IMU/MARG data but can also be employed with 6DoF data obtained using marker-based motion capture. This is further discussed and implemented chapter 9, specifically in equation 9.2, where standard QoM and QoR computed using 6DoF motion capture data are aggregated.

Contraction/Expansion and Symmetry

Contraction and expansion of the body can be computed in different ways, for example by calculating the area of bounding shapes (Glowinski et al., 2011), using Contraction Index (Fenza et al., 2005), or measuring the volume of a convex hull that encloses the body (Hachimura et al., 2005).

When wearing two IMU/MARG sensors on the forearm, it is possible to project the orientation values over a hypothetical 2D plane in front of the subject and thus obtain approximate coordinates of the points in the plane the arms are pointing to. First, the yaw values for both arms have to be centred while the subject is pointing both arms forward. Then, given θ_{yaw} and θ_{pitch} as the yaw and pitch angles respectively (expressed in radians) the coordinates of the point in the plane corresponding to one arm can be calculated as follow:

$$(x, y) = \left(\frac{x_{max}}{2} + \frac{x_{max}\, \theta_{yaw}}{2\pi}, \frac{y_{max}}{2} + \frac{y_{max}\, 2\theta_{pitch}}{2\pi} \right). \tag{7.4}$$

By calculating the Euclidean distance between the points corresponding to each arm it is possible to estimate whether or not the arms are pointing in the same direction. When arms are spread wide open (thus pointing at opposite directions), the distance between the two points will be at its maximum. This way it is possible to have a value that depends on whether arms are wide open or are resting close to the body. This value can be used as an expressive feature, even though it is probably not as precise as the contraction indexes obtained using optical motion capture, since the values are based on the orientation of the arms and not on their actual position.

By comparing the coordinates we can also see if there is horizontal or vertical symmetry between the arms, which is another useful postural feature that has been previously used for the analysis of expressive movements (Camurri and Volpe, 2011).

7.3.3 Periodicity and Rhythmic Qualities: Periodic Quantity of Motion from Multiple Sources

While the descriptors mentioned above could be described as *spatial* or *spatiotemporal* features of the movement, periodicity is a purely *temporal* quality.

In chapter 5, I introduced Periodic Quantity of Motion (PQoM) as a means to analyse movement periodicities in relation to musical rhythmic subdivisions. It was defined it as follows:

$$PQoM[t, f_{c_k}] = \sum_{n=t-N}^{t} H_{f_{c_k}} \{ (\boldsymbol{w}(|\boldsymbol{x}[n] - \boldsymbol{x}[n-1]|)) \}, \tag{7.5}$$

where f_c are centre frequencies of the bandpass filters calculated following the music tempo and its sundivisions, N is the window size in samples, $H_{f_{c_k}}$ is the k^{th} bandpass filter operator, vector $\boldsymbol{w}$ contains the weight coefficients for each tracked point, and vector $\boldsymbol{x}$ contains the Euclidean norms for each tracked point (for further information about PQoM refer to section 5.3.3).

The same equation can be used to estimate PQoM using motion data from various sources, including IMU/MARG sensors. When combining multiple sensors, a specific $\boldsymbol{w}$ vector of coefficients must be defined for each sensor type. For example, if using IMU/MARG data, the coefficients in $\boldsymbol{w}$ determine the weights the data from the accelerometer, gyroscope, and magnetometer have for the estimation of PQoM.

Figure 7.2 presents the PQoM values of a recorded arm movement computed using motion capture data and the IMU/MARG data from a Myo armband. Figure 7.2a shows the locations of the passive markers and the Myo armband. In this example, the musician performs a periodic movement with his arm, synchronising it with the half note rhythm (from 1 to 3 seconds) and with the quarter note rhythm (from 5 to 7.5 seconds). Figure 7.2b shows the each tracked $\boldsymbol{x}$ vectors: motion capture marker (green line), acceleration (blue line) and orientation (red line).

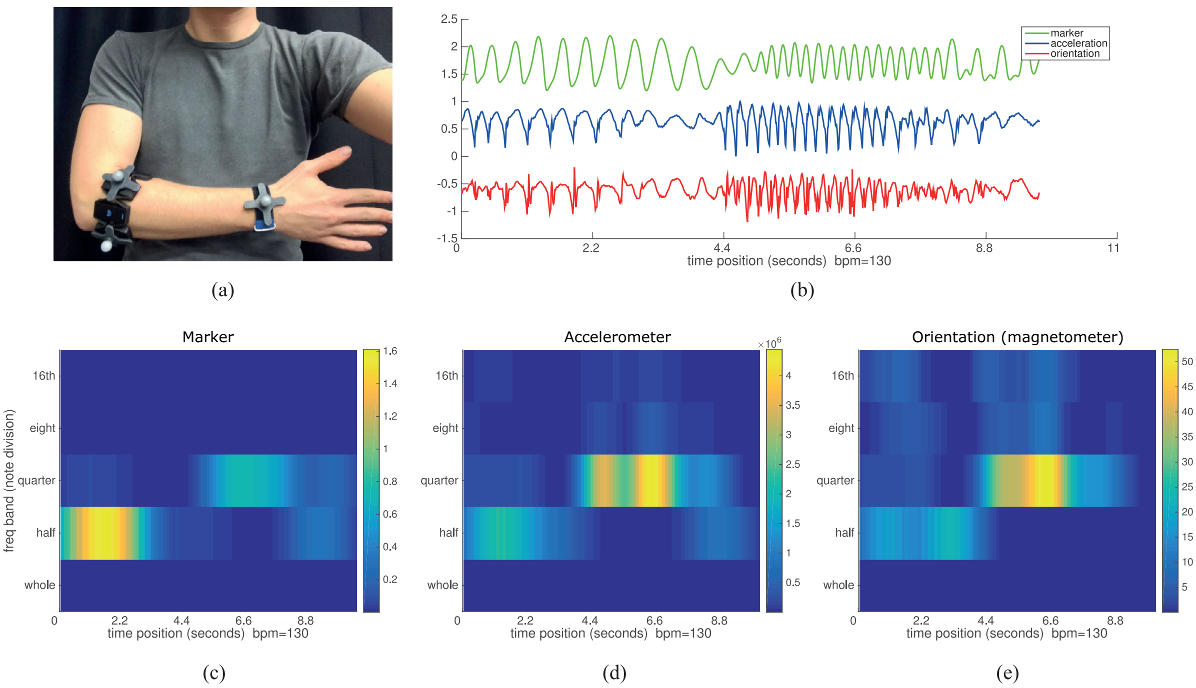


Fig. 7.2 (a) Locations of reflective markers and Myo armband. (b) Data from motion capture (green), accelerometer (blue), and magnetometer (red) (rescaled for better visualisation). (c) PQoM from motion capture data. (d) PQoM from accelerometer data. (e) PQoM from orientation data (magnetometer).

The bottom part of the figure shows the PQoM measures extracted from the tracked data. The amount of gesture periodicity over time is indicated by the light blue and yellow regions (higher values). The PQoM estimates from the motion capture marker (Figure 7.2c) follow a similar pattern of the PQoM extracted from the acceleration

(Figure 7.2d) and the orientation data (Figure 7.2e). However, magnitude may differ considerably, since each sensor measures a different kinematic feature of the movement (e.g. acceleration, angular velocity, etc.) and thus may respond differently. Therefore, the weight coefficients in $\boldsymbol{w}$ should be adjusted in order to balance the PQoM values calculated from different sensor data. Such coefficients can be adjusted arbitrarily, in order to obtain an aggregated PQoM estimate based on a mix of kinematic features or a specific feature, depending on the purpose of the analysis.

7.3.4 Machine Learning: Mapping Postural and Sonic Topologies

Motion descriptors are useful for extracting meaningful features from the raw data and they also allow to aggregate information relative to all the axes. This helps to move away from a low-level movement representation constrained by the Cartesian coordinate system and obtain motion data that is less dependant on it. A system based on orthogonal axes is indeed a convenient way to digitise movement. However, meaningful conceptualisation that help in interpreting the expressivity that body movement conveys may be hindered if subordinated to a highly disciplined method of quantitative representation. In his article about topology and data, Carlsson argues that "coordinates [. . .] are not natural in any sense, [. . .] therefore we should not restrict ourselves to studying properties of the data which depend on any particular choice of coordinates." (Carlsson, 2009, p. 256). Moreover, describing the characteristics of topological methods, he states that in order to obtain *knowledge* about the data, qualitative information is needed, and this has to be established before proceeding with quantitative analysis. Topology studies intrinsic geometric properties of the objects, which do not depend on a chosen set of coordinates and it has also been employed in the analysis of dance patterns (Naveda and Leman, 2010). This approach provides very useful notions for interpreting movement data generated by music performance gestures. In fact, such body movements are bound to multimodal expressive features, which are inherently qualitative.

To put these concepts into practice, machine learning algorithms were used to define interaction models based on different postures a musician may adopt during a performance. This was done by asking the performer to play freely while wearing two sensors armbands. A small number of postures (4–5) are then defined. This was done by observing recurrent idiosyncrasies and peculiarities of the performance and discussing the qualities of the movements with the musicians themselves, to better understand how certain movements relate to each respective instrumental techniques and with musical features of the pieces performed[11]. Sensor data is then sampled repeatedly during each pose in order to train a Support Vector Machine classifier. This was implemented using the *ml.lib* library (Bullock and Momeni, 2015) for Max, which is itself based on the Gesture Recognition Toolkit by Gillian (Gillian and Paradiso, 2014). Every posture is then associated with a set of parameters of a digital sound processing engine. During the performance, the machine learning classifier compares the incoming sensor data stream with the recorded examples, returning the values for the probability (or *likelihood*) that the current posture of the musician matches each of the defined classes. The values of the probabilities are then used to interpolate between the parameter sets of the sound engine, which can be used to process the sound of the instrument in real time or synthesise electronic sounds.

This practical approach resonates with the aforementioned notions of topology, since the incoming data is not analysed quantitatively, but it is instead evaluated in terms of proximity/distance from the predefined postures[12]. From this perspective, the sampled postures themselves are topologies determined in relation to qualitative aspects of the movement of that particular performer, thus avoiding dependency from an abstract, artificial coordinate system. The system is instead defined by the idiosyncrasies of the performer.

This approach has also several practical advantages compared to more traditional sensor-to-sound parameter mapping approaches. Incoming sensor data does not need to be rescaled to the range of the sound parameters it is mapped to. The quantitative values of the sensor data can be in fact ignored, since the classifier probabilities are used to interpolate multiple sound parameters. This is another advantage, since complex mappings can be easily defined and parameter modulation is independent from any coordinate system. Instead, the system quickly adapts to different users, and this is desirable considering the substantially different movements required for playing different instruments and the significant degree of idiosyncrasy that characterises musical performance.

It is worth mentioning that in the past few years machine learning techniques have been increasingly employed for interactive computer music performance. Notable approaches include Fiebrink's Wekinator (Fiebrink et al., 2009b), Caramiaux's Gesture Variation Follower (Caramiaux et al., 2013) and Françoise's mapping by demonstration (Françoise et al., 2014).

[11] Some early tests with different musicians: `https://youtu.be/stWI43-EZGA`

[12] This idea is reinforced by how Support Vector Machines work: "An SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible. New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall on." (Bullock and Momeni, 2015, p. 4).

In early tests, orientation data and an aggregate EMG descriptor for both arms were used as inputs to train the machine learning models. Orientation was chosen as it is not an inertial measurement therefore it can be used to describe postures. In addition to that, EMG data allows us to consider other characteristics of the movement. As pointed out by Salazar Sutil (2015) the perception body movement involves sensations that go beyond displacement in space, such as interoception and proprioception. Moreover, in their extensive work on the analysis of expressive movement, Camurri and Volpe (2011) define gesture as a multimodal entity, citing Laban's Theory of Effort (Laban and Lawrence, 1947) as a central source of concepts for the understanding of expressive movement.

7.4 Case Study: Kineslimina

Kineslimina is a piece for viola, electric guitar, motions sensors and live electronics that explores the use of the musicians instrumental gestures and movements as an expressive medium. Such gestures merge with the other musical features and become an integral part of the score. While playing their instruments, the musicians wear an armband fitted with IMUs, which tracks their movements and sends the motion data to a computer. The computer then processes the movement data and sound, responding with a wide range of dynamics: from subtle timbral alterations that follow the movements of the bow during string changes to deeper resonances when more overt gestures are performed by the musicians.

The title is a portmanteau of the words *kinespheres* and *limina*. The concept of kinesphere was defined by Laban (1966) as “the space within the reach of the body”, “that is occupied by and surrounds the body”. The kinesphere is a personal space, and how an individual relates and pays attention to it contributes to delineate it. “Limina” is instead the plural of “limen”, that is “threshold”, “margin”. The piece in fact aims at pushing the boundaries of the personal spaces that surround the musicians during the performance. The sound of the instruments is altered, and synthesised sounds are engaged by exceeding the usual extent of instrumental movements. The score of the piece can be seen as a script, through which a multimodal choreography emerges as the product of learnt body schema, altered by the influence and the reactions to an interactive system. In the ritualised context of musical performance, a non-conventional technology (the sensors) interferes with conventional ones (the instruments), reconfiguring the relationships between the score, the performers and their tools.

Parviainen et al. (2013b) propose an approach to interaction design that considers choreography as the holistic, experiential continuum of human movements resulting from the interaction with artefacts. From this perspective, musical instruments, sensors and movement/sound mappings can be seen as carriers of a set of pre-choreographies. The design of these objects (whether material or not, as in the case of software) and the environment where the interaction takes place pre-choreograph the performance of the piece. All the movement opportunities that these objects afford form the basis for the actual choreography that emerges as the score is enacted. A pre-existent, overarching design that influences the movements of the kinesphere. Eventually, the movements that shape the performance exceed what the individual kinespheres can capture. The relations between the two musicians and between the musicians and the audience and the dynamics that arise from these connections are what Parviainen et al. call "local-level movements".

As noted by Wilson (2013, p. 426), a traditional musical instrument is not a mere piece of technology, as our relationship to it is shaped by "the way it is 'meant' to be played, the canonic tradition that stands behind it as repertoire, and the normative expressive gestures that are 'input' by the player and 'output' sonically by the instrument". Bodily relationships with these cultural artefacts are mediated historically and become part of a shared knowledge. Introducing motion sensor technology in this picture adds another layer of complexity, tightly woven to the already established gesture-sound relationships. Figure 7.3 shows an example of how these different aspects of the piece are interrelated. Towards the end of the piece, the score requires the viola player to repeat an arpeggiated pattern with increasing dynamics, until a synthesiser part is heard. The part entails repeating bow strokes, and the movement pattern is captured by the sensors placed on the right wrist (blue line). The peaks in the acceleration data control a granular synthesis engine that samples and alters the timbre of the instrument at each peak. At the same time, Quantity of motion is computed (red line) and the increase of motion activity introduces other electronic parts, until the QoM data reaches a predefined threshold (green cross) at which point the closing synthesiser part is triggered and the musicians can move onto the closing notes of the piece. The amount of repetitions required to reach this point depend on the movements of the performer, which may vary according to how different musicians interpret the score. The score engenders the instrumental movements required to perform the piece, the movement data alters the sound of the instrument and has also an impact on the structure of the score itself (the number of repetitions required). This closes a feedback loop in which every part mutually influences the other. The body of the performer is the is the medium, the *locus*, where this dynamic entanglement takes place.

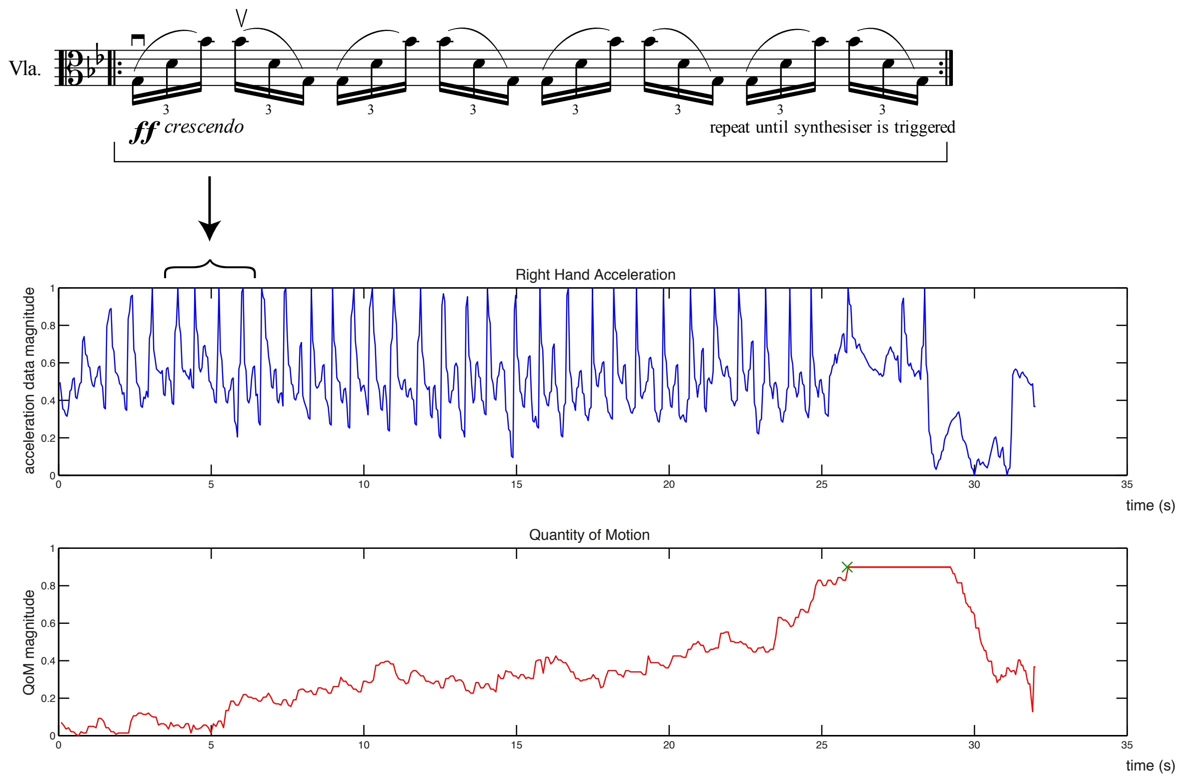


Fig. 7.3 Kineslimina: scored pattern, right hand acceleration of the viola player and Quantity of Motion. The green cross indicates when the QoM threshold is crossed and the synthesised part that closes the piece is played back.

This interdependency also affects how motion data are mapped to sound parameters. Mapping sensor parameter to sounds has long been a debated topic in the HCI and NIME research communities (Hunt et al., 2003). As also noted by Leman (2008a, p. 164), freedom of mapping that characterise digital musical interfaces "may disturb the sense of contact and of non-mediation". Drawing from established vocabulary of gesture of a traditional musical instrument and exploiting the constraints that instrumental techniques pose on the body can result in an advantage for obtaining meaningful interactions for expressive music performance. This approach takes advantage of the ecology (Clarke, 2005) of musical instruments in order to obtain expressive transparency in gesture-sound mapping.

Kineslimina premiered at the Gala Concert of CMMR 2015 – the 11th International Symposium on Computer Music Multidisciplinary Research[13] (see Fig. 7.4) and was later performed also at MuSA 2015 – Sixth International Symposium on Music/Sonic Art[14], held at the Institut für Musikwissenschaft und Musikinformatik (IMWI) in Karlsruhe, Germany. From the perspective of the performers[15], the piece re-configured the relationship between musician and instrument, extending expressive possibilities through their instrumental movements tracked by the sensors. However, this also required the performers to learn new skills and embed them in their existing instrumental techniques. This process became evident during the rehearsals. The performers experienced an increased awareness of the fundamental body schema of their instrument-playing, as subtle movements created new sonic results through the motion sensors. This made them pay renewed attention to movements they learnt in the early days of their formation as musicians, essential parts of the vocabulary of gestures of their respective instruments. Whilst the musicians learn and get more familiar with the sensors, the system itself gets constantly adapted and adjusted to accommodate the needs of the performers and better follow their performance styles. As the rehearsals go on, relationships between body movements, instrumental gestures and sensor data are re-negotiated. This does not just imply mere parameter adjustments and technical improvements to the sensor system. The process elicits and entails a careful analysis of the relationship between movements, sound and score from the privileged perspective of the performers themselves, thus resulting in a useful contribution to research from practice-led perspective (Sullivan, 2009).

[13] http://cmr.soc.plymouth.ac.uk/cmmr2015/index.html

[14] http://zilmusic.com/musa2015/

[15] Esther Coorevits: viola, motion sensors, live electronics; Federico Visi: electric guitar, motions sensors, live electronics.

The performers then progressively get to know how the mappings of movement features to sound work, and how they can explore this unconventional technology in a meaningful way. More than just sonifying the movements made by the performers, the sensor system also induces the musicians to reconsider the relationship to the performance space. Such space is where we locate the relationships between players, instruments and audience, which encompasses a set of conventions and cultural practices that are established and embodied in the performer. This can be compared to what Ervin Goffman refers to as "the frame", which is the perceptual mechanism that indicates the nature and purpose of a behaviour, and how it is to be interpreted. It is a tool for understanding the implicit agreement between performer and audience on the symbolic status of the performance (Goffman, 1974). The performance space is the framework in which we understand a performance. This frame is established through cultural practice, traditions and conventions. From the perspective of a musician, this frame consists of the historically established relationships that are found between players and instruments. Performance techniques and experiencing instruments and instrumental music are habitualised through historical practice, conventions and education. These relationships are thus part of the embodied knowledge of the performers.

In Kineslimina, performing with reconfigured instrument/body/space relationships has made the musicians more aware of other qualities of their movements and their kinespheres.

> Laban identified space, weight, time and flow as motion factors toward which performers of movement can have different attitudes depending on temperament, situation, environment and many other variables. The attitudes toward the motion factors he called [. . .] Effort. [. . .] Choices are continuously being made by all people in motion, consciously or unconsciously, to determine what combinations of these Effort elements will best serve the purposes of their intentness or modify their behaviour. [. . .] Whatever the action in which the effort combinations appear, the whole biological/psychological system is involved. (Bartenieff and Lewis, 1980, p. 51)

Intentionality is a key aspect in the study of musical gestures, as the fact that they are goal-directed actions is an essential quality for the understanding of their expressive qualities (Godøy and Leman, 2010). The Effort qualities of a movement are very much the result of this intentionality and they play an important role in the perception and understanding of body movement. During the performances, effort qualities and intentionality appeared amplified by the presence of the motion sensors

Fig. 7.4 Kineslimina: performance during the Gala Concert of the 11th International Symposium on Computer Music Multidisciplinary Research (CMMR). 16 June 2015, Plymouth, UK. Esther Coorevits (viola, motion sensors) and Federico Visi (electric guitar, motion sensors).

and their effect on the conventional performance gestures. This was – possibly[16] – also perceived by the audience, as the interplay between the performers and the role of their intersubjective space was made more transparent through the augmentation of their musical intentionality.

7.5 Summary and Discussion

In this chapter I presented techniques for interpreting motion data and discussed the implications that arise when employing motion sensors together with traditional instruments in musical practice. Movement descriptors designed to be used with IMU/MARG data were defined and a case study of a piece for viola, electric guitar, and motion sensors was then described, discussing the relationship between body movement, musical score, and motion data.

[16] Some members of the audience have informally commented after the performances, stating that strong interplay between the performers and engaging gestural narrative were perceived.

7.5.1 Beyond Control

Within this context, it is clear that body movements and gestures go well beyond simply being activators of technological objects, whether these are motion sensors or musical instruments. Body movement is considered a key element for the formation of embodied musical meaning (Leman, 2010), however its role as an important cog of the engine that engenders signification and cognition is obviously not limited to the musical context. Technological objects on the other hand have the capacity to entail gestures and store their potential meaning. As observed above and also in chapter 3, a musical instrument can be seen as a receptacle of gestures, of *kinemes* that – through a performance – give rise to a multimodal choreography. From a wider perspective, we could say that objects extend our selves, they are co-substantial, continuous, and co-extensive parts of minds in action (Malafouris, 2013). Moreover, Wilson (2013, p. 430) stresses the importance of the relationship between instrument technology and instrumentalist's pedagogy: "technology – what the instrument *is* – is inherently entangled with pedagogy, the historically established relationships found between instrument and instrumentalist.". In the case study presented here, this "inherent entanglement" encompasses also the score of the piece, which elicits the instrumental movements necessary to execute its parts and – at the same time – it is affected by them through the use of motion descriptors. Within this layered process of signification – situated in a cultural ecology and shaped by shared knowledge – the body is the medium where everything takes place. This perspective is akin to Merleau-Ponty's phenomenological approach (Merleau-Ponty, 2002).

7.5.2 Beyond the System Used to Represent It

Once the centrality of the body and its movements in the ways we make sense of the world is recognised, it is clear that – in an increasingly pervasive digital semiosphere – being able to digitise movement and interpret motion data becomes of primary interest. However, movement seems to have properties that exceed the system used to represent it. I have discussed the limitations emerging from representing movements exclusively through visual media, and the ubiquity of visual record is certainly a factor in this process. However, solving this bias is one of the challenges that contemporary researchers and practitioners have to face in order to make progress in the discourse on human movement. The development of different computational techniques to describe the qualities of body motion are a necessary step towards more meaningful interpretation of data generated by human movement. However, it is also vital to consider the constraints posed by rigid methods of representation and move towards approaches that allow to address the complex, non-linear phenomena that characterise expressivity.

Using inexpensive and unobtrusive devices such as IMU/MARG sensors may also help to move the research outside of laboratories. As seen previously, ecology plays an important role in the way we make sense of music (Clarke, 2005), similarly, being able to study movement "in the wild" may have considerable implications, as shown in previous studies (Woolford, 2014).

IMU/MARG sensors provide data that is morphologically very different from the one obtained through optical motion capture. However, it is possible to obtain analogous meaningful information if the data is correctly interpreted. In this context, using machine learning techniques is not only a quicker method to obtain complex interaction models that adapt to different individuals. It is also a way to study and reflect upon the topological qualities of human movement through applied research and practice. More sophisticated algorithms to interpret motion data can help to address its complexity, reclaiming the centrality of the body over a rigid representation of data structures.

7.5.3 Beyond Musical Interaction

Extracting expressive movement features from IMU/MARG data can lead to many other applications, well beyond the field of musical interaction. The ubiquity of IMU/MARG sensors – which is a very common feature of recent mobile devices, tablets, VR gear, game controllers, and various other wearable devices – brings about a vast number of potential scenarios were the techniques described here can be implemented. Using higher-level descriptors to estimate expressive qualities of body movements is a way towards implementing dynamic HCI designs that handle gestures not only as isolated objects of application but as part of longer experiential chains, with multiple layers of significance. This goes beyond the traditional use-oriented approach and is akin to the use of choreographies for interaction design proposed by Parviainen et al. (Parviainen et al., 2013b) and Pirhonen et al. (Pirhonen, 2013), and the work on affective computing carried out by the researchers at InfoMus/Casa Paganini (Glowinski et al., 2011; Piana et al., 2013). Both approaches avoid limiting HCI design to goal-directed actions and adopt a more holistic approach that take into consideration a wider ecology of human movement. Moreover, IMU/MARG sensors coupled with sound synthesis techniques have already found applications in the field of rehabilitation of stroke patients (Bevilacqua et al., 2013).

In a scenario where computing is becoming ubiquitous and embodiment is considered a fundamental factor for designing interactions with technology (Parviainen et al., 2013a), the implications of being able to extract meaningful information from motion data are manifold. The study of music-related motion and the analysis of motion descriptors certainly have applications beyond music making. As an example, Bennet et al. 2016 recently measured periodicity in the data obtained from motion sensors applied to rocking chairs in care homes. This was done to help improve the quality of life of residents in dementia care by creating subtle, engaging interactions that support and stimulate memory through music and movement (Bennett et al., 2016). This is but one example of current applications of motion data analysis in conjunction with music. As computing gets embedded in the environment people live in, interpreting the data produced by human activity in a meaningful way is a key issue. This is central for recent trends in computer science, such as affective computing, calm technology, and human-centred machine learning.

7.5.4 Music as a Test Bed

As suggested by a topological approach (Carlsson, 2009), gaining new higher-level knowledge from motion data also requires qualitative insights. To access more complex, structural, and subjective qualities that are more difficult to model quantitatively, data needs to be interpreted also through qualitative methods. Intuitions arising from qualitative approaches can contribute to the understanding of how *body schema* and *kinemes* work in generating embodied meaning. This can successively inform more advanced computational models able to recognise complex and meaningful qualities of human movement.

Practice as research can address the need of qualitative insights in the interpretation of motion data. Particularly, music can be an effective test bed, given its multi-layered complexities and rich cultural, multimodal qualities. As other projects have previously shown, music and performing arts can be effective test beds for new modalities of expressive human-computer interaction (Camurri et al., 2004a), and practice-led approaches have yielded technical and conceptual material useful for the development of motion capture technologies (Norman and Blackwell, 2010). Moreover, musicians are often early-adopters of new paradigms of interaction that eventually become mainstream (Kirn, 2013). Notable examples are gestural controllers and multi-touch technologies, which were adopted by musicians long before they become widespread.

Practice-led approaches are also helpful for carrying out the conceptual work required to make sense of motion data and understand the meanings it can potentially convey. Certain artistic works seek to affirm the irreducibility of the corporeal presence whilst simultaneously sublimating it through digital processing (Norman, 2015). This resonates with the rationale behind *Kineslimina*, and this creative tension can lead to new insights into how movement can carry meaning across physical and digital environments. In Kineslimina, relating motion data to a musical score has shown how multimodal qualities of music are entangled, mutually affecting each other.

7.5.5 Future Scenarios

Future work will adopt this mixed methodology in order to address technical and conceptual aspects related to body movement, motion data, and meaning formation. Other machine learning algorithms will be tested in order to map instrumental gestures to sounds synthesised by means of physical modelling. Sound synthesis techniques based on physical modelling allow to generate sounds that resemble those of certain musical instrument families. Working with synthesis parameters allows one to go beyond the timbral ranges and sonic capabilities of the actual instruments while preserving timbral resemblances to the instrument family they belong to. This poses interesting conceptual challenges, as the recognitions of timbral qualities of musical instruments relies on a shared knowledge. As discussed previously, the relationships between instruments and instrumental movements is also something encoded in a shared gestural vocabulary. The ecology around musical instruments, their timbres, and their instrumental gestures offers a rich conceptual framework for developing meaningful cross-modal mappings between motion data and synthesis parameters. Cross-modal relations between performance movements and the sonic outcomes will also be inspired by the concept of Uncanny Valley, which was previously adopted in a composition that involved tension and relaxation structures in timbrally varied musical phrases generated by physical models (Bessell, 2011).

Music is constituted by abstract structures and performance movements in continual interaction, entangled with technological and cultural knowledge. The concept of choreography appropriately describes the process of multimodal signification that emerges from the performance of a musical score. *Body scheme* and *kineme* are useful tools to gain a better understanding of how the body is a central medium in human communication. Movement is a modality of knowledge, therefore being able to interpret it and represent it through technology has potentials that are difficult to imagine, but that should certainly be further explored.

Chapter 8

Designing Tools and Composing Constraints

> [. . .] it takes time to appreciate the limitations of the instrument, the boundaries of its expression, the pushback on one's intents.
>
> LAETITIA SONAMI

> In theory, there is no difference between theory and practice. But, in practice, there is.
>
> JAN L. A. VAN DE SNEPSCHEUT

8.1 Overview

This chapter presents a collection of Max[1] tools for motion-sound interaction named *KineToolbox*[2] and two instrumental music pieces that make use of said tools. By describing both the technical solutions and their implementation, this chapter addresses the challenge of employing body movements of the musicians as a musical feature in composition and performance. Different hardware and software solutions and approaches to parameter mapping for motion-sound interaction are discussed, and a method for integrating movement in traditional music notation is described.

[1] https://cycling74.com
[2] https://github.com/federicoVisi/KineToolbox

8.2 Hardware

In the two pieces described further in this chapter, Myo[3] armbands were employed to sense the movement of the instrumentalists. The choice of this off-the-shelf device came after working with other hardware solutions.

The initial experiments described in chapter 6 were carried out using gloves cut below the knuckles. This was done in order to leave the fingers as free as possible, since they were supposed to be worn while playing a traditional musical instrument, thus unobtrusiveness was desirable. As described more in detail in section 6.3, these gloves included a wireless microcontroller, inertial sensors, and a flex sensor on the wrist (see left section of Fig. 8.1).

Fig. 8.1 The wearable devices used: fingerless gloves with accelerometers and flex sensors (left), wristband with IMU/MARG sensor (centre), and Myo sensor armband (right).

Optical motion capture technologies such as the RGB-D camera adopted in chapter 6 were discarded in performance situations for several reasons. Firstly, systems such as Kinect require a minimum distance between the device and the performer, which makes its usage more problematic with limited stage space and makes it more difficult to perform close to the audience. Also, the presence of different musical instruments makes the

[3] https://www.myo.com

skeletal mapping algorithm behave erratically, as it is not natively trained to recognise the musical instruments held by the performers. Other issues were encountered when performers got very close to each other, causing temporary occlusion of some parts of the body. Apart from these issues, wearables were preferred as they proved more reliable for sensing small-scale movements in a performance environment, where lighting and space may vary.

To further reduce obtrusiveness, the gloves and the flex sensors were successively abandoned in favour of an IMU/MARG[4] sensor (Adafruit 9-DOF IMU Breakout - L3GD20H + LSM303[5] mounted on the same controller and housed inside the pocket of an elastic wrist band (see centre section of Fig. 8.1). This configuration was used for the first version of the piece *Kineslimina*[6], which premiered at the Gala Concert of the 11th International Symposium on Computer Music Multidisciplinary Research (CMMR) held at Plymouth University, Plymouth, UK, in June 2015; and was subsequently performed at MuSA 2015 Sixth International Symposium on Music/Sonic Art: Practices and Theories held in Karlsruhe, Germany, in July 2015 (Visi et al., 2015a).

The launch of the Myo[7] armband designed by Thalmic Labs[8] introduced a relatively affordable device that combines an IMU/MARG with dry electromyogram (EMG) sensors. As described in section 7.3.1, the device allows to sense the forearm muscular activity through 8-channel EMG data sampled at 200 Hz. The Myo armband is worn on the forearm, thus moving further away from the hand in order to interfere even less with the manipulation of the musical instrument (see right section of Fig. 8.1).

Interestingly, by processing the EMG signal from selected channels it is possible to obtain data that describe the movement of the wrist in a similar or more detailed fashion compared to the flex sensor used in the fingerless glove. This done by splitting the EMG channels of the Myo armband into two subgroups, one with the sensors placed on the inner arm and one with the sensors on the outer arm. The flexor muscles used for bending the wrist inward are located in the inner arm, whereas the extensors that are engaged to bend the wrist outwards are located on the outer forearm. In KineToolbox (which will be described in more detail in section 8.3.1) the sum of the absolute values of the sensor data of two predefined groups (1, 8, 7 for the for the inner arm and 3, 4, 5 for the outer arm, using the numbering shown in Fig. 8.2) are assigned to two specific send

[4]IMU/MARG stands for *Inertial Measurement Unit / Magnetic, Angular Rate, and Gravity.* See section 7.3.1 to read more about this technology.

[5]`https://www.adafruit.com/products/1714`

[6]For more information about *Kineslimina* see section 7.4.

[7]The Myo developer kit was made available in late 2014, while the consumer version was released in early 2015.

[8]`https://www.thalmic.com`

objects in order to easily access the values[9]. This allows to obtain information about the wrist movement without flex sensors placed directly onto it. To get data describing more specific wrist movements, these values – along other descriptors – can be used to train a machine learning model, as it will be described in section 8.3.

Fig. 8.2 Numbering of the Myo armband EMG channels.

Even though using EMG data for tracking wrist movements requires a fair amount of processing, this solution is much less obtrusive than using the fingerless gloves described in chapter 6. Flex sensors have been a popular solution among new instrument builders for several years, with notable examples such as the 'Lady's glove' by Sonami (2006) and the more recent 'mi.mu glove'[10]. These sensors return a relatively stable signal without the need of much processing, but they have several downsides. They are relatively fragile and they need to be placed directly on the joint where the motion is taking place, which might be problematic in situations where obtrusion needs to be avoided, such as when playing traditional musical instruments.

EMG data also adds a sensing modality that does not depend on displacement in space (as with motion capture or – in most cases – with inertial sensors), which constitutes another layer of information useful for understanding gesture expressivity (Caramiaux et al., 2015b).

[9]To use this default groups the Myo needs to be worn with the LED logo pointing outwards and the USB port towards the wrist, as specified in the installation tutorial by Thalmic Labs

[10]http://mimugloves.com

8.3 Software and Mapping Approaches

For the mappings described in chapter 6 and the first version of Kineslimina, the data from the wearable sensors was processed and mapped using junXion[11] and Max[12]. Developed at STEIM (STudio for Electro-Instrumental Music) in Amsterdam by Frank Baldé, junXion incorporates the signal processing and mapping workflows that were developed through decades of groundbreaking practice at STEIM (Torre et al., 2016). JunXion graphical interface (Fig. 8.3) makes it easy to define explicit (Hunt et al., 2000) one-to-one, one-to-many, or many-to-one (Hunt and Kirk, 2000) mappings and accomplish basic mapping tasks such as scaling and thresholding. However, junXion also enables to programme complex, multi-layered mapping behaviours as it allows to define variables, complex mapping functions, and logical and mathematical operations. By grouping multiple mappings in states and using conditionals, it is also possible to alter several mappings at the same time. Indeed, these operations can be accomplished in other programming environments with relative ease and added flexibility. However, including junXion among the tools employed for composing Kineslimina has proved useful also from a research standpoint, since it concretely exposed and allowed to directly engage with some of the mapping procedures that have emerged from years practice-based research carried out at STEIM. As pointed out by Magnusson (2006) in his article about screen-based musical instruments, affordances and constraints can open up for different mental models in the musician and therefore affect compositional practice. The constraints and opportunities that characterise junXion are a product distilled from the practice-based research ethos promoted by STEIM researchers (Norman et al., 1998), still influential to this day (Torre et al., 2016). Baldé and the inventor and performer Michel Waisvisz (director of STEIM from 1981 to 2008, the year of his untimely death) collaborated on several projects, including The Hands (Waisvisz, 1985), one of the earliest and most influential digital musical instrument (Torre et al., 2016).

Using the terminology proposed by Van Nort et al. (2014), junXion (like other similar mapping tools or approaches) affords a *systems-oriented* perspective on mapping, which assumes that mapping is “a series of correspondences, or the out-of-time snapshot of input/output (I/O) control potential [. . .] represented by the classical “flowchart” paradigm that is ubiquitous in engineering” (Van Nort et al., 2014, p. 6–7).

The software solutions successively developed in KineToolbox (see section 8.3.1) and used for the pieces *11 Degrees of Dependence* (Visi and Miranda, 2016) and *Tuned Constraint* (Visi, 2016), favours instead a holistic approach akin to the *functional* view

[11] http://steim.org/product/junxion/
[12] https://cycling74.com

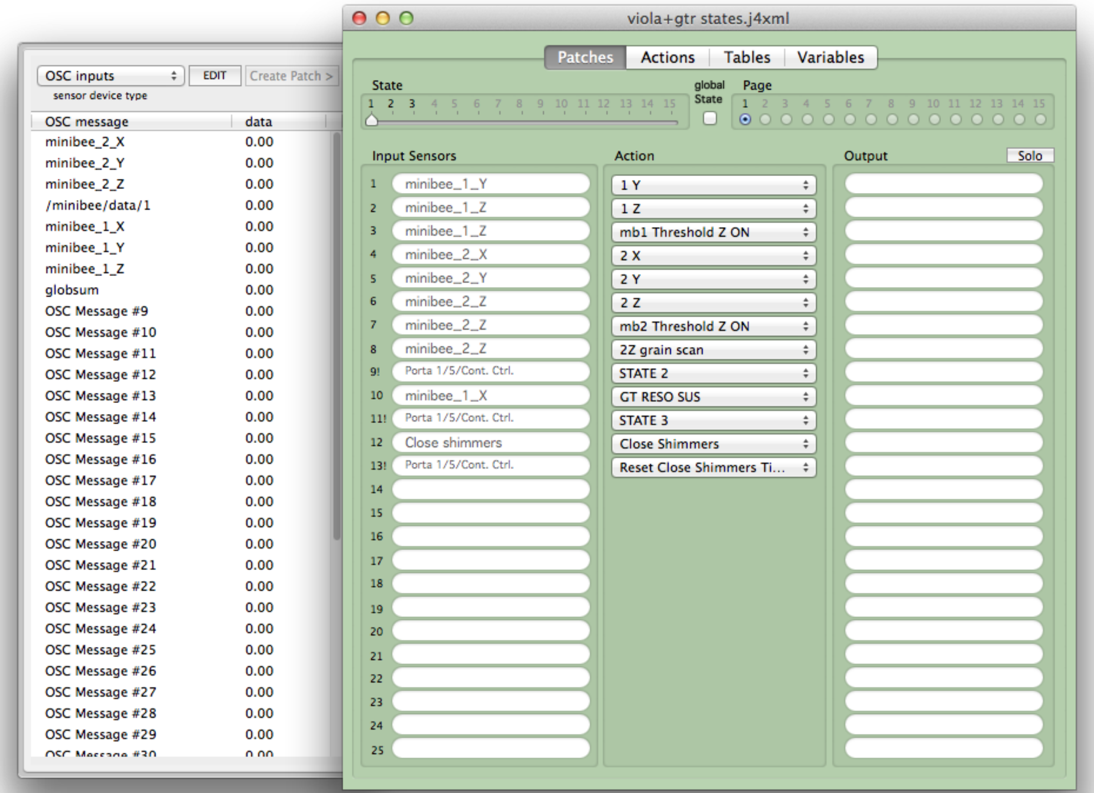


Fig. 8.3 Graphic user interface of junXion.

of mapping described by Van Nort et al. (2014). The functional point of view is defined by "the operations that associate different sets of variables, such as between control and sound-synthesis parameters." (Van Nort et al., 2014, p. 7). Here, this approach is implemented using machine learning to map a set of motion features to sets of waveguide synthesis (Smith, 1996, 1992) parameters, thus creating implicit mappings (Arfib et al., 2002; Hunt et al., 2000). In recent years, machine learning techniques have been increasingly adopted for creating mappings for the design of digital musical instruments (DMI) (Caramiaux and Tanaka, 2013, for an overview). Adopting a holistic workflow based on machine learning allows to pay much less attention to the actual value of synthesis and control parameters, and skip the calibration, rescaling, and explicit mapping phases typical of systems-oriented approaches. The software used in *Kineslimina* was completely rewritten to take advantage of this workflow. The piece was performed a third time at the Peninsula Arts Contemporary Music Festival[13] in February

[13] http://www.pacmf.co.uk

2016, this time using Myo armbands and the KineToolbox patches. Even though the piece was conceived using the older system, during rehearsals this approach proved to be more robust and quicker to configure.

8.3.1 KineToolbox

KineToolbox[14] is a collection of tools for sound interaction. It includes a set of Max patches and abstractions to quickly implement a holistic mapping workflow using machine learning and access various motion descriptors computed from IMU/MARG data. The patches in the toolbox are designed to be modular and to easily share data amongst each other.

The main input patch (fv.myofeatures2.maxpat) contains a set of abstractions useful for parsing the raw data from the Myo sensors and extract several features such as IMUQoM (see section 7.3.2), Jerkiness (see section 7.3.2), EMG MAV (EMG Mean Absolute Value, see section 9.6.2) and others[15]. The algorithm that handles the Myo is built upon the MuMYO[16] patch by Nymoen et al. (2015), adding several improvements over the original system. Firstly, the KineToolbox input patch includes various abstractions to calculate the aforementioned features. The data from each sensor on the armband and each extracted feature are also sent to a unique send object with a descriptive name. This way each feature can be easily accessed in other patches, making it easier to create a modular system that can be expanded and repurposed. Moreover, the patch allows to parse the data coming from multiple Myo armbands using ID numbers. The main abstraction uses changeable arguments (# sign) to parse the data of the Myo with the corresponding ID and prepend the ID number to all the send objects, thus allowing to easily access features of a specific Myo in other patches. Other improvements include the possibility to centre the yaw angle of the orientation data and the substitution of all the *multislider* objects used for visualisations with buttons for simple input monitoring. This was done in order to make the patch less CPU-intensive and optimise it for performance. Data visualisation is instead implemented in two separate patches: one to visualise all the raw data coming from the Myo corresponding to the selected ID number and another one to monitor specific features by typing the name of the corresponding send object in a text box (see Fig. 8.4).

[14] `https://github.com/federicoVisi/KineToolbox`

[15] The patch can be easily used to extract the same features also from other IMU/MARG sensors, provided that the acceleration, gyroscope, and orientation data is parsed correctly.

[16] `https://github.com/krisny/MuMYO`

A separate patch implements the machine learning procedure based on Support Vector Machines (SVM) (Burges, 1998) described in section 7.3.4. The user interface in presentation mode allows to monitor the input features sent from the input patch, and train the SVM classifier (part of the ml.lib library by Bullock and Momeni (2015)). Keyboard shortcuts to advance class numbers and record sample data, and a short delay time between the recording command and the actual data collection make the training procedure easier, especially when holding a musical instrument during rehearsals or before a performance. Once the model is trained and new input data is sent to the SVM, the patch returns class numbers and probability distributions that can be used to control sound parameters.

The functional, implicit mapping to the parameters of a flute physical model is implemented by loading JSON files containing sets of waveguide synthesis parameters using Max *pattr* system. Each set of parameters is associated to one of the classes of the machine learning model. The values of the probability distribution are used as weights to interpolate between the values of each parameters set, thus creating a many-to-many functional mapping from an m-dimensional input space (defined by the selected motion features) to an n-dimensional sound synthesis space (defined by the synthesis parameter set)[17].

Thanks to the modular design of KineToolbox, the synthesis engine could be easily replaced with other algorithms by exposing synthesis parameters to Max *pattr* system. The data from the SVM classifier can be sent to interpolate JSON files loaded in multiple sound synthesis and signal processing patches. External synthesis and sound processing algorithms can be also controlled in a similar fashion as long as they can communicate with Max through Open Sound Control (OSC) or MIDI protocols.

The features extracted in the input patch can also be used for more traditional explicit, systems-oriented mapping in conjunction with the functional approach based on machine learning. For example, the EMG MAV value (see section 9.6.2) is explicitly mapped to the breath pressure of the physical model, while all the other parameters are controlled as described above.

8.4 Case study: *11 Degrees of Dependence*

This section presents a composition for saxophone, electric guitar, wearable sensors, and live electronics that makes use of the tools described in this chapter so far. The piece, entitled *11 Degrees of Dependence*, explores the relationship between the performers and

[17] For more information on functional mapping approaches based on interpolation see section 9.7

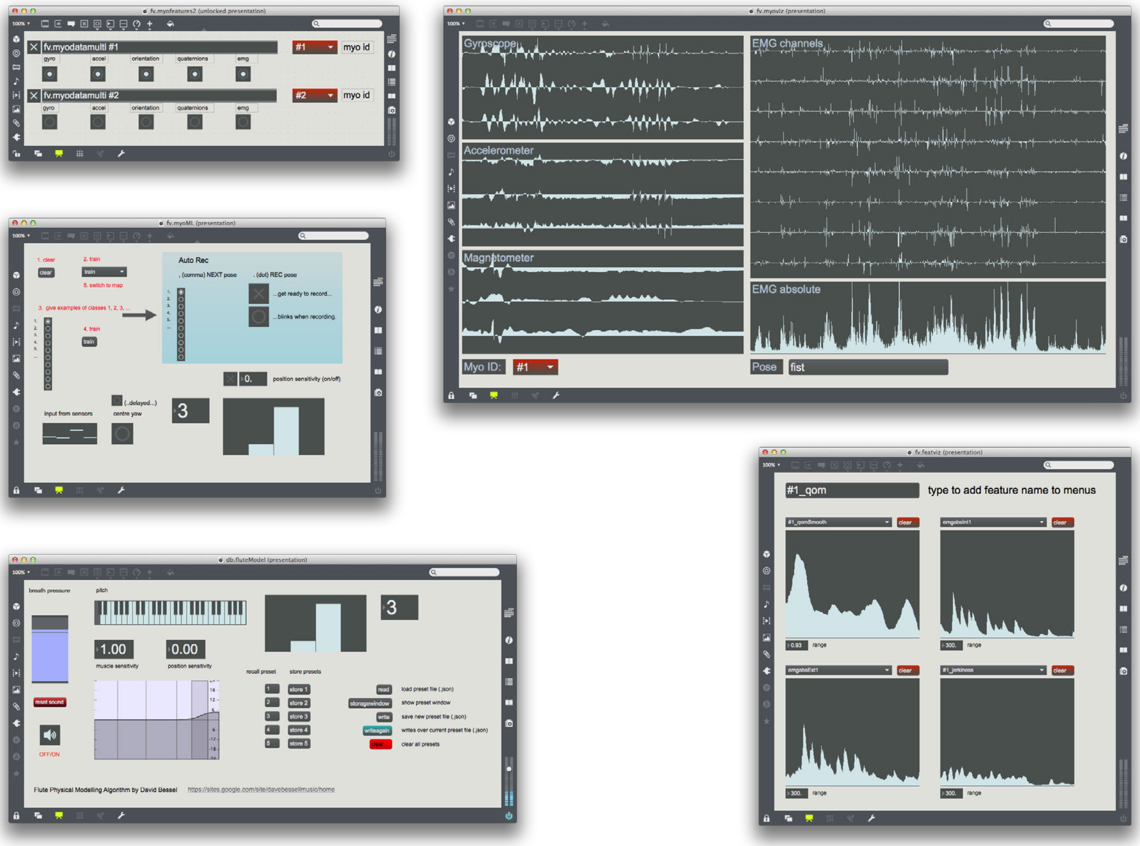


Fig. 8.4 Four patches of KineToolbox in presentation mode. Clockwise from top left: main input, Myo monitoring, features monitoring, physical model, machine learning.

their instruments, focusing on the constraints that instrumental practice imposes on body movement and a topological interpretation of the musician's kinesphere (Laban, 1966). The score includes symbols to notate movements, designed to be easily interpreted by musicians familiar with standard notation.

8.4.1 Instrumentation and Design

The piece is a duet for alto or soprano sax and electric guitar tuned in Drop C[18]. The sax player wears two Myo armbands using the KineToolbox patches to control the physical model whereas the guitarist wears the same devices to control granular synthesis and an electroacoustic resonator placed on the guitar headstock. Parameter mapping is done using the SVM implementation in KineToolbox. The data from the lateral (pitch) and longitudinal (roll) axes of the magnetometer are used as input to train the machine

[18]Open strings tuned CGCFAD from low to high.

learning model. Four ‘postures’ are then defined for both musicians. In the case of the sax player, these are:

- a ‘default’[19] performance position (named ‘Rest’) with arms comfortably by the side of the chest,
- gently leaning back, raising the saxophone with the elbows slightly open (named ‘Open’),
- leaning to the left with the right elbow slightly pointing outwards (named ‘Left’),
- leaning to the right with the left elbow slightly pointing outwards (named ‘Right’).

Each posture is then coupled with a set of synthesis parameters of the flute physical model. The Rest posture is paired with a clean sound with a clear fundamental frequency, the Open posture with a louder sound rich of breath noise, the Left posture adds overtones, and the right posture with a flutter tongued ‘frullato’ sound. This synthesised wind instrument sounds are designed to blend with the saxophone sound to generate a timbre with both familiar and uncanny qualities. The pitch played by the flute model is a C1, which is also the tonic of the piece.

The amount of noise fed into the physical model (or breath pressure) is controlled by the sum of the EMG MAV values of both arms. This implies that the amount of synthesised sound is constrained by the movement of the fingers operating the saxophone keys. Notes that require more tone holes to be closed – such as low notes for example – cause more muscular activity and thus louder sounds from the physical model. This design choice adds a component of interdependent, semi-conscious control to the performance creating a tighter coupling between the sounds of the saxophone and those of the flute model.

8.4.2 Score and Structure

11 Degrees of Dependence is structured in 3 parts, each of which contains scored themes at the beginning and the and a middle improvised section. The full score of the alto saxophone part can be found in appendix A. The score adopts conventional notation along with some custom symbols (printed in red) used to notate movement.

The main ideas for the piece were outlined in summer 2015, during my stay in New York City for a collaborative project at NYU Steinhardt. The software employed in the

[19]The sax players were asked to stand and hold the instrument comfortably as they were playing with no particular effort.

Fig. 8.5 Performance of 11 Degrees of Dependence, Harvestworks, New York City, 8 May 2016. Ana García Caraballos (Alto Sax, Myo armbands) and Federico Visi (electric guitar, Myo armbands). Full video of the performance: `https://youtu.be/QtWTq15phL4`.

piece was written during that period, and part of it is included in KineToolbox (see section 8.3.1). For the composition of the piece, I was inspired by music I have discovered while exploring New York music scene. In particular, the use of guitar harmonics and saxophone keyclicks to form repeating rhythmic patterns was inspired by the album *Dysnomia*[20] (Thirsty Ear Recordings, 2013, Erased Tapes Records, 2015) by Dawn of Midi[21], a Brooklyn-based acoustic jazz trio with members from India, Morocco, and Pakistan. In that album, Dawn of Midi employ rhythmic structures from North and West African folk traditions and extended techniques on piano, double bass, and percussion to create complex polyrhythmic structures. The idiosyncratic sound of *Dysnomia* is sometimes reminiscent of minimal music and experimental electronic music. The scales I used for the scored themes and the frequent use of odd time signatures were instead inspired by some of the live performances I attended at The Stone[22], which I frequently visited while in New York. Located in the East Village, about a twenty-minute walk from NYU Steinhardt, The Stone is a performance space dedicated to experimental and avant-garde music. The artistic programme is curated by John Zorn in collaboration with

[20] `https://dawnofmidi.bandcamp.com/album/dysnomia`
[21] `http://www.erasedtapes.com/artists/biography/27/Dawn+of+Midi`
[22] `http://www.thestonenyc.com`

various other musicians (mostly associated with the avant-garde music scene) doing weekly residencies. At the time of my stay in New York, Zorn was presenting his *Bagatelles*, a series of 300 tunes which were performed at the Stone every Sunday for six months by a different ensemble.

In 11 Degrees of Dependence, I used a minor mode with a minor second, an augmented fourth, and a major seventh. The electric guitar parts are mostly ostinato rhythmic patterns that frequently change time signature between 4/4, 5/8, and 6/8. When the guitar takes the lead in section 2, the saxophone plays a percussive pattern in 5/8 using keyclicks to accompany the guitar, thus switching roles. Movement was scored following the melodic phrasing, in order to pair a complete movement to a phrase or a small group of notes. This had the added benefit of making the piece easier to memorise, since it would be impractical to use a music stand during performance. The work on body movement was inspired by some of the concerts I have attended at The Stone, and also by the duo performances of Colin Stetson and Sarah Neufeld. In particular, the live rendition of the piece "The Rest of Us"[23] prompt me to pay attention to the cyclical body weight shifting movements of the musicians and the physical interplay between the two performers.

Fig. 8.6 shows an excerpt of the score of 11 Degrees of Dependence useful to give an example of how movement notation works in the piece. The red symbols represent the four postures defined during the training process: the circle corresponds to the Rest posture, the triangle to the Open posture, the arrowhead pointing left corresponds to the Left posture, and finally the arrowhead pointing right corresponds to the Right posture. The custom symbols used in the score were initially sketched by hand, then drawn in Adobe Illustrator[24] and saved as Scalable Vector Graphics (SVG). The SVG files were then imported in Avid Sibelius[25] as custom symbols to be easily used along with conventional notation symbols.

While the symbols indicate at which point in time the posture should be reached, the red lines show how the transition between the different postures should be articulated. These lines resemble other lines commonly found in conventional music notation. A straight line between two symbols means that the performer should start from the posture represented by the first symbol and progressively move towards the posture represented by the second symbol. The movement resulting from the transition between the postures should end in correspondence with the second symbol, thus following the rhythmic

[23] Video of Stetson and Neufeld's performance of "The Rest of Us" for Billboard: https://www.youtube.com/watch?v=ZNnKemIYoWs

[24] http://www.adobe.com/products/illustrator.html

[25] http://www.avid.com/sibelius

Fig. 8.6 Section 3a of the score of 11 Degrees of Dependence, showing the same sax phrase repeated twice with different movement articulations.

subdivision indicated in the staff. This is similar to a glissando, also notated using straight lines between note heads. A curved line between the posture symbols works instead analogously to a legato, meaning that the indicated posture quickly tied with the following one. For example, in the first bar shown in Fig. 8.6 the saxophonist starts playing the phrase from a Rest posture, progressively transitioning to a Left posture, which is reached in correspondence with the dotted minim in the third bar. The posture is held until the end of the note and then quickly tied to Rest posture. A sforzando under a posture symbol denotes that the posture is forcefully accentuated. For instance, in bar 32 shown in Fig. 8.6, the player raises the saxophone higher than usual, thus generating a louder synthesised sound due to the increased effort. The sinusoidal line between two vertically aligned posture symbols – as in bar 28 – indicates a smooth, cyclical alternation between the two postures for the duration of the line. This is somewhat similar to a trill or a mordent in conventional music notation, although the speed of the alternation can be much slower and is left to the performer's interpretation. In the case of bars 28 and 29 of 11 Degrees of Dependence, the sax players that performed the piece tended to perform the movement with cycles equivalent to a dotted minim, following the tactus of the music.

The excerpt in Fig. 8.6 includes the same saxophone phrase repeated twice, but with different movement articulations. This shows how different gestural performances of each repetition and different sounds generated by the physical model can create musical variation while maintaining the same melodic material. A video showing a rendition of this segment by Katherine Williams on soprano saxophone can be found here: `https://youtu.be/KC98tkkYse8?t=7m1s` (7:01 – 7:30).

8.4.3 On Scoring Movement

Several composers included performer's movements in their scores adopting various techniques. Notable examples are Mauricio Kagel music-theatre work (Laskewicz, 2008) and Karlheinz Stockhausen's *HARLEKIN* for clarinet (Marczak, 2009). Kojs (2011) discusses the notation of 'action-based music' grounded in ecological perception and enactive music cognition (see chapter 3 to read more about these theoretical accounts), noting that "A number of performers have reported to me that performing action-based music has facilitated a better understanding of their instrument and informed their interpretation of the standard repertoire." (Kojs, 2011, p. 66). Unconventional music notation has recently received a renewed research interest in academia (Battier et al., 2015; Hoadley et al., 2016). Particularly relevant are recent works on notating the gestural aspect of piano performance (Maestri and Antoniadis, 2015), and a custom notation system for a DMI inspired by wind instruments and with motion sensing capabilities (Mays and Faber, 2014). Communicating musical ideas involving both established practice and idiosyncratic designs is also an issue that has been recently addressed by researchers in the NIME[26] community (Green, 2016).

From a research standpoint, writing and performing 11 Degrees of Dependence as a short étude on instrumental music electronically augmented through the musicians' movements has several implications. As discussed above, a system for notating the performer's movements that is sufficiently accessible to be presented to musicians familiar with conventional notation was needed. This issue was addressed by designing a set of symbols that work in tandem with the motion-sound mapping strategy and tools. However, there are other implications that go beyond addressing this practical necessity. In their article about bringing the Magnetic Resonator Piano (an electronically augmented acoustic grand piano) to a larger community of musicians, Mcpherson and Kim (2012) address several issues related to the development of novel instrumental techniques that are based both on newly designed instrument augmentations and on established practice.

[26] International Conference on New Interfaces for Musical Expression, `http://www.nime.org`.

Fig. 8.7 Performance of 11 Degrees of Dependence, Peninsula Arts Contemporary Music Festival 2016, Plymouth, UK, 28 February 2016. Katherine Williams (Soprano Sax, Myo armbands) and Federico Visi (electric guitar, Myo armbands). Full video of the performance: `https://youtu.be/KC98tkkYse8`.

Rather than discussing the design and evaluation of a novel musical instrument, they focus on how novel instrumental practices can be brought to broader musical communities, and how this can have an impact on the design of the instruments itself. One of the main motivations for addressing these issues is effectively summarised in a remark by Jordà (2004, p. 326): "Many new instruments are being invented. Too little striking music is being made with them.". Establishing a continuing musical life for novel instrumental practices is crucial for for their success, and notation is certainly a means to facilitate this. To achieve this Mcpherson and Kim (2012, p. 26) recommend to "Demonstrate uniqueness, but connect to familiar models" and "Design for the first performance; then iterate."

11 Degrees of Dependence was composed and notated following similar motivations and criteria, as it is aimed at testing and further improving some of the theoretical assumptions and technical solutions described in this dissertation. Therefore, scored parts involving both conventional notation and body movements that can be read and performed consistently by different musicians are a means for connecting to familiar models, allowing for different interpretations and iterations. The piece premiered at the Peninsula Arts Contemporary Music Festival 2016, Plymouth, UK, 2016, with Dr

Katherine Williams[27] on soprano sax. This approach to notating movement has proved effective as the piece was performed by three different performers (see section 8.4.4). In the two most recent iterations, the score was sent to the performers beforehand, explaining the movement notations verbally via a video call. Two of the performers reported that they found it easier to memorise the score after they began to consider the melodic and gestural phrases as whole syntactic elements of the piece rather than two separate streams of information running in parallel.

This notation system can also be employed with musical instruments other than the saxophone. The fact that the symbols refer to points in the kinesphere of the musician and not to specific instrumental gestures allows for the applications of the same concepts to other instruments and pieces. As an example, the saxophone part of 11 Degrees of Dependence could be transposed for violin by rewriting the notes in a key that is appropriate for the violin range and by defining four new postures in relation to the violin performance movements. Each of the four new postures corresponds to one of the four symbols in the score, and the mappings with the synthesis parameters of the flute physical model are easily defined through machine learning. Thus, the notation system is flexible and can be employed with other instruments and in other compositions. The software tools are conceptually paired with the movement notation system, and allow to quickly create new mappings using the synthesis parameters defined previously, preserving the original sonic palette of the piece while changing the performance movements involved.

8.4.4 List of Performances of 11 Degrees of Dependence

- With Katherine Williams on soprano saxophone:
  - Peninsula Arts Contemporary Music Festival 2016, Plymouth, UK, 28 February 2016 (Fig. 8.7).
  - Nonclassical Club Night curated by Gabriel Prokofiev, International Festival For Artistic Innovation – iFIMPaC 2016, Leeds College of Music, Leeds, UK, 10 March 2016.
- With Ana García Caraballos on alto saxophone:
  - Creative Tech Week, Harvestworks, New York, US, 8 May 2016 (Fig. 8.5).
- With Lara Jones on alto saxophone:

[27] Katherine Williams' input as a performer and musicologist has been essential for the development of the piece.

– Music of Motion and Presence, Plymouth Art Weekender / Peninsula Arts, The House, Plymouth University, Plymouth, UK, 25 September 2016 (with the addition of Marco Frattini on percussion).

8.5 Further Applications: *Tuned Constraint*

The KineToolbox patches were also used for a piece for analogue synthesiser and Myo armbands entitled *Tuned Constraint* (Visi, 2016). The piece has some characteristics in common with 11 Degrees of Dependence as they use similar hardware and software configurations to make use of the movement of the instrumentalist to control waveguide synthesis parameters. Moreover, both works make use of the affordances and constraints of the instruments to design movement-sound interactions based on a many-to-many holistic mapping strategy. Whereas the interaction with the physical model by the saxophonist was constrained by the instrumental gestures involved in playing the saxophone, in Tuned Constraint the movement of the performer is instead constrained by classic modalities of interaction involved in traditional analogue synthesis. A fully analogue synthesiser usually affords the control of a limited number of parameters by operating knobs and switches on its front panel. Similarly to the example of the DJ turntable mentioned in section 3.3.3, this is arguably part of an emerging vocabulary of gestures involved in electronic music performance that is becoming part of a shared knowledge. As also noted by Magnusson (2010), some physical objects have easily detectable affordances – for example a knob invites turning – and affordances and constraints can be used for designing mappings for DMIs. The concept of constraint is also used by Mulder (2000) for the design of virtual control surfaces for DMIs, and Gurevich et al. (2012) analyse the role of interface constraint in facilitating the development of style and virtuosity in DMI performance. One of the constraints of fully analogue synthesisers is that each knob is usually hardwired to a single synthesis parameter. In other words, parameter mapping is defined by the manufacturer and cannot be altered without modifying the instrument.

The training procedure and synthesis parameters used for the saxophonist in 11 Degrees of Dependence (see section 8.4.1) were also adopted for Tuned Constraint, this time using different postures related to actions on the synthesiser front panel:

- both hands resting on the front panel of synthesiser,
- both hands lifted up at about the level of the head,
- turning a knob with the right hand,

- turning a knob with the left hand.

Tuned Constraint aims at exploring the constraints that this interface paradigm impose on the performer's gestural behaviour, and to use such constraints as constitutive expressive elements. The performer's body is the site were different electronic music interaction paradigms (the classic analogue paradigm based on knobs and switches and the holistic mapping strategy based sensors and machine learning) are actualised and juxtaposed. The performer's actions emerging from this process of synergy and contrast are integral parts of the musical experience of the audience, having an impact on liveness, immediacy, presence, and flow of the performance.

8.5.1 List of Performances of Tuned Constraint

- Practice Research Symposium, Plymouth University, Plymouth, UK, 5 February 2016 (Fig. 8.8).
- International Metabody Forum, Brunel University / Artaud Performance Centre, London, UK, 7 – 9 April 2016.
- ICLI 2016 – International Conference on Live Interfaces, University of Sussex, 2 July 2016 (Visi, 2016).

8.6 Summary and Comments

This chapter presented a set of software tools and hardware solutions employed to augment instrumental performance through the body movements of the musicians. These tools were employed in the composition and performance of two musical pieces, here used as case studies. A set of symbols to notate movement was designed to be used within the conventional western music notation system in order to be more accessible to musicians.

The piece *11 Degrees of Dependence* was performed in several occasions by three different sax players, showing that the notation system allows for consistent reiterations of the piece, leaving reasonable room for the interpretation of each individual performer. In fact, a purely unconventional graphical score was avoided in favour of an approach that allowed a certain degree of interpretative freedom given a set of constraints. In 11 Degrees of Dependence, the improvised parts in the middle of each scored section allow the performers to freely elaborate the melodic and gestural material of the scored parts.

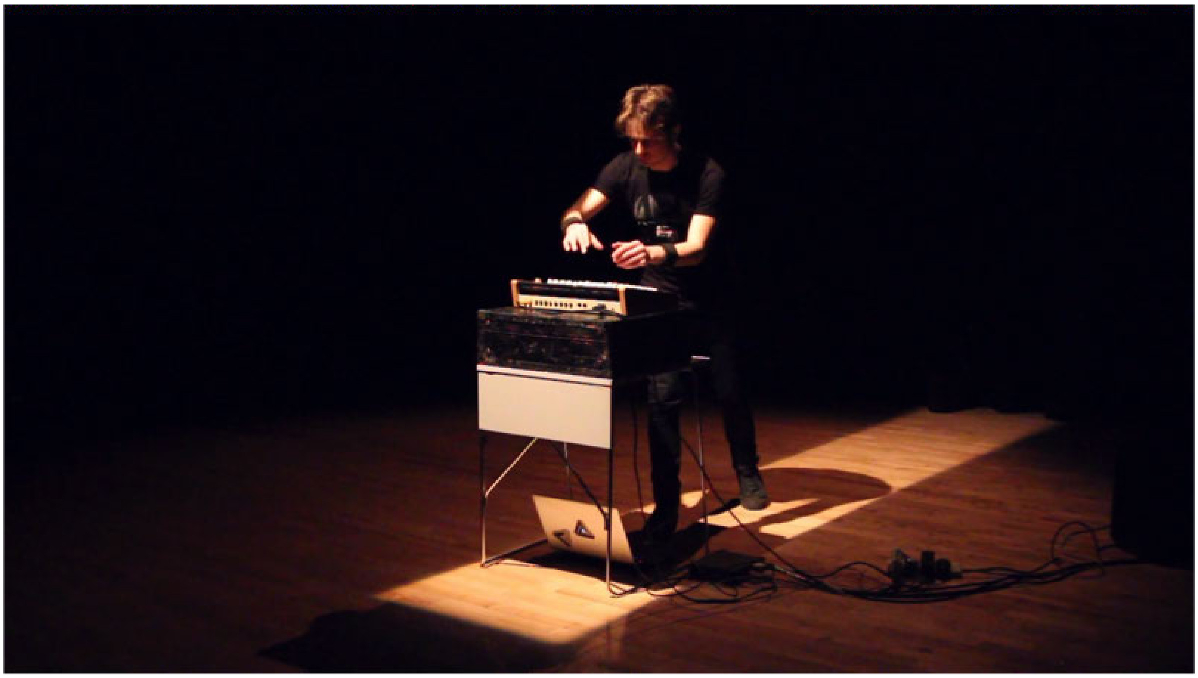


Fig. 8.8 Performance of Tuned Constraint, Practice Research Symposium, Plymouth University, Plymouth, UK, 5 February 2016. Federico Visi (analogue synthesiser, Myo armbands). Full video of the performance: `https://youtu.be/jdVw22D3NNM`.

The hardware/software system was used in two very different pieces. This shows that the system is flexible, as it can be repurposed and was not designed to address the needs of a single piece or performance. In fact – even though the choices in terms of hardware, software, and notation system were made to work in synergy – the key concepts of this holistic approach to mapping can be applied to other systems involving other hardware solutions, machine learning algorithms, and synthesis engines. The postures used for training the machine learning models can be interpreted as topologies of the kinesphere of the musicians, portions of the space around them with spatial and sonic relationships defined through the training procedure (Visi and Miranda, 2016). Moving between each postures constantly varies the synthesis parameters, generating movement/sound articulations. Here, a posture is defined as a point in the kinesphere towards which the body is moved, adding *goal-directedness* (cf. chapter 2) and thus giving intentionality and expressivity to the movement. From a different perspective, this resonates with HCI accounts that see posture as static conditions and gestures as dynamic transitions (Mulder, 1996).

The feedback from the musicians involved in the performance of 11 Degrees of Dependence also proved very useful for improving the usability of the KineToolbox patches and revisiting concepts subject of the theoretical and analytical parts of this dissertation, shedding light on issues of interaction design and the role of body movement in music performance.

Chapter 9

A Knowledge-based, Data-driven Method for Motion-sound Mapping

> There is more wisdom in your body
> than in your deepest philosophy.
>
> ---
>
> Friedrich Nietzsche
> *Thus Spoke Zarathustra*

9.1 Introduction and Motivation

This chapter presents a knowledge-based, data-driven method for using data describing action-sound couplings collected from a group of people to generate multiple complex mappings between the performance movements of a musician and sound synthesis. This is done by using a database of multimodal motion data collected from multiple subjects coupled with sound synthesis parameters. A series of sound stimuli is synthesised using the sound engine that will be used in performance. Multimodal motion data is collected by asking each participant to listen to each sound stimulus and move as they were producing the sound using the musical instrument they are given. Multimodal data is recorded during each performance, paired with the synthesis parameters used for generating the sound stimulus. The dataset created this way is then used to build a topological representation of the performance movements of the subjects. This representation is then employed to interactively generate training data for machine learning algorithms and define mappings for real-time performance. To better illustrate each step of the procedure, I describe an implementation involving clarinet, motion capture, wearable sensor armbands, and waveguide synthesis.

The main motivation behind this method is to make use of the music-related movement knowledge of a group of people to define motion-sound mappings for live interaction. The resulting mappings take advantage of the *ecological knowledge of action-sound couplings* of the group of people that participated to the multimodal motion data collection.

A topological representation of the motion data aims at providing an interpretation of what is shared and what is idiosyncratic among the participants, thus allowing to take into account commonalities and individualities when generating the training data for the machine learning model. The interactive method based on this representation allows to generate training data that either preserves certain peculiarities of individual subjects or is based on features shared by many participants. This gives more control over the transparency and intuitiveness of the resulting movement-sound mapping. This is particularly desirable for expressive applications such as music performance, where idiosyncrasies and non-obvious mappings could be deliberately employed for expressive purposes.

9.2 Background

This method is informed by several assumptions and approaches described throughout this thesis. From a theoretical point of view, the idea of an ecological knowledge in listening put forward by Godøy (2010) is central, particularly the assumption that studying the relationships between gestures and sound might contribute to our knowledge of how gestures help structure our experience of music. Moreover, Jensenius (2007) claims that ecological knowledge about action-sound couplings guide our perception of artificially created action-sound relationships. Thus, intuitive action-sound mappings should be modelled after action-sound couplings that have similar properties and are part of this ecological knowledge.

This method employs procedures and techniques for data collection and analysis typically adopted in experiments of music cognition and systematic musicology. Analogous procedures were adopted in the experiment described in chapter 5. Relevant previous works include the study by Godøy et al. (2006a) where they analyse video recordings of people miming piano-playing movements while listening to musical excerpts. This was done to explore the ability of listeners with different musical backgrounds to reproduce the geometry of and the dynamics of movements related to piano performance. Palmer et al. (2009) use motion capture to study performance expressivity and ancillary movements of clarinetists and Teixeira et al. (2015) use similar techniques to evaluate the gesture consistency of a group of clarinetists during several performances. However, in this case experimental procedures serve a different purpose. The main goal of this method is to obtain data describing how a group of people associate instrumental movements to certain musical sounds and observe shared and individual features of this movements. This data can be seen as a snapshot of the ecological knowledge that a group of individuals has of certain action-sound couplings related to the musical instrument they are 'performing' the given sound stimuli with. Rather than performing statistical analyses aimed at corroborating a hypothesis or exploring recurrent patterns in music-related movement, here the data is used for defining movement-sound interaction models that take advantage of specific information regarding the ecological knowledge of clarinet performance movements of a group of people. Differently from the work by Godøy et al. (2006a), here an actual musical instrument is used also by non-experts. This is done in order to obtain movements that are constrained by the physical affordances of the clarinet.

One early study focused on the movements of clarinet players was carried out by Wanderley (1999), who pointed out that ancillary gestures of the performer affect the sound produced by the instrument and therefore should also be taken in consideration in sound synthesis. Other studies dedicated to movement in clarinet performance were carried out by Desmet et al. (2012b) and Caramiaux et al. (2012) among others. Topological approaches to musical analysis include Topological Gesture Analysis by Naveda and Leman (2010) and the use of persistent homology for the analysis of musical score by Sethares and Budney (2014).

The work on the piece *11 Degrees of Dependence* (see chapter 8) brought my attention to several conceptual issues related to body movement in instrumental music performance and musical meaning. This method can be seen as an extension or generalisation of the insight on music and movement that working on that compositions has led to. However, the purpose of this method is not to substitute interaction design choices based on intuition with decisions informed by quantitative data. Rather, these techniques are aimed at providing a method for interpreting and utilising movement information for musical interaction. In fact, this approach may also be used in conjunction with other mapping strategies. To attempt an analogy with other techniques common in electronic music production, this method can be seen as a way of *sampling movement knowledge* from a group of individuals. This information can then be manipulated and repurposed, as it is common practice with audio samples. The size and composition of the group where this information is sampled from can also be considered a factor that can be deliberately manipulated by the composer. For example, one might be interested in working with movement data collected from a small group of individuals of specific ethnicity, gender, age group, etc. as opposed to analyse large databases that describe movements of a vast population.

9.3 Procedure Overview

The structure of this method is outlined in Fig. 9.1. The main steps of the procedure are:

- Design the sound stimuli using the synthesis engine that will be employed in performance.
- Present the sound stimuli to the group of participants, asking them to move as they were producing the sounds with the instrument they are given. Collect multimodal data during each performance.
- Extract features from the multimodal data and define a topological representation of the performances.
- Select a point in the topology to generate the corresponding motion data.
- Train a machine learning model for real-time interaction with the generated data and the synthesis parameters used for producing the sound stimulus.

In the following sections, I will describe the procedure in detail, both in general terms and by illustrating how the method was implemented using a clarinet, a synthesis engine based on a flute physical model, motion capture, and armbands with EMG and IMU/MARG sensors.

9.4 Design of Sound Stimuli

The set of sound stimuli to play back to the participants during data collection is designed by recording and editing sequences of synthesis parameters. Playing back these sequences allow the physical model to generate the desired sound.

In this implementation, the sound stimuli were designed and synthesised using the same physical modelling algorithm previously employed for composing and performing the piece *11 Degrees of Dependence* (see chapter 8). The flute model algorithm is based on waveguide synthesis (Smith, 1996, 1992) and was originally designed by Bessell (2011) and used in his piece *Ophidian*. Parameter sequences were deliberately designed to obtain

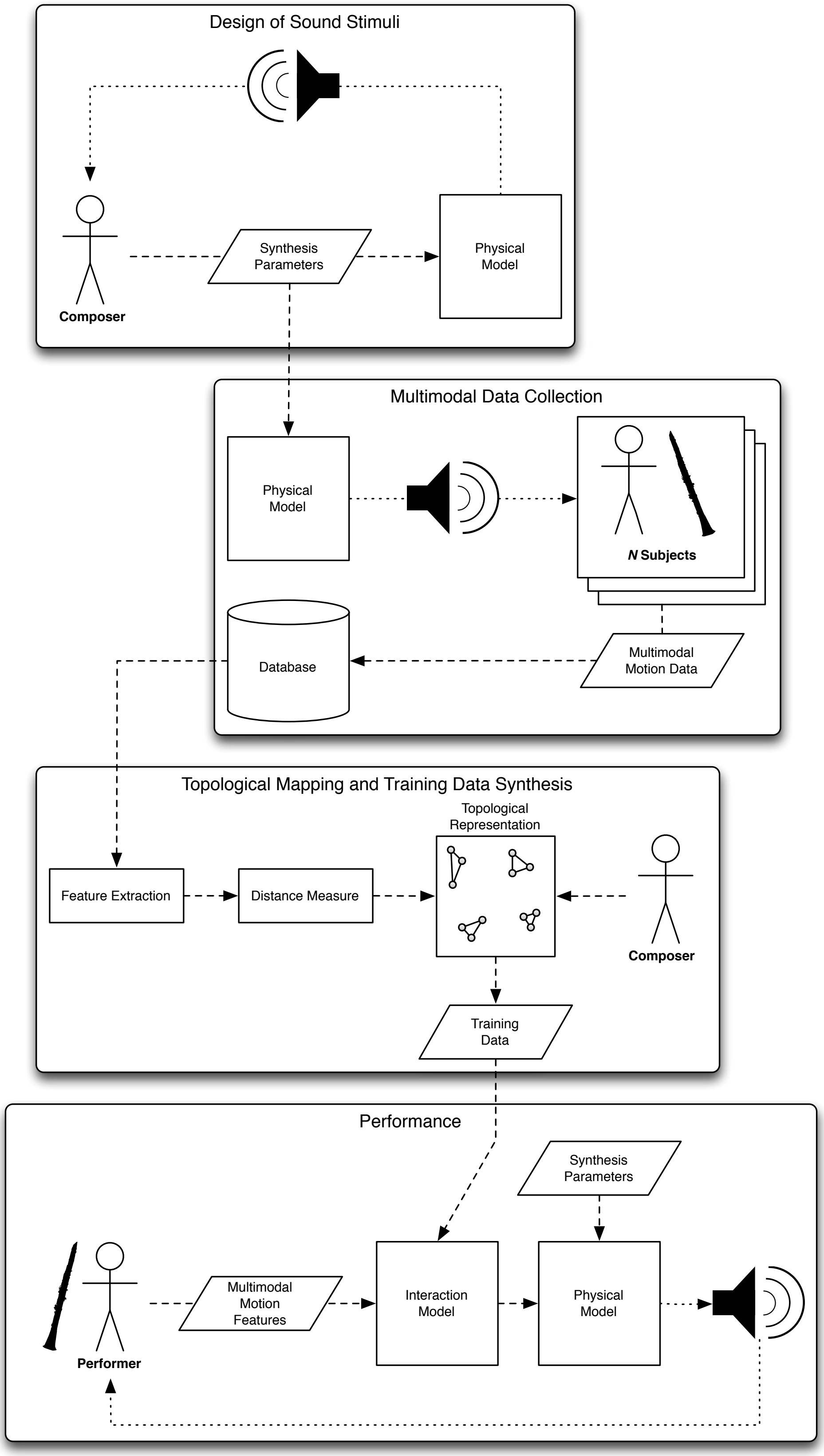


Fig. 9.1 Method structure and workflow.

some examples of various articulations that can be achieved using the flute model. It is important to note that obtaining sounds that closely resembled a real flute was not our goal. Rather, I aimed at obtaining sounds that preserve some timbral qualities of wind instruments but go beyond what a conventional wind instrument can achieve in terms of pitch, tone, and dynamics.

The sound stimuli were synthesised in Max[1]. Nineteen parameters of the physical model were exposed for control, therefore each sound sequence is made up of nineteen parameter envelopes. In order to obtain complex articulations, four sets of synthesis parameters were stored in a JavaScript Object Notation (.json) file. Each parameter set correspond to the values necessary to obtain a specific timbre: clean tone with clear fundamental frequency, tone with higher breath noise, overtones, and flutter tongue. The stimuli were created by interpolating said parameter sets over time. To do so, I used the orientation data from the IMU/MARG sensors of two Myo[2] armbands as input to a Support Vector Machine (Burges, 1998). The SVM was trained to interpolate between the parameter sets and obtain complex multi-parameter modulations. The audio output of the model, all the synthesis parameters, and the interpolation ratios of the parameter sets were recorded in separate tracks of a MuBu container (Schnell et al., 2009) and then saved as SDIF files. Doing so allows to re-synthesise the stimuli in real time using the parameter data and store the recorded audio for reference. Six stimuli of length between 10 and 28 seconds were eventually selected.

It is worth mentioning that any other strategy can be adopted to record the synthesis parameters of the stimuli: from manually designing each parameter envelope with a graphical editor to using other controllers to perform and record parameter modulations.

9.5 Multimodal Data Collection

Once the parameter sequences for the synthesis of the stimuli are defined, we can then ask a group of people to mime a performance of each sound using a musical instrument. Multimodal motion data is recorded during each trial. This results in a multimodal database containing data describing the performance movements each participant associated to each stimulus. This data aligned with the synthesis parameters used to generate the sound, also stored in the multimodal database.

In this section I will describe the procedure and the architecture of the system I used for collecting motion capture, IMU/MARG, and EMG data aligned with the synthesis

[1] https://cycling74.com
[2] https://www.thalmic.com

parameters of the physical model. Following that, I will explain the motivations behind the adoption of a multimodal approach, the marker configuration, and other design choices aimed at addressing technical and conceptual issues.

9.5.1 Task and Data Collection Procedure

Each participant was informed about the purpose of the data collection and asked to wear the Myo armbands and the rigid body markers as described in section 9.5.3.

The participant was then given a clarinet fitted with three reflective markers and with the reed removed. The embouchure of the clarinet was protected with a layer of food grade cling film, which was replaced after every individual session. This was done for hygienic purposes and in order to allow each participant to comfortably use the embouchure.

For each stimulus, each participant was first asked to carefully listen to the sound and imagine the movements they would do if they were to perform that sound using the clarinet. This listening phase could be repeated however many times the participant wanted (in most cases two times were sufficient). The participants were allowed to rehearse the movements while listening to the sound in order to find the movements and actions that, in their opinion, best matched the idea of performing that sound using the clarinet. After having sufficiently familiarised with the sound and decided the movements, the participant is asked to mimic a performance along the sound for three times. Participants were also instructed to perform each stimulus consistently (i.e. trying to perform the same performance movements they devised during the listening phase the best way they can throughout the three takes). In order to help the participant to start the performance synchronised with the sound, each stimulus playback was introduced by a four-beat count in at 120 bpm tempo[3]. During each take, all the data from the rigid bodies motion tracking and the EMG and IMU data of the Myo armbands were recorded in a single, multitrack MuBu container in Max. The synthesis parameters and audio of each stimulus were also recorded in the same container synchronously, and so was the click track the produced the count in before the stimulus. All the performances were also filmed using a Canon DSLR.

[3]The stimuli do not have evident rhythmic qualities therefore this tempo value is common to all the stimuli

9.5.2 Participants

Eight participants took part in the data collection phase of this project (7 male, 1 female, aged 24-53, average age: 31, SD of age: 9.0), which took place in the ICCMR Studio at Plymouth University in June 2016. All the participants had some musical background. An individual recording session lasting approximately half an hour was scheduled for each participant. During the session, the subject performed along each of the six stimuli three times, for a total eighteen takes per participant. A stool was placed nearby the capture area in case the participant needed to rest between takes.

It this important to point out that in this context whether the sample is representative of a larger population is not of major concern. The aim of the study is not to obtain statistically relevant results in order to validate a hypothesis. Rather, the goal is to generate mappings based on how a group of people moved to mime the performance of certain sounds. In fact, as pointed out in section 9.2, selecting a biased sample could be done deliberately in order to obtain datasets that yield peculiar mappings.

9.5.3 Apparatus

The movement of the participants were recorded using a multimodal set-up involving a six-camera optical motion capture system (OptiTrack Flex 3) and two Myo sensor armbands. The motion capture system was used to track seven rigid bodies, each one constituted by three or four reflective markers. The locations of the rigid bodies were as follows: head, left upper arm, right upper arm, left hand, right hand, sacral wand (hips), and clarinet (see Fig. 9.3 and section 9.5.4 for further details).

The 6 DOF data (3D position coordinates and orientation quaternions) was streamed to Max via Open Sound Control[4] protocol (OSC) using a custom MATLAB script. The participants also worn two Myo armbands, on either forearms. The devices streamed IMU/MARG 9DoF data and EMG data over a dedicated OSC port. The IMU data is constituted by 3D acceleration, 3D angular velocity, and 3D orientation Euler angles. The EMG data has eight channels per armband, which are numbered as described in Fig. 9.2. All the data was recorded synchronously in Max as described in section 9.5.1 at a sampling rate of 50 Hz.

The sound stimuli were re-synthesised in real time during the recording session using the previously recorded parameters and were played back via a pair of Genelec 8020C loudspeakers.

[4]http://opensoundcontrol.org/introduction-osc

Fig. 9.2 Numbering of the EMG channels on a Myo sensor armband.

9.5.4 Rigid Bodies

In motion capture terminology, a rigid body is a geometric arrangement of three or more markers that is defined as a single asset within a motion capture project. A rigid body is usually designed to have unique spatial relationships between its markers. This is to allow the system to recognise and track each rigid body as an individual labelled trajectory. A rigid body can be used to obtain position and orientation[5] relative to the rigid body's pivot point, whereas a single marker only returns position. The pivot point of a rigid body is by default located on the centroid of the markers that constitute it. However, the pivot point can also be adjusted by translating it to another location of the rigid body.

Advantages over single markers

I decided to use only rigid bodies and avoid single markers for several reasons. First of all, the purpose of this method is to create an interaction model for controlling a synthesis engine in real time. Single passive reflective markers are not suitable for the purpose, since the trajectory of a single marker would likely lose its label in case of a minor occlusion. Rigid bodies on the contrary retain their label also in case of occlusion as long as they are not deformed. Moreover, tracking of rigid bodies is generally more stable and less prone to dropouts since it relies on multiple markers that can help the tracking system in case one of the marker gets occluded (NaturalPoint Inc., 2016). Most motion

[5]Combined position and orientation is also known as 6 degrees of freedom (6DoF) data.

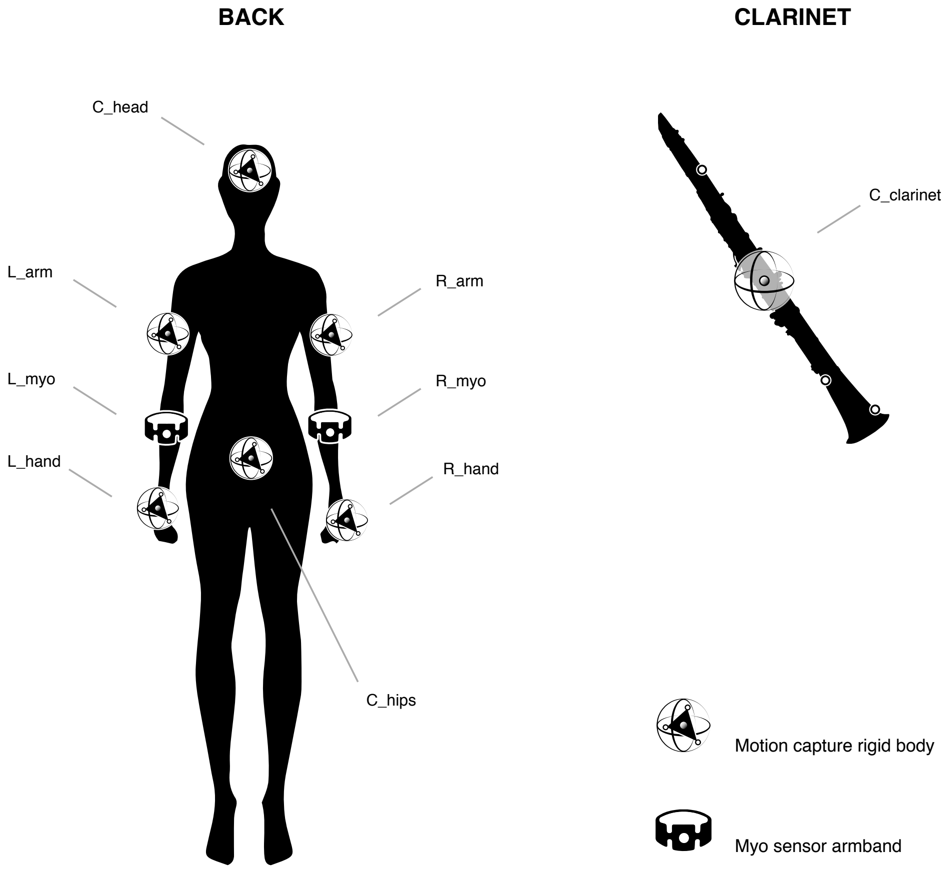


Fig. 9.3 Multimodal configuration: locations of the rigid bodies and the Myo sensor armbands. The clarinet was fitted with three single markers and then defined as a single rigid body with the pivot point at the centre.

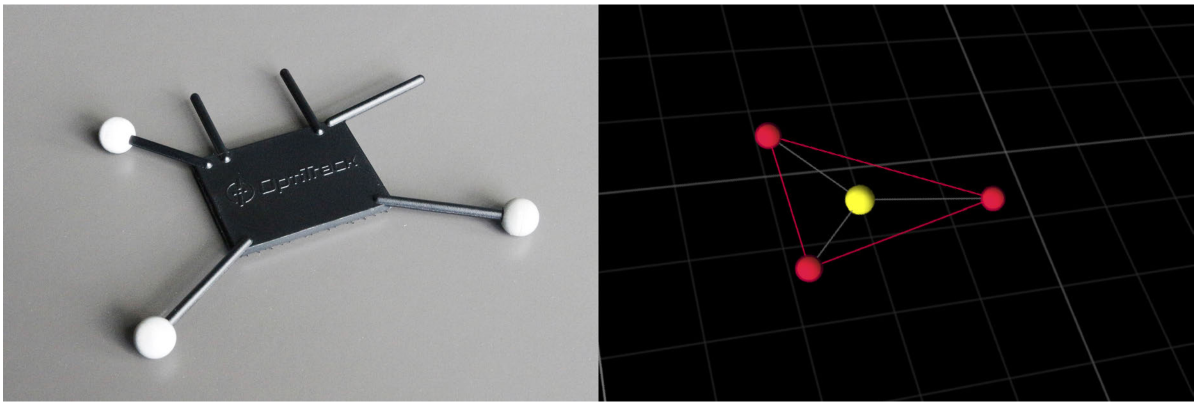


Fig. 9.4 OptiTrack rigid body and corresponding asset as visualised in OptiTrack Motive software. The pivot point is displayed in yellow.

capture systems are also capable of tracking so called 'skeletons' which are standard marker configurations designed for tracking a full human body. Using skeletons should minimise marker mislabeling in case of occlusion but again this solution is not suitable for the purpose of this method, since it would require the performer to wear a full mocap suit with a considerable number of markers attached. This could be acceptable when collecting data in a laboratory context but would be rather unwieldy in an actual performance situation. Moreover, when using skeletons, the marker configuration cannot be changed and therefore there would not be sufficient flexibility for designing an interaction model that is based on few selected body locations and movement features. Rigid bodies also have the advantage of providing 6DoF information. While single marker trajectories are defined by 3D positional data only, rigid body assets also carry information about orientation.

Design

Since this project required the use of multiple unique rigid bodies, I designed and 3D-printed our own custom rigid body bases to be used along with the standard OptiTrack hand rigid bodies. One of the main criteria for designing unique rigid bodies is to avoid geometrical congruency with marker configurations of other rigid bodies (NaturalPoint Inc., 2016). Also, as noted in the OptiTrack documentation (NaturalPoint Inc., 2016), the rigid body solver benefits from having a point not on the same plane as all the other markers. This makes the overall shape more unique and easier to track. Symmetry should be avoided as it may cause the rigid body asset to flip during capture.

Following these directions, I designed some custom rigid body bases using a 3D modelling software[6]. The size of the central rectangular base is similar to the one of OptiTrack rigid bodies, whereas all the beams have different angles and lengths and are not on the same plane as the base (see Fig. 9.5). The design includes six beams, therefore different marker configurations using three to six markers can be made using the same model. I used four retroreflective markers, whereas the standard OptiTrack rigid bodies I used mounted three. This was done for a dual purpose: marker count is another useful discriminant that helps the rigid body solver to recognise each rigid body. Moreover, having a redundant marker is helpful, as the system is less likely to lose tracking of the rigid body in case of occlusion (NaturalPoint Inc., 2016).

During tests, I noticed that our rigid bodies performed to some extent better than the standard OptiTrack hand rigid bodies. I assume that this is due to the use of a fourth marker and a non-planar design, which makes the markers less likely to align and

[6] Autodesk 123D: `http://www.123dapp.com`

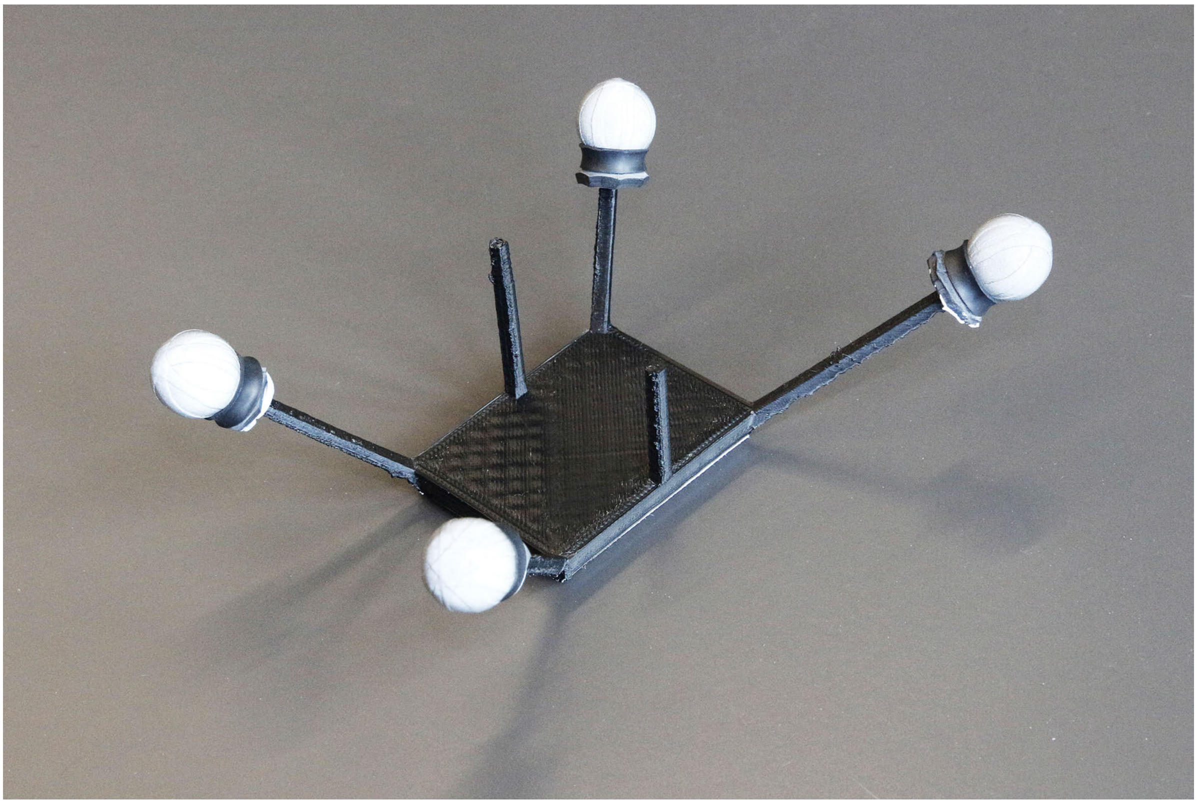


Fig. 9.5 Custom 3D-printed rigid body

occlude each other. Another factor that might have improved the performance is the slight distance of the reflective markers from the body and clothes of the participants, which may have helped to keep the rigid body visible by multiple cameras. Benchmarking the performance of these custom rigid bodies against that of the standard OptiTrack ones is beyond the scope of this project. However, useful information about the design of custom rigid bodies can be found in this article by Pintaric and Kaufmann (2008).

Locations

The marker configuration I adopted is constituted by seven rigid bodies, which were located on the head, left upper arm, right upper arm, left hand, right hand, sacral wand, and clarinet, as shown in Fig. 9.3. This configuration is deliberately focused on upper body, arms, and hips marker locations. This was done as the purpose of this implementation is to create an interaction model dedicated to instrumental and ancillary movements occurring in specific regions of the performer's body. Indeed, the method described here can also be implemented with other marker configurations in order to

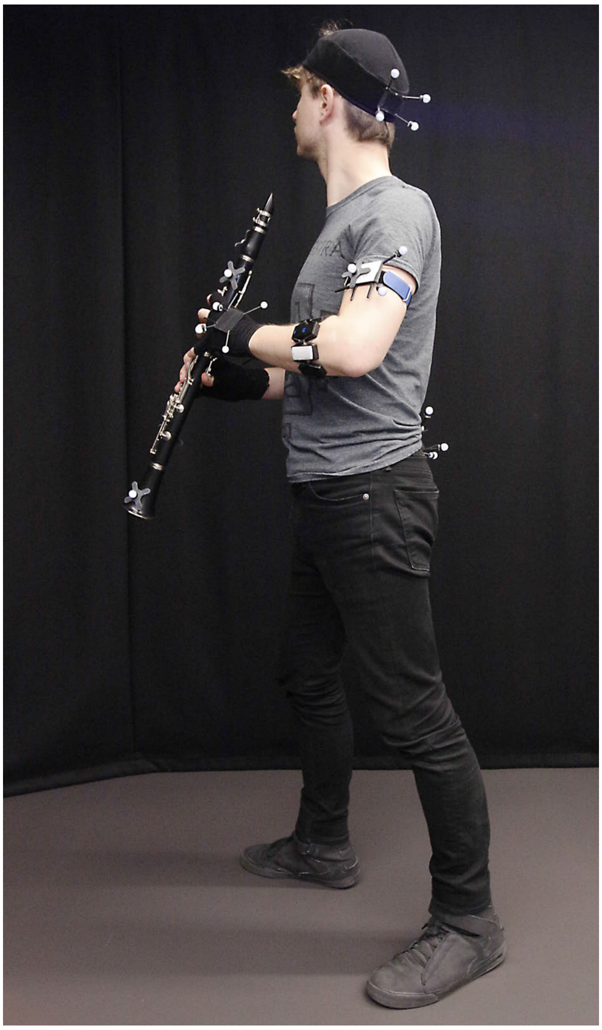


Fig. 9.6 Placement of rigid bodies and Myo armbands for multimodal data recording. The custom rigid bodies are on the back of the head and on the lower back.

create interaction models based on other movements, regions of the body, and motion features.

The marker configuration adopted for the studies described in chapter 5 is based on single markers, which are then processed offline. Using only rigid bodies has some practical implications. For example, a common marker placement for tracking head movements – which was adopted in chapter 5 and is also part of the standard Plug-in-Gait marker configuration (Vicon®, 2006) – is composed by four markers placed approximately on either temple and on the back of the head, roughly in a horizontal plane. It is common practice then to calculate the centroid (known as 'joint' in motion capture terminology) of these four markers and use the resulting data as a single 'head' marker. The same four markers can also be used to calculate rotation angles around their centroid, thus providing 6 DoF information. A single rigid body placed on the head of a subject returns 6 DoF data as well, with the added advantage of real-time trackability. In addition, a rigid body is less likely to generate mislabeled trajectories that need to be fixed manually during data preprocessing. If – for specific measurement reasons – the tracking point needs to be placed near the centre of mass of the head (i.e. inside the subject's skull, where the centroid of the four single markers would also be), this can be achieved by translating the pivot point of the rigid body to the desired position. However, using many rigid bodies at the same time may require higher computational power, since the solver algorithm has to calculate the position and angle of the pivot point from the incoming positional data of the rigid body markers at every frame. Moreover, rigid bodies may be a more compact solution in certain cases, but they cannot be used for smaller-scale applications such as finger and face tracking.

9.5.5 Multimodal Approach

In this case, tracking finger movements to determine when the clarinet keys are pressed would be challenging, as this would require placing small adhesive markers[7] on the fingers of the subjects. An additional marker on the metacarpal would be necessary as a relative reference, used to separate finger movements from arms and full body movements. This would be a rather obtrusive solution with several other downsides. Tracking small finger markers requires a higher-end – and therefore more expensive – motion capture camera set up, with either high-resolution cameras capable of tracking smaller markers from a distance or a higher camera count with few extra cameras placed closer to the subject (thus reducing the actual capture volume). Rigid bodies cannot be used, thus making it very difficult to track labeled trajectories in real time, as explained in section 9.5.4. Moreover, applying adhesive on the hands of every participant would be time-consuming, might feel uncomfortable, and is overall quite impractical in a performance situation.

Placing sensors on the instrument to track either keys or fingers movements would be a cumbersome and invasive solution, which would also limit the portability of the method and its implementation with other musical instruments and performers. Instead, the EMG data obtained from the Myo armbands can be used to estimate hands and fingers actions on the instrument with minor obtrusiveness and using a modality that is not dependent on displacement in space.

[7] In standard motion capture applications, adhesive markers are usually employed for tracking facial expressions. Finger tracking is often done using gloves or rings fitted with reflective markers. This solution is excessively obtrusive for instrumental music performance as it interferes considerably with instrument manipulation. An overview of standard motion capture markers produced by Natural Point can be found at http://www.optitrack.com/products/motion-capture-markers/.

The action of pushing keys on a clarinet involves small-scale displacement and is one of the main sound-modifying gesture of the clarinet's gestural vocabulary. Therefore, it makes sense to choose EMG data over motion capture to track finger movements on the clarinet, regardless of the other technical issues that using finger markers would imply.

9.6 Data Processing and Distance Measure

Following the data collection phase, the content of the multimodal database is processed and analysed. This section describes the steps necessary to represent the performance movements for each sound stimulus as a point in a feature space. This is done by selecting and extracting motion features from the data and adopting a distance measure to locate each trajectory on the feature space. The distance relationships between all the movements the participants performed in response to a single stimulus define a map: a gestural topology of the movement reactions to that sound. This representation is useful to understand differences and similarities between each performance in relation to the selected features. Points clusters in regions of the feature space would indicate that a group of participants have performed similar movements, whereas outliers might suggest a more idiosyncratic performance.

9.6.1 Multimodal Data Preprocessing

Each take was saved as a single SDIF file containing all the data from motion capture, sensor armbands, and physical modelling synthesis engine (sampled at 50 Hz), including the audio information of both the stimulus and the click track (sampled at 44.1 kHz). All the takes were then loaded in MATLAB for preprocessing. Each take was trimmed in order to remove the data recorded during the initial count in and after the the sound stimulus ended.

The quaternions describing the orientation of the rigid bodies were converted to Euler angles measured in radians. The phase of all the angles was then unwrapped. This was done in order to avoid big jumps in the data, which would affect the calculation of derivatives and confuse machine learning algorithms if the orientation data is used as input. Gaps in the motion capture data (which usually occur due to temporary occlusion of the markers) were filled using cubic interpolation. All the MoCap data was then smoothed using a Savitzky Golay smoothing FIR filter with a window length of seven samples and a polynomial order of two in order to obtain optimal combination of precision and smoothness as suggested by Burger et al. (2014).

The data from the Myo sensor armbands did not require gap filling and smoothing. The orientation data were centred at beginning of the stimulus and unwrapped as with the Euler angles of the rigid bodies.

9.6.2 Feature Selection and Extraction

After applying the preprocessing routine to all the data, I selected the locations of the body and the motion features on which I wanted to focus on for creating the interaction model for real time performance. In this implementation, the features selected are the Quantity of Motion of the clarinet rigid body and the envelope of the mean absolute value of the right arm EMG data.

Quantity of Motion (QoM) is a motion feature widely used in the study of body movement in music (Godøy et al., 2006b) and it is also employed for detecting affective states and emotion (Piana et al., 2013). Fenza et al. (2005) define Quantity of Motion (QoM) as the sum of Euclidean distances between successive points in a time window. Quantity of Motion was computed using the following equation:

$$QoM(t) = \sum_{k=0}^{N-1} \mid \|p_{t-k}\| - \|p_{t-k-1}\| \mid \tag{9.1}$$

Similarly to the version of Quantity of Motion for Inertial Measurement Units (IMUQoM) described in chapter 7, rotation angles can also be considered in order to obtain a descriptor that takes advantage of 6DoF information:

$$6DoFQoM(t) = \sum_{k=0}^{N-1} \beta_1 \mid \|q_{t-k}\| - \|q_{t-k-1}\| \mid + \beta_2 \mid \|p_{t-k}\| - \|p_{t-k-1}\| \mid . \tag{9.2}$$

As in IMUQoM, $\|q\|$ is the norm of the orientation quaternion, and β_1 and β_2 are weights to balance the contributions of translational and rotational motion data[8].

In either case, the values for each frame are summed over a time window of length N samples.

In their article on the evaluation of five different features extracted from the Myo EMG data, Arief et al. (2015) show that the EMG Mean Absolute Value (MAV) is the best performing feature for time series analysis.

[8]Note that the first addend of equation 9.2 is equivalent to Quantity of Rotation (QoR), defined previously in equation 7.3.

The EMG feature used in this implementation is the mean of the absolute values of all the eight EMG channels of the armband. This value can be considered an index of the overall muscular activity in the forearm. It is calculated as follows:

$$MAV = 1/N \sum_{k=1}^{N} |X_k|. \tag{9.3}$$

X is the vector with the EMG data and N is equals to the size of X, which in out case is equal to 8 since the Myo has eight EMG channels.

To smoothen this feature and make it more usable for real-time interaction, I calculated the upper envelope using shape-preserving piecewise cubic interpolation.

9.6.3 Distance Measure and Feature Space

A distance measure needs to be adopted in order to locate each performance in a two-dimensional feature space defined by the selected features. I used Dynamic Time Warping (DTW) to measure the distance between the feature vectors of each take. DTW returns the smallest distance between trajectories if warped, therefore it accounts for the fact that sequences might shift slightly forward or backward in time. It is widely used for time series analysis and classification tasks (Salvador and Chan, 2007; Senin, 2008) and for real-time gesture recognition (Gillian et al., 2011).

As described in section 9.5.1 above, the participants performed along each stimulus three times. Fig. 9.7 shows the locations in the feature space of all the performances along stimulus 5. The locations of the points were obtained by placing the mean of all the performances at origin of the axes and using DTW to calculate the distances of each point from the mean. In Fig. 9.7, the trials performed by the same participant are displayed with the same colour. By connecting the respective takes and filling the resulting triangular area we obtain a visual representation of the consistency of each participant across the three trials. The circle inside each triangle is the centroid obtained by averaging the locations of the three performances. Areas of the feature space with higher concentration of points suggest shorter distances between participants and therefore more similarity in the selected features.

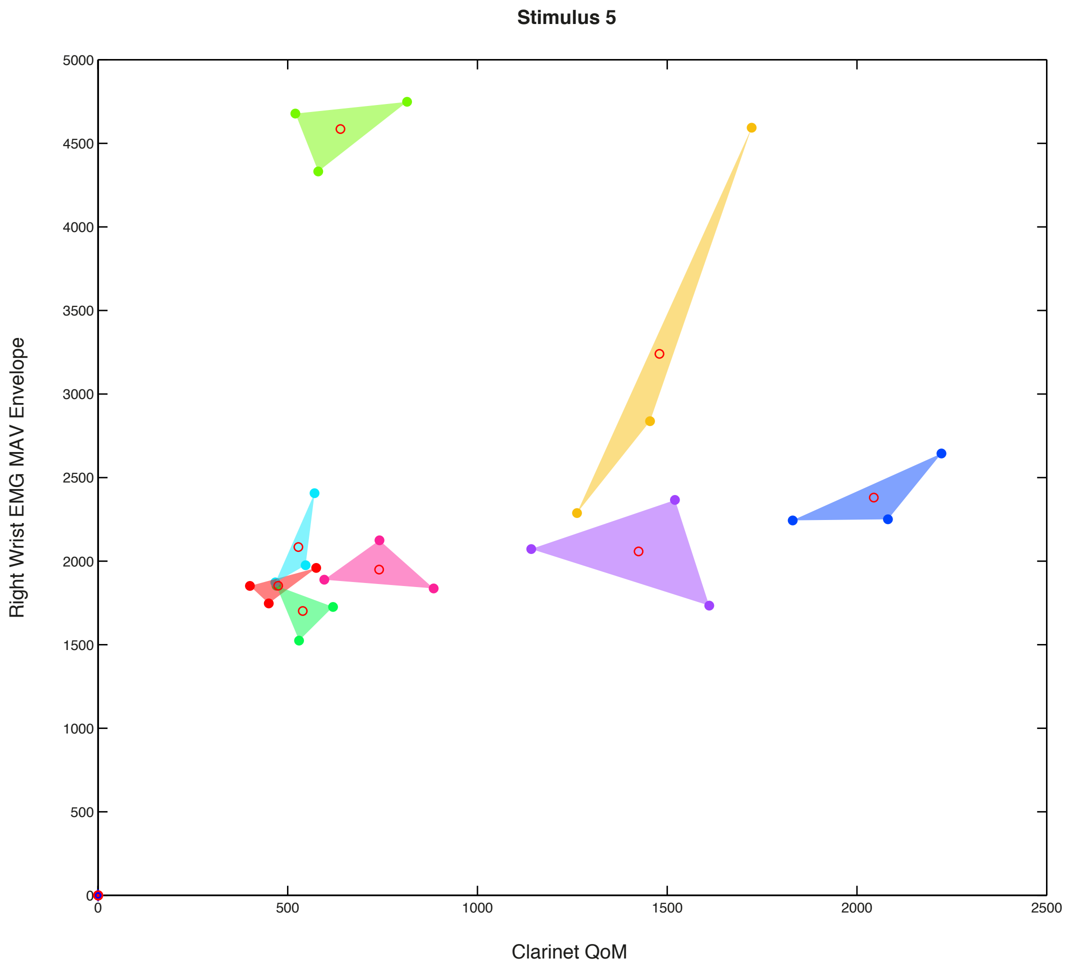


Fig. 9.7 Two-dimensional topological representation of the performances along stimulus 5. The vertices of each triangle correspond to the respective performances of each participants. The circle inside each triangle indicates the centroid.

9.7 Topological Representation, Interpolation, and Synthesis of Training Data

Van Nort et al. (2014) describe the topological perspectives of adopting a holistic conceptual approach to mapping between control and sound parameters. In particular, they focus on "functional properties related to a mapping's geometric and topological structure in the case of continuous, many-to-many mappings". Rather than focusing on the interconnections between individual parameters, a functional view of parameter mapping is concerned with the structural properties of the set of input and output parameters. This properties determine a *mapping topology*, which defines "the nature of the continuity, connectedness, and boundary definition in the mapping association (or associations) between control and sound sets.” (Van Nort et al., 2014, p. 7).

Functional approaches to mapping for controlling sound synthesis parameters have been implemented in several ways. Bowler et al. (1990) use a topological grid of points in a N-dimensional space and simplicial interpolation. Similar topological approaches were successively implemented in C++ by Goudeseune et al. (2001) and in Max by Van Nort et al. (2014). Arfib et al. (2002) adopt a mapping strategy that uses high-level parameters based on perceptual spaces to control many low-level synthesis model parameters. Interpolating sets of parameter values in order to allow the control of a high-dimensional sound model via a low-dimensional control space is also a key point of the method proposed by Momeni and Wessel (2003).

A holistic, topological approach to parameter mapping using machine learning was adopted in the pieces *11 Degrees of Dependence* and *Tuned Constraint* described in chapter 8. In this case, the topology defined using the distance measure described in the previous section is used to synthesise the training data for the machine learning model that will be used for real-time interaction. In order to represent each participant as a single point in the feature space, the features describing the three performances along the stimulus were averaged. The resulting time series and corresponding distances in the feature space are exported from MATLAB and loaded in Max.

Using the radial basis function interpolation (Freed et al., 2010) Max tool RBFI, each participant is represented as the centre of a Gaussian kernel (Momeni and Wessel, 2003) in the feature space. The largest graph on the top left corner of Fig. 9.8 shows the Gaussian kernels obtained from the same data used to plot the triangles shown in Fig. 9.7. Each point in this feature space corresponds to a temporal feature set obtained by continuously interpolating between the features extracted from the performances of the participants. This is done by using the distances from the selected point to calculate the contribution (weight) of the feature set of each participant at that point of the feature space. The grey sliders in the lower left of Fig. 9.8 display the weights at the cursor position. The two graphs on the left (blue and red) show the interpolated features at the corresponding point in the feature space selected using the cursor. In this case, the blue graph shows the interpolated QoM of the clarinet, whilst the red graph the interpolated envelope of the right arm EMG MAV.

From a practical point of view, this interactive display of the data allows to intuitively create a feature set that can be used in conjunction with the synthesis parameters of the stimulus to train a real-time interaction model. Displaying the data of each participant as a location in a feature space has the purpose of communicating certain topological qualities of the data. The distance relationships between the Gaussian kernels give higher-level information about how the participants moved to the sound that is useful for defining the interaction. For example, placing the cursor close to a cluster with several participants would result in training data that is closer to how the participants in that cluster performed along the stimulus. From this perspective, clusters can be considered as different 'styles' of performance movements along the same sound stimulus. Positioning the cursor away from clusters and closer to outliers would instead result in training data that is more *idiosyncratic*: representative of how a single individual reacted to and performed along the sound stimulus. Clusters – on the other hand – suggest that a group of participant performed the sound stimulus with a movement that has certain *shared* features. Ostensibly, moving closer to a highly-populated cluster would result in training data that would produce interaction models that are more transparent and intuitive for the population of that cluster. Conversely, data closer to less populated areas and outliers would instead lead to less predictable interactions. However, the characteristics of the resulting interaction model also depends on the chosen features and the machine learning algorithm used.

9.8 Interaction Model and Performance

The data generated by selecting a point in the feature space paired with the synthesis parameters of the corresponding sound stimulus can then be used to train a machine learning model for real-time interaction. Various algorithms can be adopted to do this. Some of the established supervised learning methods that make use of multiple examples to recognise gesture classes are based on Hidden Markov Models (HMM) (Françoise et al., 2012; Lucchese et al., 2012) and Support Vector Machines (SVM) (Nymoen et al., 2010; Piana et al., 2014). In particular, hybrid methods based on gesture templates and statistical recognition are employed for real-time continuous control using a limited number of training samples (Bevilacqua et al., 2010; Rajko et al., 2007).

Machine learning is also widely used for implementing interactive approaches to gesture-sound mapping. Fiebrink et al. (2009a) use supervised learning to build a training dataset from the gestures users perform along a musical score. Caramiaux et al. (2014a) use a perception-action loop as design principle for gesture-sound mapping in digital musical instruments. Françoise et al. (2013) employ HMM to conjointly model control and synthesis parameters.

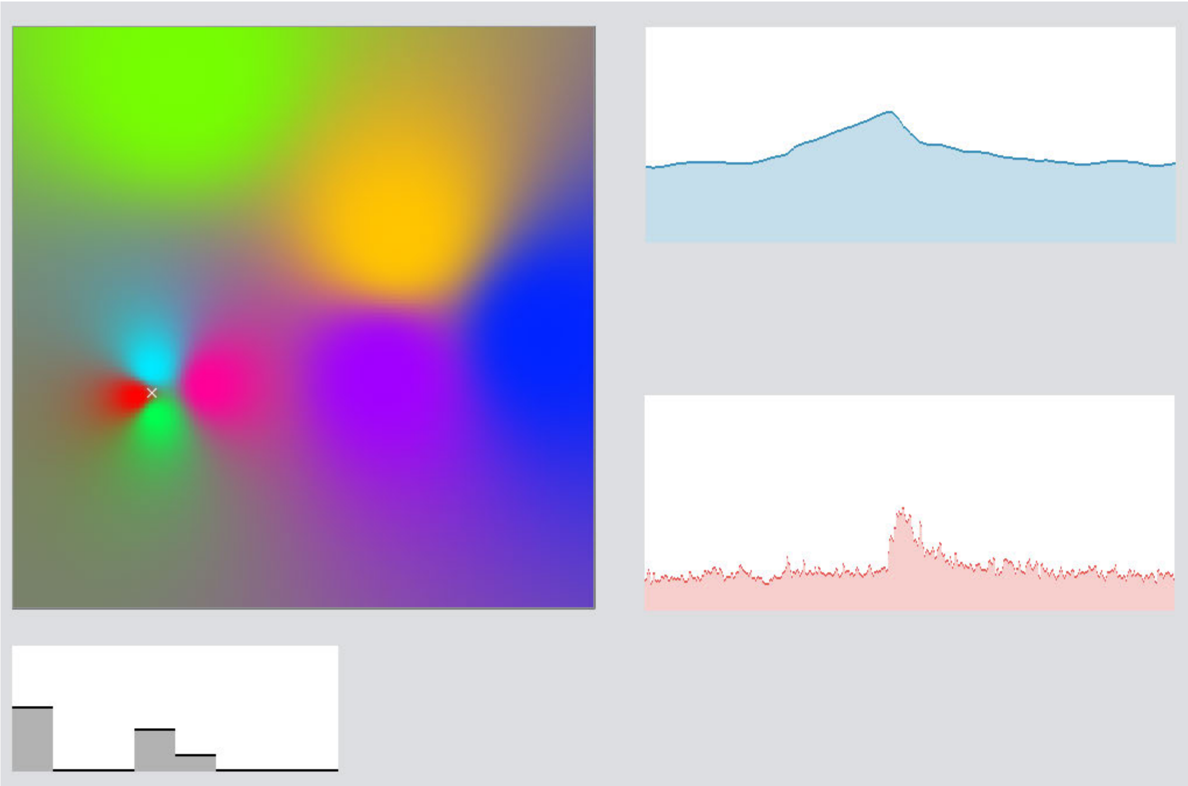


Fig. 9.8 Screenshot of the user interface in Max. Each Gaussian kernel in the upper left graph corresponds to a participant (compare with the locations of the triangles of the respective colour in Fig. 9.7). The white cross is the cursor used to select a point in the feature space. The grey sliders show the weights at the selected point. The two graphs on the right side show the interpolated features corresponding to the selected point.

In this instance, I used the Max implementation of the Gesture Variation Follower (GVF)[9] by Caramiaux et al. (2014b). The training data obtained with the procedure described in the previous section is used as a gesture template for GVF. In performance, the two features used for analysing the movements of the participants are calculated in real time and fed to GVF. The synthesis parameters that produced the stimulus the participants performed along to are loaded in the physical modelling patch. During performance, the GVF continuously outputs the temporal alignment with the gesture template. This information is used to move through the temporal dimension of the synthesis parameters of the sound stimulus, thus mapping the movements of the performer to sounds generated by the physical model.

For this implementation, GVF was chosen over other algorithms also because – by modelling the temporal information of the gesture template – it is able to detect

[9] https://github.com/bcaramiaux/ofxGVF

when the movement described in the template is performed backwards. Moreover, it allows for continuous interaction with the physical model without having to define the beginning and end of a gesture. These characteristics allow to obtain different sounds and articulations by interacting with the synthesis parameter space defined by the sound stimulus. Gesture-sound mapping can be easily redefined by repositioning the cursor on another region of the topological representation to generate new data to train the interaction model with. This allows to interactively explore the different mappings that the feature space defined by the movement of the participants affords.

The sound palette can be expanded beyond what can be achieved using the parameters of a single sound stimulus. This is done by repeating the procedure described in this and the previous sections for the other stimuli, thus generating additional templates for GVF paired with the synthesis parameters of the corresponding stimuli.

In performance, the amount of rigid bodies and sensor armbands worn by the performer can be reduced to the ones that are strictly necessary for extracting the selected motion features in real time. This results in a less cumbersome performance setup. For performing with the two features selected in this example, only the right Myo armband and the clarinet's rigid body markers would be necessary. However, the data from the other rigid bodies and sensors is not superfluous, as it can be stored and used for other compositions based on the same instrument family and for generating other interaction models based on other motion features and locations of the body.

9.9 Summary and Comments

I described a knowledge-based, data driven method for mapping music-related body motion data to sound synthesis. This method uses multimodal data obtained by tracking the movements of a group of people asked to move as they were 'performing' a set of sounds using a silent musical instrument given to them. The sound stimuli are produced using the synthesis engine that will be used for real-time performance and the synthesis parameters paired with the multimodal motion data collected during the participants' performances. This information is used to build a topological representation of the performance movements in a two-dimensional feature space. This representation provides information about shared and individual characteristics of the performances. Each point of the feature space corresponds to a set of features that – paired with the synthesis parameters of the corresponding sound stimulus – are used to train a machine learning model for real-time interaction with the synthesis engine.

This knowledge-based, data-driven approach stems from some of the assumptions resulting from the theoretical, analytical, and practical work carried out throughout this thesis. If movement affects our experience of music and therefore can be used as a musical feature, movement data contains information that can aid musical composition and performance. Considering the topic and scope of this thesis, this method was specifically designed for instrumental music performance. However, I do not exclude that a similar approach could be employed for creating empty-handed interactions or cross-modal mappings involving non-musical media.

The motion features selected for the clarinet example are relatively simple. The use of more sophisticated and higher-level features and more complex synthesis engines could in principle lead to more complex motion-sound interactions. The same procedure can be applied using features that involve more body locations or full-body motion descriptors. The implementation I described is limited to two features in order to allow for a clear representation of the performances in a two-dimensional feature space. This was done also for the purpose of creating a simple Graphical User Interface for selecting the areas in the feature space. Working with three or more features is also possible in principle. However, different ways of presenting the topology of the performance data should be adopted. A 3D representation is certainly a straightforward solution but more complex, multidimensional relations could be represented using alternative methods such as topological networks (Lum et al., 2013).

If using a larger dataset with many subjects, clustering algorithms can be employed to further generalise and simplify the model. Conceptually, clusters could be seen as different performance styles adopted by a number of individuals. Other distance measures can also be adopted to estimate the relationships between the different performances and unsupervised or semi-supervised training of deep networks might help to select the best features to train the model with. Fried and Fiebrink (2013) describe implementations of deep learning for cross-modal mapping. Future work can also go towards implementing coarticulation between gesture templates and synthesis parameters corresponding to different sound stimuli. Coarticulation in music has been extensively analysed by Godøy (2014) and implementations for real-time interaction are currently being developed by Bevilacqua et al. (2016).

Chapter 10

Postlude

Consciousness is only possible through change; change is only possible through movement.

ALDOUS HUXLEY
The Art of Seeing

I move, therefore I am.

HARUKI MURAKAMI
1Q84

This chapter concludes the dissertation by providing a summary of the results achieved throughout the research project in relation to the research aims and questions presented in chapter 1. Implications and future research perspectives are also discussed.

10.1 Summary and Contributions to Knowledge

This thesis have addressed the topic of body movements in instrumental music performance from multiple perspectives, adopting an interdisciplinary approach and a mixed methodology. Throughout the three main parts of the dissertation, methods techniques for the analysis of instrumentalists' movements and their use in composition and performance have been discussed. More in detail, the research questions presented in section 1.2.1 have been addressed as follows.

To what extent the movements involved in instrumental music performance are part of a shared knowledge of musical gestures, and how do they relate to other musical features?

This research question has been addressed from a theoretical standpoint and through empirical analysis. The experiments described in part II were designed on the grounds of the embodied accounts of music cognition delineated in part I.

The first exploratory experiment discussed in chapter 4 was aimed at observing how the variation of a musical feature within the piece affects the body movements of the performer. This initial study has shown non-obvious links between scored musical features and body movements of a viola player. In particular, the study has shown that ancillary movements do not always resonate with the variations of instrumental gestures. This suggests that other factors come into play, such as the difficulty of the task. The relationship between scored musical features and body movement was further explored in the composition used as case study in chapter 7. This practical study has shown how the multimodal qualities of the music mutually affect each other; the body of the performer is the medium where this dynamic entanglement takes place.

The results of the larger-scale experiment described in chapter 5 suggest that there is a shared knowledge of instrumental gestures shared also among people with no experience playing the violin. These results provide empirical evidence to support some of the theories discussed in part I, particularly the assumption that there is a shared *ecological knowledge* (Godøy, 2010) of an instrument's repertoire of sound-producing gestures (see section 3.3.3). In addition, the analysis of motion data collected during the experiment required the adoption and development of several motion data analysis techniques, which were described and documented.

If gestures and body movement play a key role in how we experience and understand music, how can they be employed as expressive elements in composition and performance?

In light of the theoretical framework described in part I and the empirical analysis in part II, the chapters in part III addressed this research question from several viewpoints. Chapter 7 addressed technical and compositional challenges by defining motion descriptors useful for expressive applications involving wearable motion-sensing devices. The solutions adopted in a piece for viola, electric guitar, and motion sensors entitled *Kineslimina* are then described and the piece is used as case study to further discuss the relationship between body movement, musical score, and motion data. Chapter 8 presented a set of software tools and hardware solutions employed to augment instrumental performance through the body movements of the musicians. These tools were employed in the composition and performance of two musical pieces, here used as case studies. In addition, a set of symbols to notate movement was designed to be used within the conventional western music notation system in order to be more accessible to classically trained musicians.

Can a multimodal embodied approach to musical meaning formation that take into consideration the ecological knowledge of a traditional musical instrument be used to inform effective mapping strategies?

Chapter 6 introduced the issue of parameter mapping in augmented instrumental music performance and presented an approach to mapping inspired by embodied music cognition and the studies of musical gestures. Concurrently, the chapter presented a layered mapping strategy informed by functional aspects of musical gestures. This strategy was implemented using various computable motion descriptors including Periodic Quantity of Motion (PQoM), a novel feature used to measure the resonance of body movement with musical rhythmic subdivisions.

Chapter 9 introduced a knowledge-based, data driven method for mapping music-related body motion data to sound synthesis. This method is informed by the theoretical assumptions presented in part I and the analysis techniques described in part II. This interactive tool uses a multimodal dataset obtained by tracking the movements of a group of people. The participants were asked to move as they were 'performing' a set of sounds using a silent musical instrument. The sound stimuli were produced using the same synthesis engine used in performance and the synthesis parameters were paired with the multimodal motion data collected during the participants' performances. This information is used to build a topological representation of the performance movements in a feature space. This provides information about shared and individual characteristics of the performances and it is a way of exploring aspects of the ecological knowledge of the instrument among the group of participants.

In order to address the research questions, the dissertation followed the broader research aims defined in section 1.2 by:

- Contributing to the theoretical discourse delineated in part I. Strengths and limitations of current theories in relation to the topic of instrumental music performance were discussed, emphasising the centrality of the *situatedness* of music performance and the key role of the environment in embodied accounts of music cognition. Functional categorisation of musical gestures has also been reviewed, suggesting the adoption of the concept of *functional components* in order to make the porous and flexible nature of functional categories more evident and explicit.

- Illustrating techniques and hardware/software solutions useful to analyse body movements in instrumental music performance. Part II focused on offline data analysis, describing the use of various motion descriptors – including Periodic Quantity of Motion – to study the relationship between body movements and other musical features. In addition to quantitative analysis, qualitative aspects related to the musical score and other characteristics of the music have also been included in the analysis.
- Composing brief pieces for traditional instruments and electronics that are used as case studies to explore the role of gestural aspects of instrument playing in the formation of musical meaning. These were also used as testing ground for concepts and tools. Chapters 7 and 8 describe the development of these pieces, documenting how conceptual and technical challenges have been addressed. Chapter 8 also describe a system for notating body movement within conventional music notation influenced by concepts of topology.
- Developing tools for composition and performance involving motion-sensing devices. In addition to the solutions employed for the study in chapter 5 and the pieces in chapters 7 and 8, chapter 9 illustrated a novel method for synthesising training data for machine learning models and solutions for multimodal motion sensing, including custom made 3D-printed rigid bodies for marker-based motion capture.

10.2 Implications and Future Research

Collectively, the these findings suggest that there is a shared embodied knowledge of body movements related to instrumental music performance. Attempting to use movement as a musical feature in composition and performance has brought to the fore the implications of dealing with the affordances and constraints of the medium itself, whether these are dictated by the relationship between the body of the performer and the musical instrument or defined by the conventions of standard music notation. Some of these constraints can certainly be eluded by designing entirely new instruments or notation systems. However, connecting to familiar models allowed to take advantage of the expertise of trained musicians and at the same time learn more about the role of body movement in established practice. Moreover, exploring the shared embodied knowledge of non-experts allowed to find out more about the musical experience of perceivers. Combining quantitative and qualitative observations has also led to the design of interaction models that take into consideration the environment in which they are situated.

Tackling the issue of creating meaningful cross-modal mappings has certainly wider implications. As computing comes to be a key part of the semiotic environment in which we live and paradigms such as mixed reality, pervasive computing, and calm technology become more widespread, interpreting data generated by human movement becomes a central issue. In these scenarios, persistent semi-conscious interactions are increasingly common and obtaining meaningful information from high-dimensional data is crucial for implementations human-centred machine learning. Exploring these issues in musical contexts has made evident that body movement is a modality of knowledge and expression, itself entangled to other modalities of human cognition. Movement carries information that can be transduced into digital data and research has shown that the inverse process is also viable (Lopes et al., 2015a,b).

Ultimately, the complex, multimodal qualities of music make it a remarkably effective testbed for concepts and technologies that will eventually define how we experience the world through movement. At the same time, wholly embracing movement as a musical feature will significantly impact our relationship with musical expression.

References

Altavilla, A., Caramiaux, B., and Tanaka, A. (2013). Towards Gestural Sonic Affordances. In Yeo, W., Lee, K., Sigman, A., H., J., and Wakefield, G., editors, *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 61–64, Daejeon, Republic of Korea. Graduate School of Culture Technology, KAIST. 51

Amelynck, D., Maes, P.-J., Martens, J. P., and Leman, M. (2014). Expressive body movement responses to music are coherent, consistent, and low dimensional. *IEEE transactions on cybernetics*, 44(12):2288–301. 51, 69, 90

Arfib, D., Couturier, J. M., Kessous, L., and Verfaille, V. (2002). Strategies of mapping between gesture data and synthesis model parameters using perceptual spaces. *Organised Sound*, 7(02):127–144. 120, 153

Arief, Z., Sulistijono, I. A., and Ardiansyah, R. A. (2015). Comparison of five time series EMG features extractions using Myo Armband. In *2015 International Electronics Symposium (IES)*, pages 11–14. IEEE. 150

Bååth, R. (2015). Estimating the distribution of sensorimotor synchronization data: A Bayesian hierarchical modeling approach. *Behavior Research Methods*, pages 1–12. 73

Bartenieff, I. and Lewis, D. (1980). *Body Movement: Coping with the Environment*. Routledge, London, 2002 edition. 108

Battier, M., Bresson, J., Couprie, P., Davy-Rigaux, C., Fober, D., Geslin, Y., Genevois, H., Picard, F., and Tacaille, A., editors (2015). *Proceedings of the First International Conference on Technologies for Music Notation and Representation - TENOR2015*. Institut de Recherche en Musicologie, Paris, France. 128

Bennett, P., Hinder, H., and Cater, K. (2016). Rekindling Imagination in Dementia Care with the Resonant Interface Rocking Chair. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems - CHI EA '16*, pages 2020–2026, New York, New York, USA. ACM Press. 112

Bergamini, E., Ligorio, G., Summa, A., Vannozzi, G., Cappozzo, A., and Sabatini, A. M. (2014). Estimating orientation using magnetic and inertial sensors and different sensor fusion approaches: Accuracy assessment in manual and locomotion tasks. *Sensors (Switzerland)*, 14(10):18625–18649. 94

Bessell, D. (2011). Ophidian and the Uncanny Valley. *International Computer Music Conference Proceedings*, 2011. 113, 138

Bevilacqua, F., Caramiaux, B., and Françoise, J. (2016). Perspectives on Real-time Computation of Movement Coarticulation. In *Proceedings of the 3rd International Symposium on Movement and Computing - MOCO '16*, pages 1–5, New York, New York, USA. ACM Press. 157

Bevilacqua, F., Van Zandt-Escobar, A., Schnell, N., Boyer, E. O., Rasamimanana, N., Françoise, J., Hanneton, S., and Rasamimanana, N. (2013). Sonification of the coordination of arm movements. *Proceedings of the International Conference on Multisensory Motor Behavior: Impact of Sound 2013*, 107(5):1. 111

Bevilacqua, F., Zamborlin, B., Sypniewski, A., Schnell, N., Guédy, F., and Rasamimanana, N. (2010). Continuous realtime gesture following and recognition. In *In Embodied Communication and Human-Computer Interaction, volume 5934 of Lecture Notes in Computer Science*, GW'09, pages 73–84. Springer Verlag, Berlin, Heidelberg. 154

Bowler, I., Purvis, A., Manning, P., and Bailey, N. (1990). On Mapping N Articulation Onto M Synthesiser-control Parameters. *International Computer Music Conference Proceedings*, 1990. 153

Broughton, M. and Stevens, C. (2009). Music, movement and marimba: an investigation of the role of movement and gesture in communicating musical expression to an audience. *Psychology of Music*, 37(2):137–153. 15, 18, 35

Brunner, T., Lauffenburger, J.-P., Changey, S., and Basset, M. (2015). Magnetometer-augmented IMU simulator: in-depth elaboration. *Sensors (Basel, Switzerland)*, 15(3):5293–310. 95

Bullock, J. and Momeni, A. (2015). ml.lib: Robust, Cross-platform, Open-source Machine Learning for Max and Pure Data. In *Proceedings of the 2015 NIME conference*, Baton Rouge, USA. 103, 122

Burger, B., Saarikallio, S., Luck, G., Thompson, M. R., and Toiviainen, P. (2013a). Relationships Between Perceived Emotions in Music and Music-induced Movement . *Music Perception: AnInterdisciplinary Journal*, 30(5):517–533. 15

Burger, B., Thompson, M. R., Luck, G., Saarikallio, S., and Toiviainen, P. (2013b). Influences of rhythm- and timbre-related musical features on characteristics of music-induced movement. *Frontiers in Psychology*, 4(183). 43

Burger, B., Thompson, M. R., Luck, G., Saarikallio, S., and Toiviainen, P. (2013c). Influences of rhythm- and timbre-related musical features on characteristics of music-induced movement. *Frontiers in psychology*, 4:183. 68

Burger, B., Thompson, M. R., Luck, G., Saarikallio, S. H., and Toiviainen, P. (2014). Hunting for the beat in the body: on period and phase locking in music-induced movement. *Frontiers in Human Neuroscience*, 8(November):1–16. 54, 56, 90, 149

Burger, B. and Toiviainen, P. (2013). MoCap Toolbox – A Matlab toolbox for computational analysis of movement data. In Bresin, R., editor, *Proceedings of the 10th Sound and Music Computing Conference*, pages 172–178, Stockholm, Sweden. KTH Royal Institute of Technology. 43, 56, 93, 95

Burges, C. J. C. (1998). A Tutorial on Support Vector Machines for Pattern Recognition. *Data Mining and Knowledge Discovery*, 2:1–43. 122, 140

Cadoz, C. (1988). Instrumental gesture and musical composition. In *Proceedings of the International computer music conference*, Cologne, Germany. 26

Cadoz, C. and Wanderley, M. M. (2000). Gesture - Music. In Wanderley, M. M. and Battier, M., editors, *Trends in Gestural Control of Music*, pages 71–94. IRCAM - Centre Pompidou, Paris. 26, 27, 28, 35

Camurri, A., Coletta, P., Demurtas, M., Peri, M., Ricci, A., Sagoleo, R., Simonetti, M., Varni, G., and Volpe, G. (2007). A Platform for Real-Time Multimodal Processing. In *Proceedings International Conference Sound and Music Computing*, number July, pages 11–13. 97

Camurri, A., Lagerlöf, I., and Volpe, G. (2003). Recognizing emotion from dance movement: comparison of spectator recognition and automated techniques. *International Journal of Human-Computer Studies*, 59(1-2):213–225. 90

Camurri, A., Mazzarino, B., Ricchetti, M., Timmers, R., and Volpe, G. (2004a). Multimodal Analysis of Expressive Gesture in Music and Dance Performances. In *Gesture-based Communication in Human-Computer Interaction, LNAI 2915*, number 1990, pages 20–39. 112

Camurri, A., Mazzarino, B., and Volpe, G. (2004b). Analysis of Expressive Gesture: The EyesWeb Expressive Gesture Processing Library. In Camurri, A. and Volpe, G., editors, *Gesture Workshop*, volume 2915 of *Lecture Notes in Computer Science*, pages 460–467. Springer. 60, 83

Camurri, A., Mazzarino, B., and Volpe, G. (2004c). Expressive interfaces. *Cognition, Technology & Work*, 6(1):15–22. 91

Camurri, A., Poli, G. D., Leman, M., and Volpe, G. (2001). A Multi-layered Conceptual Framework for Expressive Gesture Applications. *Proceedings of MOSART: Workshop on Current Directions in Computer Music*, (1):29–34. 31, 80

Camurri, A. and Volpe, G. (2011). Multimodal Analysis of Expressive Gesture in Music Performance. In Solis, J. and Ng, K., editors, *Musical Robots and Interactive Multimodal Systems*, pages 47–66. Springer Berlin Heidelberg. 60, 98, 100, 104

Camurri, A., Volpe, G., De Poli, G., and Leman, M. (2005). Communicating expressiveness and affect in multimodal interactive systems. *IEEE Multimedia*, 12(1):43–53. 83

Caramiaux, B., Altavilla, A., Pobiner, S. G., and Tanaka, A. (2015a). Form Follows Sound: Designing Interactions from Sonic Memories. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems - CHI '15*, pages 3943–3952, New York, New York, USA. ACM Press. 98

Caramiaux, B., Bevilacqua, F., and Tanaka, A. (2013). Beyond Recognition: Using Gesture Variation for Continuous Interaction. *CHI '13 Extended Abstracts on Human Factors in Computing Systems*, pages 2109 – 2118. 103

Caramiaux, B., Donnarumma, M., and Tanaka, A. (2015b). Understanding Gesture Expressivity through Muscle Sensing. *ACM Transactions on Computer-Human Interaction*, 21(6):1–26. 25, 118

Caramiaux, B., Françoise, J., Schnell, N., and Bevilacqua, F. (2014a). Mapping Through Listening. *Computer Music Journal*, 38(3):34–48. 154

Caramiaux, B., Montecchio, N., Tanaka, A., and Bevilacqua, F. (2014b). Adaptive Gesture Recognition with Variation Estimation for Interactive Systems. *ACM Transactions on Interactive Intelligent Systems*, 4(4):1–34. 155

Caramiaux, B. and Tanaka, A. (2013). Machine Learning of Musical Gestures. *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 513–518. 120

Caramiaux, B., Wanderley, M. M., and Bevilacqua, F. (2012). Segmenting and Parsing Instrumentalists' Gestures. *Journal of New Music Research*, 41(1):13–29. 137

Carlsson, G. (2009). Topology and data. *Bulletin of the American Mathematical Society*, 46(2):255–308. 102, 112

Chadefaux, D., Wanderley, M. M., Le Carrou, J.-L., Fabre, B., and Daudet, L. (2012). Experimental study of the musician / instrument interaction in the case of the concert harp. In *Proceedings of Acoustics 2012*. 41, 48

Chemero, A. (2009). *Radical embodied cognitive science*. MIT Press. 14

Clark, A. (2008). *Supersizing the Mind*. Oxford University Press. 30

Clarke, E. F. (2005). *Ways of Listening: An Ecological Approach to the Perception of Musical Meaning*. Oxford University Press, New York City, NY, United States. 33, 50, 51, 107, 111

Coorevits, E., Moelants, D., Maes, P.-J., and Leman, M. (2014). Studying the effect of tempo on music performance: a multimodal approach. In *9th Conference on Interdisciplinary Musicology CIM14*, Berlin. 68

Coorevits, E., Moelants, D., Östersjö, S., and Gorton, D. (2015). Decomposing a Composition: On the Multi-layered Analysis of Expressive Music Performance. In Aramaki, M., Kronland-Martinet, R., and Ystad, S., editors, *Proceedings of the 11th International Symposium on Computer Music Multidisciplinary Research*, Plymouth. 91

Costalonga, L. (2009). *Biomechanical Modeling of Musical Performace: A Case Study of the Guitar*. PhD thesis, School of Computing, Communications and Electronics, University of Plymouth. 81

Cox, A. (2001). The Mimetic Hypothesis and Embodied Musical Meaning. *Musicæ Scientiæ*, 5(2). 30

Dahl, L. (2014). Triggering Sounds from Discrete Air Gestures: What Movement Feature Has the Best Timing? In Caramiaux, B., Tahiroglu, K., Fiebrink, R., and Tanaka, A., editors, *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 201–206, London, United Kingdom. Goldsmiths, University of London. 51

Dahl, S., Bevilacqua, F., Bresin, R., Clayton, M., Leante, L., Poggi, I., and Rasamimanana, N. H. (2010). Gesture in performance. In Godøy, R. I. and Leman, M., editors, *Musical Gestures: Sound, Movement, and Meaning*. Routledge. 48, 81, 85

Davidson, J. W. (1993). Visual Perception of Performance Manner in the Movements of Solo Musicians. 91

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(3):381–401. 91

Davidson, J. W. (2012). Bodily movement and facial actions in expressive musical performance by solo and duo instrumentalists: Two distinctive case studies. *Psychology of Music*, 40(5):595–633. 91

Davies, H. (2017). Electronic instruments. *Grove Music Online*. 31

Delalande, F. (1988). La gestique de Gould: éléments pour une sémiologie du geste musical. *Glenn Gould Pluriel*, pages 85–111. 26, 27

Dempster, W. T. and Gaughran, G. R. L. (1967). Properties of body segments based on size and weight. *American Journal of Anatomy*, 120(1):33–54. 69

Desmet, F., Nijs, L., Demey, M., Lesaffre, M., Martens, J.-P., and Leman, M. (2012a). Assessing a clarinet player's performer gestures in relation to locally intended musical targets. *Journal of New Music Research*, 41(1):31–48. 41

Desmet, F., Nijs, L., Demey, M., Lesaffre, M., Martens, J.-P., and Leman, M. (2012b). Assessing a Clarinet Player's Performer Gestures in Relation to Locally Intended Musical Targets. *Journal of New Music Research*, 41(1):31–48. 90, 91, 137

Di Scipio, A. (2003). 'Sound is the interface': from interactive to ecosystemic signal processing. *Organised Sound*, 8(03). 80

Doati, R. (2004). L'apparizione di tre rughe (2001-2004), for guitar, tape, live electronics and EyesWeb interactive system (composition). 31

Dogantan-Dack, M. (2015). *Artistic Practice as Research in Music: Theory, Criticism, Practice*. Ashgate Publishing, Ltd. 6

Elmezain, M., Al-Hamadi, A., and Michaelis, B. (2009). Hand Gesture Spotting Based on 3D Dynamic Features Using Hidden Markov Models. In Śl\k ezak, D., Pal, S., Kang, B.-H., Gu, J., Kuroda, H., and Kim, T.-h., editors, *Signal Processing, Image Processing and Pattern Recognition*, volume 61 of *Communications in Computer and Information Science*, pages 9–16. Springer Berlin Heidelberg. 84

Fenza, D., Mion, L., Canazza, S., and Rodà, A. (2005). Physical movement and musical gestures: a multilevel mapping strategy. In *Proceedings of Sound and Music Computing Conference*, Salerno, Italy. 83, 99, 150

Fiebrink, R., Cook, P. R., and Trueman, D. (2009a). Play-Along Mapping of Musical Controllers. In *Icmc 2009*, pages 61–64. 154

Fiebrink, R., Trueman, D., and Cook, P. R. (2009b). A Meta-Instrument for Interactive , On-the-fly Machine Learning. pages 280–285. 103

Fink, G. A. (2008). *Markov Models for Pattern Recognition: From Theory to Applications*. Springer-Verlag, Berlin Heidelberg. 84

Flash, T. and Hogan, N. (1985). The coordination of arm movements: an experimentally confirmed mathematical model. *J. Neurosci.*, 5(7):1688–1703. 97

Françoise, J., Caramiaux, B., and Bevilacqua, F. (2012). A Hierarchical Approach for the Design of Gesture-to-Sound Mappings. *9th Sound and Music Computing Conference*, pages 233–240. 154

Françoise, J., Schnell, N., and Bevilacqua, F. (2013). A multimodal probabilistic model for gesture-based control of sound synthesis. *Proceedings of the 21st ACM international conference on Multimedia - MM '13*, pages 705–708. 154

Françoise, J., Schnell, N., Borghesi, R., and Bevilacqua, F. (2014). Probabilistic Models for Designing Motion and Sound Relationships. In Caramiaux, B., Tahiroglu, K., Fiebrink, R., and Tanaka, A., editors, *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 287–292, London, United Kingdom. Goldsmiths, University of London. 103

Freed, A., MacCallum, J., Schmeder, A., and Wessel, D. (2010). Visualizations and Interaction Strategies for Hybridization Interfaces. In Beilharz, K., Bongers, B., Johnston, A., and Ferguson, S., editors, *Proceedings of the 2010 Conference on New Interfaces for Musical Expression (NIME 2010)*, number Nime, pages 343–347, Sydney, Australia. 153

Freedman, N. and Grand, S., editors (1977). *Communicative Structures and Psychic Structures*. Springer US, Boston, MA. 93

Fried, O. and Fiebrink, R. (2013). Cross-modal Sound Mapping Using Deep Learning. *Proceedings of the 2013 New Interfaces for Musical Expression Conference (NIME2013)*. 157

Gallese, V. (2009). Motor abstraction: A neuroscientific account of how action goals and intentions are mapped and understood. *Psychological Research*, 73(4):486–498. 20

Gibbs, R. W. J. (2005). *Embodiment and Cognitive Science*. Cambridge University Press, Cambridge. 14

Gibson, E. J. (2000). Where Is the Information for Affordances? *Ecological Psychology*, 12(1):53–56. 33, 51

Gibson, J. J. (1977). The Theory of Affordances. In *Perceiving, Acting, and Knowing*, volume Perceiving, pages 127–142 (332). 32, 51

Gibson, J. J. (1986). *The Ecological Approach to Visual Perception*, volume 39. Houghton Mifflin. 33

Gibson, W. (2006). Material culture and embodied action: Sociological notes on the examination of musical instruments in jazz improvisation. *Sociological Review*, 54(1):171–187. 34

Gillian, N., Knapp, B., and O'Modhrain, S. (2011). Recognition Of Multivariate Temporal Musical Gestures Using N-Dimensional Dynamic Time Warping. In Jensenius, A. R., Tveit, A., Godøy, R. I., and Overholt, D., editors, *Proceedings of the 2011 International Conference on New Interfaces for Musical Expression (NIME2011)*, number June, pages 337–342, Oslo, Norway. University of Oslo and Norwegian Academy of Music. 151

Gillian, N. and Paradiso, J. A. (2014). The Gesture Recognition Toolkit. *Journal of Machine Learning Research*, 15:3483–3487. 103

Glowinski, D., Baron, N., Grandjean, D., Ott, T., Shirole, K., Torres-Eliard, K., and Rappaz, M.-A. (2014a). Analyzing expressive styles and functions of bodily movement in violinist performance. In *Proceedings of the 2014 International Workshop on Movement and Computing*, MOCO '14, pages 154:154–154:155, New York, NY, USA. ACM. 47, 48

Glowinski, D., Dael, N., Camurri, a., Volpe, G., Mortillaro, M., and Scherer, K. (2011). Toward a Minimal Representation of Affective Gestures. *IEEE Transactions on Affective Computing*, 2(2):106–118. 97, 99, 111

Glowinski, D., Dardard, F., Gnecco, G., Piana, S., and Camurri, A. (2014b). Expressive non-verbal interaction in a string quartet: an analysis through head movements. *Journal on Multimodal User Interfaces*. 15, 68

Glowinski, D., Gnecco, G., Piana, S., and Camurri, A. (2013a). Expressive Non-verbal Interaction in String Quartet. *2013 Humaine Association Conference on Affective Computing and Intelligent Interaction*, 0:233–238. 41

Glowinski, D., Mancini, M., Cowie, R., Camurri, A., Chiorri, C., and Doherty, C. (2013b). The movements made by performers in a skilled quartet: a distinctive pattern, and the function that it serves. *Frontiers in Psychology*, 4. 90

Godøy, R. I. (2003). Motor-Mimetic Music Cognition. *Leonardo*, 36(4):317–319. 30

Godøy, R. I. (2006). Gestural-sonorous objects: embodied extensions of schaeffer's conceptual apparatus. *Organised Sound*, 11(2):149–157. 15

Godøy, R. I. (2010). Gestural Affordances of Musical Sound. In Godøy, R. I. and Leman, M., editors, *Musical gestures: Sound, movement, and meaning*. Routledge. 2, 5, 14, 32, 136, 160

Godøy, R. I. (2014). Understanding Coarticulation in Musical Experience. pages 535–547. Springer International Publishing. 157

Godøy, R. I., Haga, E., and Jensenius, A. R. (2005). Playing "Air Instruments": Mimicry of Sound-producing Gestures by Novices and Experts. *Gesture in Human-Computer Interaction and Simulation, 6th International Gesture Workshop, GW 2005*, 3881/2006:256–267. 51, 58, 64

Godøy, R. I., Haga, E., and Jensenius, A. R. (2006a). Playing 'Air Instruments': Mimicry of Sound-producing Gestures by Novices and Experts. In Gibet, S., Courty, N., and Kamp, J.-F., editors, *Gesture in Human-Computer Interaction and Simulation, 6th International Gesture Workshop*, volume 3881 of *LNAI*, pages 256–267. Springer, Berlin Heidelberg. 28, 136, 137

Godøy, R. I., Haga, E., and Jensenius, A. R. (2006b). Playing Air Instruments : Mimicry of Sound-Producing Gestures by Novices and Experts. pages 256–267. 150

Godøy, R. I., Jensenius, A. R., Nymoen, K., Gody, R. I., Jensenius, A. R., and Nymoen, K. (2010). Chunking in Music by Coarticulation. *Acta Acoustica united with Acoustica*, 96(4):690–700. 15, 36, 82

Godøy, R. I. and Leman, M., editors (2010). *Musical Gestures: Sound, Movement, and Meaning*. Routledge. 2, 13, 15, 90, 108

Goebl, W. and Palmer, C. (2009). Synchronization of Timing and Motion Among Performing Musicians. *Music Perception: An Interdisciplinary Journal*, 26(5):427–438. 90

Goffman, E. (1974). *Frame analysis: An essay on the organization of experience.* Harvard University Press. 108

Goudeseune, C., Garnett, G., and Johnson, T. (2001). Resonant Processing of Instrumental Sound Controlled by Spatial Position. In Poupyrev, I., Lyons, M. J., Fels, S. S., and Blaine, T., editors, *Proceedings of the 2001 International Conference on New Interfaces for Musical Expression (NIME-01)*, pages 24–26, Seattle, WA, United States. 153

Green, O. (2016). The Situation of Practice-Led Research Around NIME, and Two Methodological Suggestions for Improved Communication. *Leonardo*, 49(1):78–79. 6, 128

Gritten, A. and King, E., editors (2006). *Music and Gesture*. Ashgate. 2, 13

Gritten, A. and King, E. (2011). *New Perspectives on Music and Gesture*. SEMPRE studies in the psychology of music. Ashgate. 2, 13

Gurevich, M., Marquez-Borbon, A., and Stapleton, P. (2012). Playing with Constraints: Stylistic Variation with a Simple Electronic Instrument. *Computer Music Journal*, 36(1):23–41. 131

Hachimura, K., Takashina, K., and Yoshimura, M. (2005). Analysis and evaluation of dancing movement based on LMA. In *ROMAN 2005. IEEE International Workshop on Robot and Human Interactive Communication, 2005.*, pages 294–299. IEEE. 99

Hackney, P. (2002). *Making Connections: Total Body Integration Through Bartenieff Fundamentals*. Routledge. 98

Hansen, K. F. (2002). The Basics of Scratching. *Journal of New Music Research*, 31(4):357–365. 32

Hardjowirogo, S.-I. (2017). Instrumentality. On the Construction of Instrumental Identity. In Bovermann, T., de Campo, A., Egermann, H., Hardjowirogo, S.-I., and Weinzierl, S., editors, *Musical Instruments in the 21st Century*, pages 9–24. Springer Singapore, Singapore. 29, 30, 32

Henbing, L. L. and Leman, M. (2007). A Gesture-Based Typology of Sliding-Tones in Guqin Music. *Journal of New Music Research*, 36(2):61–82. 19

Hinton, A. (2014). *Understanding Context*. Number 37. O'Reilly Media, Inc. xix, 17

Hirose, N. (2002). An ecological approach to embodiment and cognition. *Cognitive Systems Research*, 3(3):289–299. 30

Hoadley, R., Nash, C., and Fober, D., editors (2016). *Performance Practice of Real-Time Notation*. Anglia Ruskin University, Cambridge, UK. 128

Hommel, B. (2009). Action control according to TEC (theory of event coding). *Psychological Research Psychologische Forschung*, 73(4):512–526. 18

Hommel, B., Müsseler, J., Aschersleben, G., and Prinz, W. (2001). The Theory of Event Coding (TEC): A framework for perception and action planning. *Behavioral and Brain Sciences*, 24(05):849–878. 18

Hornbostel, E. M. V. (1933). The ethnology of African sound-instruments (continued). *Africa: Journal of the International African Institute*, 6(3):277–311. 29

Hunt, A. and Kirk, R. (2000). Mapping Strategies for Musical Performance. *Trends in Gestural Control of Music*, pages 231–258. 119

Hunt, A., Wanderley, M. M., and Kirk, R. (2000). Towards a Model for Instrumental Mapping in Expert Musical Interaction University of York Analysis-Synthesis Team. *Proc. of the Int. Computer Music Conf. (ICMC'2000), ICMA: Berlin, Germany.*, pages pp. 209–12. 119, 120

Hunt, A. A. D., Wanderley, M. M., Paradis, M., Janhan Sousa, W., Wanderley, M. M., and Paradis, M. (2003). The importance of Parameter Mapping in Electronic Instrument Design. *Journal of New Music Research*, 32(4):88–93. 80, 107

Hurley, S. L. S. L. (1998). *Consciousness in action*. Harvard University Press. xix, 16

Huron, D. and Berec, J. (2009). Characterizing Idiomatic Organization in Music: A Theory and Case Study of Musical Affordances. *Empirical Musicology Review*, 4(3):103–122. 31

Hutchins, E. (2010). Cognitive Ecology. *Topics in Cognitive Science*, 2(4):705–715. 14

Iordanescu, L., Grabowecky, M., and Suzuki, S. (2013). Action enhances auditory but not visual temporal sensitivity. *Psychonomic Bulletin & Review*, 20(1):108–114. 18

Jenkins, H. S. (2008). Gibson's Affordances: Evolution of a Pivotal Concept. *Journal of Scientific Psychology*, (December):34–45. 33

Jensenius, A. R. (2007). *Action-sound: Developing methods and tools to study music-related body movement.* Ph.d. thesis, Universitetet i Oslo. 136

Jensenius, A. R. (2014). To gesture or Not? An Analysis of Terminology in NIME Proceedings 2001-2013. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 217–220. 24

Jensenius, A. R. and Lyons, M., editors (2016). *A NIME Reader: Fifteen years of New Interfaces for Musical Expression.* Springer, Berlin, Forthcoming. 31, 80

Jensenius, A. R., Skogstad, S. A., Nymoen, K., Tørresen, J., and Høvin, M. E. (2009). Reduced displays of multidimensional motion capture data sets of musical performance. In *Proceedings of ESCOM 2009: 7th Triennial Conference of the European Society for the Cognitive Sciences of Music*. 57

Jensenius, A. R., Wanderley, M. M., Godøy, R. I., and Leman, M. (2010). Musical gestures: Concepts and methods in research. In Godøy, R. I. and Leman, M., editors, *Musical Gestures: Sound, Movement, and Meaning*, pages 12–35. Routledge. 24, 25, 27, 28, 35, 36, 81, 82

Jordà, S. (2004). Instruments and Players: Some Thoughts on Digital Lutherie. *Journal of New Music Research*, 33(3):321–341. 129

Katsman, R. (2006). Anthropoetic gesture: A key to Milorad Pavic's poetics (Landscape Painted with Tea). *Academic Electronic Journal in Slavic Studies*, 16. 25

Kendon, A. (2004). Gesture: visible action as utterance. page 400. 25

Kilner, J. M. and Lemon, R. N. (2013). What We Know Currently about Mirror Neurons. *Current Biology*, 23(23):R1057–62. 87

King, E. (2006). Supporting Gestures: Breathing in Piano Performance. In Gritten, A. and King, E., editors, *Music and Gesture*, chapter 9, pages 142–164. Ashgate, Aldershot. 91

Kirn, P. (2013). How music can predict the human/machine future. Berlin, Germany. re:publica 2013. 112

Kojs, J. (2011). Notating Action-Based Music. 128

Kurtenbach, G. and Hulteen, E. A. (1990). *Gestures in human-computer communication.* Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA. 25

Küssner, M. B., Tidhar, D., Prior, H. M., and Leech-Wilkinson, D. (2014). Musicians are more consistent: Gestural cross-modal mappings of pitch, loudness and tempo in real-time. *Frontiers in Psychology*, 5(July):1–15. 15

Kvifte, T. (2008). What is a musical instrument? *Svensk tidskrift för musikforskning*, 1:45–56. 28, 29

Laban, R. (1966). *Choreutics*. Macdonald & Evans, London. 104, 123

Laban, R. and Lawrence, F. C. (1947). *Effort*. Macdonald & Evans, London, reprinted edition. 98, 104

Lähdeoja, O., Wanderley, M. M., and Malloch, J. (2009). Instrument augmentation using ancillary gestures for subtle sonic effects. In *Proceedings of the SMC 2009 - 6th Sound and Music Computing Conference*, number July, pages 327–330, Porto, Portugal. 81, 87

Larboulette, C. and Gibet, S. (2015). A review of computable expressive descriptors of human motion. In *Proceedings of the 2nd International Workshop on Movement and Computing - MOCO '15*, pages 21–28, New York, New York, USA. ACM Press. 97

Laskewicz, Z. (2008). *The New Music-Theatre of Mauricio Kagel*. Night Shades Press, Ghent, Belgium. 128

Lavender, T. and Hommel, B. (2007). Affect and action: Towards an event-coding account. *Cognition & Emotion*, 21(6):1270–1296. 18

Leman, M. (2008a). *Embodied Music Cognition and Mediation Technology*. MIT Press. 2, 13, 14, 15, 51, 85, 90, 107

Leman, M. (2008b). Systematic Musicology at the Crossroads of Modern Music Research. *Systematic and Comparative Musicology: Concepts, Methods, Findings*, pages 89–115. 1

Leman, M. (2010). Music, gesture, and the formation of embodied meaning. In Godøy, R. I. and Leman, M., editors, *Musical Gestures: Sound, Movement and Meaning*, pages 126–153. Routledge. 15, 24, 87, 110

Leman, M. (2012). Musical gestures and embodied cognition. In Dutoit, T., Todoroff, T., and D'Alessandro, N., editors, *Actes des Journées d'Informatique Musicale (JIM 2012)*, number Jim, pages 5–7, Mons, Belgique. UMONS/numediart. 14, 15

Leman, M., Desmet, F., Styns, F., Van Noorden, L., and Moelants, D. (2009a). Sharing musical expression through embodied listening: A case study based on Chinese Guqin music. *Music Perception*, 26(3):263–278. 15

Leman, M., Desmet, F., Styns, F., Van Noorden, L., and Moelants, D. (2009b). Sharing Musical Expression Through Embodied Listening: A Case Study Based on Chinese Guqin Music. *Music Perception*, 26(3):263–278. 19

Leman, M., Lesaffre, M., Nijs, L., and Deweppe, A. (2010). User-Oriented Studies in Embodied Music Cognition Research. *Musicae Scientiae*, 14(2_suppl):203–223. 51

Leman, M. and Maes, P.-J. (2014). The role of embodiment in the perception of music. *Empirical Musicology Review*, 9(3-4):236–246. 14, 16, 19

Lesaffre, M., Leman, M., Tanghe, K., De Baets, B., De Meyer, H., and Martens, J.-P. (2003). User-dependent taxonomy of musical features as a conceptual framework for musical audio-mining technology. In *Proceedings of the Stockholm Music Acoustics Conference*, pages 635–638, Stockholm. 83

Libin, L. (2017). Acoustic. *Grove Music Online*. 32

Lopes, P., Ion, A., Mueller, W., Hoffmann, D., Jonell, P., and Baudisch, P. (2015a). Proprioceptive Interaction. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems - CHI '15*, pages 939–948, New York, New York, USA. ACM Press. 163

Lopes, P., Jonell, P., and Baudisch, P. (2015b). Affordance++. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems - CHI '15*, pages 2515–2524, New York, New York, USA. ACM Press. 163

Lucchese, G., Field, M., Ho, J., Gutierrez-Osuna, R., and Hammond, T. (2012). GestureCommander: continuous touch-based gesture prediction. *Proceedings of the 2012 ACM annual conference extended abstracts on Human Factors in Computing Systems Extended Abstracts CHI EA '12*, pages 1925–1930. 154

Luck, G. and Sloboda, J. A. (2009). Spatio-Temporal Cues for Visually Mediated Synchronization. *Music Perception: An Interdisciplinary Journal*, 26(5):465–473. 90

Lum, P. Y., Singh, G., Lehman, A., Ishkanov, T., Vejdemo-Johansson, M., Alagappan, M., Carlsson, J., and Carlsson, G. (2013). Extracting insights from the shape of complex data using topology. *Scientific reports*, 3:1236. 157

MacRitchie, J., Buck, B., and Bailey, N. J. (2013). Inferring musical structure through bodily gestures. *Musicae Scientiae*, 17(1):86–108. 75

Madgwick, S. O. H., Harrison, A. J. L., and Vaidyanathan, A. (2011). Estimation of IMU and MARG orientation using a gradient descent algorithm. *IEEE ... International Conference on Rehabilitation Robotics : [proceedings]*, 2011:5975346. 95

Maes, P.-j., Amelynck, D., Lesaffre, M., Leman, M., and Arvind, D. (2012). The Conducting Master: An interactive, real-time gesture monitoring system based on spatiotemporal motion templates. *International Journal of Human-Computer Interaction*, (March 2015):120919075414004. 90

Maes, P.-J., Leman, M., Kochman, K., Lesaffre, M., and Demey, M. (2011). The "One-Person Choir": A Multidisciplinary Approach to the Development of an Embodied Human-Computer Interface. *Computer Music Journal*, 35(2):22–35. 31, 80

Maes, P.-J., Palmer, C., Leman, M., Wanderley, M. M., Palmer, C., Wanderley, M. M., Leman, M., and Wanderley, M. M. (2014). Action-Based Effects on Music Perception. *Frontiers in Psychology*, 4:1008. 15, 17, 30, 87

Maestri, E. and Antoniadis, P. (2015). Notation as Instrument: From Representation to Enaction. In Battier, M., Bresson, J., Couprie, P., Davy-Rigaux, C., Fober, D., Geslin, Y., Genevois, H., Picard, F., and Tacaille, A., editors, *Proceedings of the First International Conference on Technologies for Music Notation and Representation - TENOR2015*, pages 210–217, Paris, France. Institut de Recherche en Musicologie. 128

Magnusson, T. (2006). Affordances and constraints in screen-based musical instruments. *Proceedings of the 4th NordiCHI*, (October):441–444. 119

Magnusson, T. (2009). Of Epistemic Tools: musical instruments as cognitive extensions. *Organised Sound*, 14(02):168. 30

Magnusson, T. (2010). Designing constraints: composing and performing with digital musical systems. *Computer Music Journal*, 34(4):62–74. 34, 131

Malafouris, L. (2013). *How things shape the mind: A theory of material engagement*. MIT Press, Cambridge, USA. 110

Manning, F. and Schutz, M. (2013). Moving to the beat improves timing perception. *Psychonomic Bulletin & Review*, 20(6):1133–1139. 18

Marczak, K. (2009). *Theatrical elements and their relationship with music in Karlheinz Stockhausen's HARLEKIN for clarinet*. Doctoral thesis, University of British Columbia. 128

Matyja, J. R. and Schiavio, A. (2013). Enactive music cognition: Background and research themes. *Constructivist Foundations*. 19, 87

Mayor, O., Llimona, Q., Marchini, M., Papiotis, P., and Maestre, E. (2013). repoVizz: A Framework for Remote Storage, Browsing, Annotation, and Exchange of Multi-modal D ata. *Proceedings of the 21st ACM international conference on Multimedia*, pages 415–416. 93

Mays, T. and Faber, F. (2014). A Notation System for the Karlax Controller. In Caramiaux, B., Tahiroglu, K., Fiebrink, R., and Tanaka, A., editors, *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 553–556, London, United Kingdom. Goldsmiths, University of London. 128

Mazzola, G. and Andreatta, M. (2007). Diagrams, Gestures and Formulae in Music. *Journal of Mathematics and Music*, 1(1):23–46. 25

Mcgrenere, J. and Ho, W. (2000). Affordances: Clarifying and Evolving a Concept. *Proceedings of GraphIcs Interface*, (May):1–8. 33

McNeill, D. (1996). *Hand and Mind: What Gestures Reveal about Thought*. Psychology/cognitive science. University of Chicago Press. 93

Mcpherson, A. P. and Kim, Y. E. (2012). The Problem of the Second Performer: Building a Community Around an Augmented Piano. *Computer Music Journal*, 36(4):10–27. 34, 128, 129

Menin, D. and Schiavio, A. (2012). Rethinking Musical Affordances. *Avant*, III(2):201–215. 51

Merleau-Ponty, M. (2002). *Phenomenology of Perception*. Psychology Press. 92, 110

Miranda, E. R. and Wanderley, M. (2006). *New Digital Musical Instruments: Control And Interaction Beyond the Keyboard (Computer Music and Digital Audio Series)*. A-R Editions, Inc., Madison, WI, USA. 31, 34, 80

Momeni, A. and Wessel, D. (2003). Characterizing and Controlling Musical Material Intuitively with Geometric Models. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 54–62. 153

Mulder, A. (1996). Hand gestures for HCI. *Hand Centered Studies of Human Movement Project, Technical Report*, pages 91–96. 133

Mulder, A. (2000). Towards a choice of gestural constraints for instrumental performers. *Trends in gestural control of music*, pages 315–335. 131

Müller, M. (2007). *Information Retrieval for Music and Motion*. Springer Verlag. 60, 84, 85

Murray-Browne, T., Mainstone, D., Bryan-kinns, N., and Plumbley, M. D. (2011). The Medium is the Message: Composing Instruments and Performing Mappings. In Jensenius, A. R., Tveit, A., Godoy, R. I., and Overholt, D., editors, *Proceedings of the International Conference on New Interfaces for Musical Expression*, number June, pages 56–59, Oslo, Norway. 80

NaturalPoint Inc. (2016). Rigid Body Tracking. In *NaturalPoint Product Documentation*. 143, 145

Naveda, L. and Leman, M. (2010). The Spatiotemporal Representation of Dance and Music Gestures using Topological Gesture Analysis (TGA). *Music Perception*, 28(1):93–111. 15, 75, 90, 102, 137

Newton, D. and Marshall, M. T. (2011). Examining How Musicians Create Augmented Musical Instruments. In Jensenius, A. R., Tveit, A., Godoy, R. I., and Overholt, D., editors, *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 155–160, Oslo, Norway. 34

Nijs, L., Leman, M., Lesaffre, M., and Leman, M. (2009). The musical instrument as a natural extension of the musician. In Castellango, M. and Genevois, H., editors, *Proceedings of the 5th Conference of Interdisciplinary Musicology*, pages 132–133. LAM-Institut jean Le Rond d'Alembert. 30, 32

Nijs, L., Lesaffre, M., and Leman, M. (2013). The musical instrument as a natural extension of the musician. In Castellegno, M., Genevois, H., and Bardez, J.-M., editors, *Music and its instruments*, pages 467–484. Edition Delatour France. 91

Norman, D. A. (1988). *The Psychology Of Everyday Things*. Basic Books. 33

Norman, D. A. (1999). Affordance, conventions, and design. *Interactions*, 6(3):38–43. 33

Norman, S.-J. (2015). Oskar Schlemmer's programmatic gesture research. In Sutil, N. S. and Popat, S., editors, *Digital movement: essays in motion technology and performance*, Palgrave studies in performance and technology, pages 21–34. Palgrave Macmillan, London. 112

Norman, S. J. and Blackwell, A. F. (2010). Gesture and Embodied Interaction: Capturing Motion / Data / Value. *Leonardo Journal*, 43(2):198–199. 112

Norman, S. J., Ryan, J., and Waisvisz, M. (1998). Touchstone. In *Touch*. 119

Nymoen, K., Glette, K., Skogstad, S. A., Torresen, J., and Jensenius, A. R. (2010). Searching for Cross-Individual Relationships between Sound and Movement Features using an SVM Classifier. In Beilharz, K., Bongers, B., Johnston, A., and Ferguson, S., editors, *Proceedings of the 2010 Conference on New Interfaces for Musical Expression (NIME 2010)*, number Nime, pages 259–262, Sydney, Australia. 154

Nymoen, K., Godøy, R. I., Jensenius, A. R., and Torresen, J. (2013). Analyzing correspondence between sound objects and body motion. *ACM Transactions on Applied Perception*, 10(2):1–22. 15

Nymoen, K., Haugen, M. R., and Jensenius, A. R. (2015). MuMYO Evaluating and Exploring the MYO Armband for Musical Interaction. In *Proceedings of the International Conference on New Interfaces For Musical Expression*, pages 215–218, Baton Rouge, LA. 121

Ortmann, O. (1929). *The physiological mechanics of piano technique: an experimental study of the nature of muscular action as used in piano playing, and of the effects thereof upon the piano key and the piano tone*. Da Capo Press, Incorporated. 91

Palmer, C., Koopmans, E., Carter, C., Loehr, J. D., and Wanderley, M. M. (2009). Synchronization of motion and timing in clarinet performance. *International Symposium on Performance Science*, (Davidson 1995):1–6. 136

Parviainen, J., Tuuri, K., and Pirhonen, A. (2013a). Drifting Down the Technologization of Life: Could Choreography-Based Interaction Design Support us in Engaging with the World and our Embodied Living? *Challenges*, 4(1):103–115. 111

Parviainen, J., Tuuri, K., Pirhonen, A., Turunen, M., and Keskinen, T. (2013b). Gestures within Human-Technology Choreographies for Interaction Design. 104, 111

Penttinen, H., Pakarinen, J., Välimäki, V., Laurson, M., Li, H., and Leman, M. (2006). Model-based sound synthesis of the guqin. *The Journal of the Acoustical Society of America*, 120(6):4052–4063. 19

Piana, S., Staglianò, A., Camurri, A., and Odone, F. (2013). A set of full-body movement features for emotion recognition to help children affected by autism spectrum condition. In *IDGEI International Workshop*, Crete, Greece. 111, 150

Piana, S., Staglianò, A., Odone, F., and Camurri, A. (2016). Adaptive Body Gesture Representation for Automatic Emotion Recognition. *ACM Transactions on Interactive Intelligent Systems*, 6(1):1–31. 97

Piana, S., Staglianò, A., Odone, F., Verri, A., and Camurri, A. (2014). Real-time Automatic Emotion Recognition from Body Gestures. 154

Pintaric, T. and Kaufmann, H. (2008). A rigid-body target design methodology for optical pose-tracking systems. In *Proceedings of the 2008 ACM symposium on Virtual reality software and technology - VRST '08*, page 73, New York, New York, USA. ACM Press. 146

Pirhonen, A. (2013). Choreographies at Home as Objects of Design. 111

Pollick, F. E., Paterson, H. M., Bruderlin, A., and Sanford, A. J. (2001). Perceiving affect from arm movement. *Cognition*, 82(2):B51–B61. 97

Prinz, W. (1990). A Common Coding Approach to Perception and Action. In *Relationships Between Perception and Action*, pages 167–201. Springer Berlin Heidelberg, Berlin, Heidelberg. 17

Prinz, W. (1997). Perception and Action Planning. *European Journal of Cognitive Psychology*, 9(2):129–154. 17

Rajko, S., Qian, G., Ingalls, T., and James, J. (2007). Real-time Gesture Recognition with Minimal Training Requirements and On-line Learning. In *2007 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8. IEEE. 154

Ramsay, J. O. (2006). *Functional data analysis*. Wiley Online Library. 51, 69

Rasamimanana, N. H., Bernardin, D., Wanderley, M., and Bevilacqua, F. (2008). String Bowing Gestures at Varying Bow Stroke Frequencies: A Case Study. In *Lecture Notes in Computer Science*, volume 5085, pages 216–226. Springer Verlag. 41

Repp, B. H. and Su, Y.-H. (2013). Sensorimotor synchronization: A review of recent research (2006–2012). *Psychonomic Bulletin & Review*, 20(3):403–452. 84

Rowlands, M. (2010). *The new science of the mind: From extended mind to embodied phenomenology*. MIT Press. 14

Salazar Sutil, N. (2015). *Motion and Representation: The Language of Human Movement*. MIT Press. 92, 104

Salvador, S. and Chan, P. (2007). FastDTW : Toward Accurate Dynamic Time Warping in Linear Time and Space. *Intelligent Data Analysis*, 11:561–580. 151

Schacher, J. C., Järveläinen, H., Strinning, C., and Neff, P. (2015). Movement Perception in Music Performance - A Mixed Methods Investigation. *Proceedings of the International Conference on Sound and Music Computing, SMC'15, Maynooth, Ireland*. 5, 28

Schaeffer, P. (1966). *Traité des objets musicaux*. Éditions du Seuil, Paris, France. 15

Schaeffer, P., Reibel, G., and Ferreyra, B. (1998). *Solfège de l'objet sonore*. INA GRM, Paris, France. 15

Schiavio, A. (2014). Commentary on Leman & Maes , The role of embodiment in the perception of music (2014). 9(3):1–9. 19, 20

Schmitt, J.-C. (1990). *La raison des gestes dans l'Occident médiéval*. 25

Schnell, N., Röbel, A., Schwarz, D., Peeters, G., and Borghesi, R. (2009). MuBu & Friends - Assembling Tools for Content Based Real-Time Interactive Audio Processing in Max/MSP. In *Proceedings of ICMC*. 140

Schoonderwaldt, E. and Chen, L. J.-Y. (2009). The player and the instrument: Analysis of bowing gestures in violin and viola performance. In *Proceedings of he Fifth Conference on Interdisciplinary Musicology (CIM09)*, pages 156–157. 41

Schutz, M. and Lipscomb, S. (2007). Hearing gestures, seeing music: vision influences perceived tone duration. *Perception*, 36(6):888–897. 30

Senin, P. (2008). Dynamic time warping algorithm review. *Information and Computer Science Department University of Hawaii at Manoa Honolulu, USA*, pages 1–23. 151

Sethares, W. A. and Budney, R. (2014). Topology of musical data. *Journal of Mathematics and Music*, 8(1):73–92. 137

Shang, J., Mao, G., and Gu, Q. (2002). Design and implementation of MIMU/GPS integrated navigation systems. In *Position Location and Navigation Symposium, 2002 IEEE*, pages 99–105. IEEE. 94

Smith, H. and Dean, R. T. (2009). *Practice-Led Research, Research-Led Practice in the Creative Arts*. Research Methods for the Arts and Humanities. Edinburgh University Press, Edinburgh. 5, 6, 87, 91

Smith, J. (1996). Physical Modeling Synthesis Update. *Computer Music Journal*, 20(2):44–56. 120, 138

Smith, J. O. (1992). Physical modeling using digital waveguides. *Computer Music journal*, 16(4):74–91. 120, 138

Snapper, J. (2004). Scratching the surface: Spinning time and identity in hip-hop turntablism. 32

Sonami, L. (2006). On my work. *Contemporary Music Review*, 25(5-6):613–614. 118

Sonnenfeld, A. and Hansen, K. F. (2016). S-notation: A complete musical notation system for scratching and sample music derived from "Theory of Motions". In Hoadley, R., Nash, C., and Fober, D., editors, *Proceedings of the International Conference on Technologies for Music Notation and Representation - TENOR2016*, number July, pages 50–57, Cambridge, UK. Anglia Ruskin University. 32

Stewart, J. R., Gapenne, O., and Di Paolo, E. A. (2010). *Enaction: toward a new paradigm for cognitive science*. MIT Press. 19

Sullivan, G. (2009). Making Space: The Purpose and Place of Practice-Led Research. In Smith, H. and Dean, R. T., editors, *Practice-led research, research-led practice in the creative arts*, chapter 2, pages 41–65. Edinburgh University Press, Edinburgh. 107

Teixeira, E. C., Loureiro, M. a., Wanderley, M. M., and Yehia, H. C. (2015). Motion Analysis of Clarinet Performers. *Journal of New Music Research*, 44(2):97–111. 136

Thompson, E. (2007). *Mind in life: biology, phenomenology, and the sciences of mind*. Belknap Press of Harvard University Press. 19

Thompson, M. (2012). *The Application of Motion Capture to Embodied Music Cognition Research Marc Thompson*. PhD thesis, University of Jyväskylä. 90

Thompson, M. R. and Luck, G. (2011). Exploring relationships between pianists' body movements, their expressive intentions, and structural elements of the music. *Musicae Scientiae*. 41

Torre, G., Andersen, K., and Baldé, F. (2016). The Hands: The Making of a Digital Musical Instrument. *Computer Music Journal*, 40(2):22–34. 119

Trojaniello, D., Cereatti, A., Pelosin, E., Avanzino, L., Mirelman, A., Hausdorff, J. M., and Della Croce, U. (2014). Estimation of step-by-step spatio-temporal parameters of normal and impaired gait using shank-mounted magneto-inertial sensors: application to elderly, hemiparetic, parkinsonian and choreic gait. *Journal of NeuroEngineering and Rehabilitation*, 11(1):152. 94

Van Dyck, E., Moelants, D., Demey, M., Deweppe, A., Coussement, P., and Leman, M. (2013). The impact of the bass drum on human dance movement. *Music Perception*, 30(4):349–359. 90

Van Nort, D., Wanderley, M. M., and Depalle, P. (2014). Mapping Control Structures for Sound Synthesis: Functional and Topological Perspectives. *Computer Music Journal*, 38(3):6–22. 119, 120, 152, 153

Verfaille, V., Wanderley, M. M., and Depalle, P. (2006). Mapping Strategies for Gestural and Adaptive Control of Digital Audio Effects. *Journal of New Music Research*, 35(1):71–93. 81

Vicon® (2006). Plug-in-Gait Marker Placement. 147

Vines, B. W., Krumhansl, C. L., Wanderley, M. M., and Levitin, D. J. (2006). Cross-modal interactions in the perception of musical performance. *Cognition*, 101(1):80–113. 18

Visi, F. (2016). Tuned Constraint. In *ICLI 2016 International Conference on Live Interfaces*, University of Sussex. 119, 131, 132

Visi, F., Coorevits, E., Miranda, E., and Leman, M. (2014a). Effects of different bow stroke styles on body movements of a viola player: an exploratory study. In *Proceedings of The joint ICMC|SMC|2014 Conference*, Athens, Greece. 39, 68, 92

Visi, F., Coorevits, E., and Miranda, E. R. (2015a). A practice-based study on instrumental gestures in music composition and performance: Kineslimina. In *MuSA 2015 - Sixth International Symposium on Music/Sonic Art: Practices and Theories*, Karlsruhe, Germany. 117

Visi, F., Coorevits, E., Schramm, R., and Miranda, E. (2015b). Instrumental Movements of Neophytes: Analysis of Movement Periodicities, Commonalities and Individualities in Mimed Violin Performance. In *Proceedings of the 11th International Symposium on Computer Music Multidisciplinary Research (CMMR)*, Plymouth, United Kingdom. 92

Visi, F., Coorevits, E., Schramm, R., and Miranda, E. (2016). Analysis of Mimed Violin Performance Movements of Neophytes. In Kronland-Martinet, R., Aramaki, M., and Ystad, S., editors, *Music, Mind, and Embodiment: 11th International Symposium, CMMR 2015, Plymouth, UK, June 16-19, 2015, Revised Selected Papers*, volume LNCS 9617, pages 88–108. Springer International Publishing. 15, 49

Visi, F., Coorevits, E., Schramm, R., and Miranda, E. R. (2017). Musical Instruments, Body Movement, Space, and Motion Data: Music as an Emergent Multimodal Choreography. *Human Technology: An Interdisciplinary Journal on Humans in ICT Environments (forthcoming)*. 89

Visi, F. and Miranda, E. R. (2016). Instrumental Movements to Physical Models: Mapping Postural and Sonic Topologies through Machine Learning. In *Porto International Conference on Musical Gesture as Creative Interface*. 119, 133

Visi, F., Schramm, R., and Miranda, E. (2014b). Gesture in performance with traditional musical instruments and electronics: Use of embodied music cognition and multimodal motion capture to design gestural mapping strategies. In *MOCO '14: Proceedings of the 2014 International Workshop on Movement and Computing*, Paris, France. ACM. 60, 79

Vuoskoski, J. K., Thompson, M. R., Clarke, E. F., and Spence, C. (2014). Crossmodal interactions in the perception of expressivity in musical performance. *Attention, Perception, & Psychophysics*, 76(2):591–604. 15, 18

Wachsmann, K., Kartomi, M. J., von Hornbostel, E. M., and Sachs, C. (2017). Instruments, classification of. *Grove Music Online*. 29

Waisvisz, M. (1985). The Hands: A Set of Remote MIDI-Controllers. In *International Computer Music Conference Proceedings*, pages 313–318. Michigan Publishing, University of Michigan Library. 119

Waisvisz, M. (2006). Manager or Musician? About virtuosity in live electronic music. Do we operate our electronic systems or do we play them? In *Proceedings of the International Conference on New Interfaces for Musical Expression*, page 415. 87

Wanderley, M. M. (1999). Non-obvious Performer Gestures in Instrumental Music. In Braffort, A., Gherbi, R., Gibet, S., Richardson, J., and Teil, D., editors, *Gesture Workshop*, volume 1739 of *Lecture Notes in Computer Science*, pages 37–48. Springer. 137

Wanderley, M. M. and Malloch, J. (2014). Editors' Notes. *Computer Music Journal*, 38(3):4–5. 80

Wanderley, M. M., Vines, B. W., Middleton, N., McKay, C., and Hatch, W. (2005). The Musical Significance of Clarinetists' Ancillary Gestures: An Exploration of the Field. *Journal of New Music Research*, 34(1):97–113. 26, 68

Williamon, A. and Davidson, J. W. (2002). Exploring co-performer communication. *Muslcae Sclentlle Spring*, VI(1):53–72. 91

Wilson, S. (2013). Building an Instrument, Building an Instrumentalist: Helmut Lachenmann's Serynade. *Contemporary Music Review*, 32(5):425–436. 105, 110

Woolford, K. (2014). Capturing Human Movement in the Wild. *Proceedings of the 2014 International Workshop on Movement and Computing - MOCO '14*, pages 19–24. 111

Appendix A

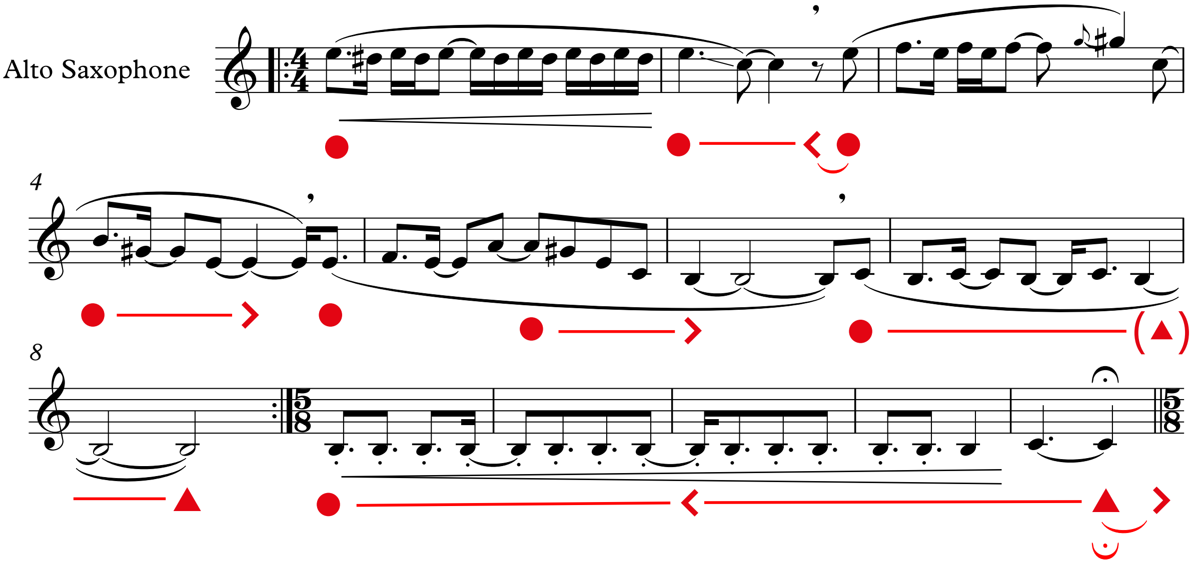
11 Degrees of Dependence: Alto Saxophone Score

11 Degrees of Dependence

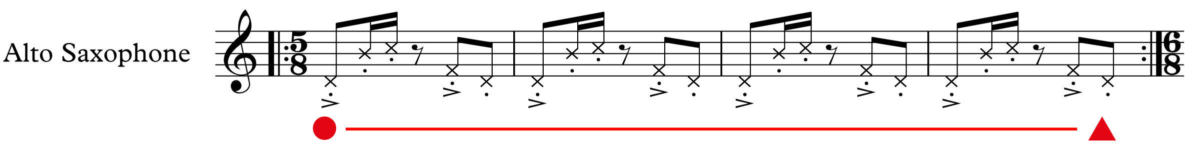
(alto saxophone part)

Federico Visi

Section 1



Section 2

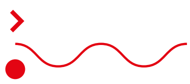


Arrangement:

Sec. 1
· Bar 1–8 two times, no movements the first time;
· Impro (16 bars);
· Bar 1-8 two times + Coda to part 2...
Sec. 2
· Guitar theme twice accompained by sax keyclicks (16 bars)
· Free Impro (adlib);
· Guitar theme twice accompained by sax keyclicks (16 bars)
Sec. 3
The time signature changes following the same pattern (four 6/8 bars; four 5/8 bars).
When you are ready, start with part 3a at the beginning of a pattern module (the first 6/8 bar), then:
· 3a two times
· Impro (32 bars, i.e. 4 full 6/8+5/8 bass patterns)
· 3b (end).

Movement Notation

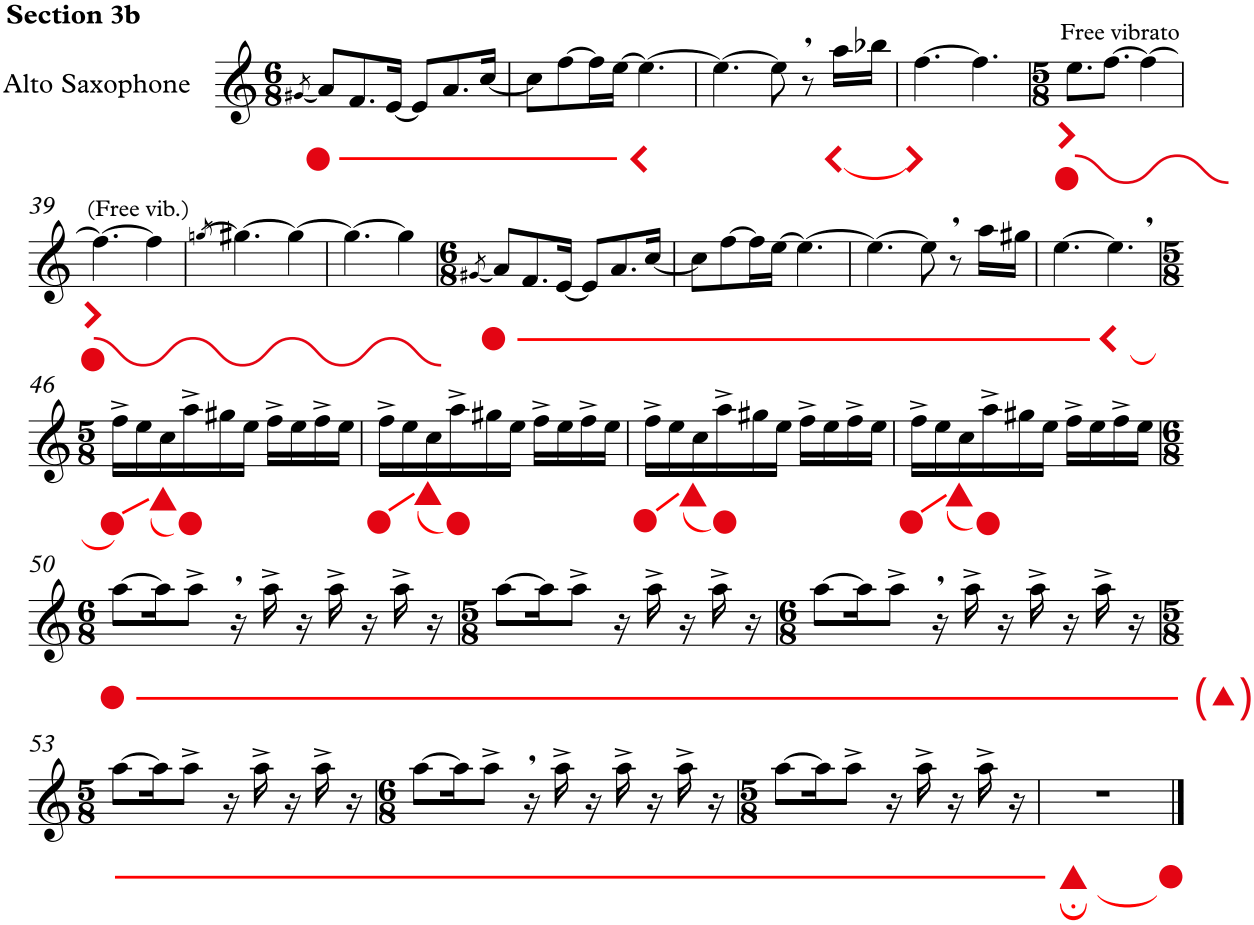
● = Rest
▲ = Open
< = Left
> = Right


= Modulate between the two positions

2

Section 3a

Section 3b



Appendix B

List of Related Publications, Performances, Talks, and Workshops

B.1 Book Chapters

- Visi, F., Coorevits, E., Schramm, R., & Miranda, E. Analysis of Mimed Violin Performance Movements of Neophytes. In *Extended Proceedings of the 11th International Symposium on Computer Music Multidisciplinary Research (CMMR)* (Vol. LNCS 9617). Springer, (in print).
- Visi, F. Augmenting Instruments and Extending Cultures: on the Overtone Violin. In A. R. Jensenius & M. Lyons (Eds.), *A NIME Reader: Fifteen years of New Interfaces for Musical Expression*. Springer, (in print).

B.2 Journal Articles

- Visi, F., Coorevits, E., Schramm, R., & Miranda, E. R. (n.d.). Musical Instruments, Body Movement, Space, and Motion Data: Music as an Emergent Multimodal Choreography. *Human Technology: An Interdisciplinary Journal on Humans in ICT Environments* (forthcoming).

B.3 Conference Proceedings

- Visi, F., & Miranda, E. R. (2016). Instrumental Movements to Physical Models: Mapping Postural and Sonic Topologies through Machine Learning. In *Porto International Conference on Musical Gesture as Creative Interface.* Porto, Portugal.

- Visi, F., Coorevits, E., & Miranda, E. R. (2015). A practice-based study on instrumental gestures in music composition and performance: Kineslimina. In *MuSA 2015 – Sixth International Symposium on Music/Sonic Art: Practices and Theories.* Karlsruhe, Germany.

- Visi, F., Coorevits, E., Schramm, R., & Miranda, E. (2015). Instrumental Movements of Neophytes: Analysis of Movement Periodicities, Commonalities and Individualities in Mimed Violin Performance. In *Proceedings of the 11th International Symposium on Computer Music Multidisciplinary Research (CMMR).* Plymouth, United Kingdom.

- Visi, F., Coorevits, E., Miranda, E., & Leman, M. (2014). Effects of different bow stroke styles on body movements of a viola player: an exploratory study. In *Proceedings of The joint ICMC/SMC/2014 Conference.* Athens, Greece.

- Visi, F., Schramm, R., & Miranda, E. (2014). Use of Body Motion to Enhance Traditional Musical Instruments: A Multimodal Embodied Approach to Gesture Mapping , Composition and Performance. In *Proceedings of the International Conference on New Interfaces for Musical Expression* (pp. 601–604). London, United Kingdom.

- Visi, F., Schramm, R., & Miranda, E. (2014). Gesture in performance with traditional musical instruments and electronics: Use of embodied music cognition and multimodal motion capture to design gestural mapping strategies. In *MOCO 14: Proceedings of the 2014 International Workshop on Movement and Computing.* Paris, France.

- Schramm, R., Nunes, H. de S., Nunes, L. de A., Visi, F., & Miranda, E. (2015). One Micro Song, Three Ends: an approach for musical composition and an interactive decision machine based on expressive live performance. In *Proceedings of the 11th International Symposium on Computer Music Multidisciplinary Research (CMMR).* Plymouth, United Kingdom.

B.4 Compositions and Performances

- *Music of Motion and Presence* – for saxophone, electric guitar, percussion, and motion sensors
  - Plymouth Art Weekender / Peninsula Arts, The House, Plymouth University, Plymouth, UK, 2016 (with Lara Jones on alto sax and Marco Frattini on percussion).
- *11 Degrees of Dependence* – for saxophone, electric guitar, and motion sensors
  - Creative Tech Week, Harvestworks, New York, US, 2016 (with Ana García Caraballos on alto sax).
  - Nonclassical Club Night curated by Gabriel Prokofiev, International Festival For Artistic Innovation – iFIMPaC 2016, Leeds College of Music, Leeds, UK, 2016 (with Dr Katherine Williams on soprano sax).
  - Peninsula Arts Contemporary Music Festival 2016, Plymouth, UK, 2016 (with Dr Katherine Williams on soprano sax).
- *Tuned Constraint* – for analogue synthesiser and motion sensors
  - ICLI 2016 – International Conference on Live Interfaces, University of Sussex, 2016.
  - International Metabody Forum, Brunel University / Artaud Performance Centre, London, UK, 2016.
  - Practice Research Symposium, Plymouth University, Plymouth, UK, 2016.
- *Kineslimina* – for viola, electric guitar, and motion sensors (with Esther Coorevits on viola)
  - Peninsula Arts Contemporary Music Festival 2016, Plymouth, UK, 2016
  - MuSA 2015 – Sixth International Symposium on Music/Sonic Art: Practices and Theories, Karlsruhe, Germany, 2015.
  - Gala Concert of the 11th International Symposium on Computer Music Multidisciplinary Research (CMMR), Plymouth University, Plymouth, UK, 2015.

B.5 Talks and Seminars

- *Methods and Technologies for the Analysis and Interactive Use of Body Movements in Instrumental Music Performance*, McGill University, Montreal, Canada, 2016.
- *Instrumental gestures in music composition and augmented music performance*, in Music Seminars Series, Music Department, Plymouth University, Plymouth, UK, 2016.
- *Gestures and Embodied Meaning in Performances with Traditional Musical Instruments*, in NYU MARL: Music and Audio Research Laboratory 2015–16 Talk Series, New York, US, 2015.
- *Body Movements and Embodied Meaning in Performances with Traditional Musical Instruments*, in fourMs seminars, Department of Musicology, Universitetet I Oslo, Oslo, Norway, 2015.
- *Gesture, Body Movement, Musical Experience*, in ICCMR Doctoral Seminar Series, Plymouth University, Plymouth, UK, 2014.

B.6 Workshops

- *Motion and Music Workshop 2: from movement to analysis*, SysMus 2016 International Conference of Students of Systematic Musicology, University of Jyväskylä, Jyväskylä, Finland, 2016. In collaboration with Marc Thompson, Birgitta Burger, Juan Ignacio Mendoza, and Esther Coorevits.
- *Kinefy Workshop: movement interaction with Max and Processing*, Creative Tech Week, Harvestworks, New York, US, 2016. In collaboration with Andrew Telichan Phillips.
- *Motion and Music Workshop: processing and performing gesture motifs for computer music*, Symposium on Computer Music Multidisciplinary Research (CMMR), Plymouth, UK, 2015. In collaboration with Luiz Naveda.