



**HAL**  
open science

# Synthesis of Music Performances: Virtual Character Animation as a Controller of Sound Synthesis

Alexandre Bouënard

► **To cite this version:**

Alexandre Bouënard. Synthesis of Music Performances: Virtual Character Animation as a Controller of Sound Synthesis. Modeling and Simulation. Université de Bretagne Sud; Université Européenne de Bretagne, 2009. English. NNT: . tel-00497292

**HAL Id: tel-00497292**

**<https://theses.hal.science/tel-00497292v1>**

Submitted on 3 Jul 2010

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



**Ph.D. THESIS**

presented by

delivered by the Université Européenne de Bretagne

For obtaining the degree of:

**DOCTOR OF PHILOSOPHY** of the **UNIVERSITÉ DE BRETAGNE SUD**

Mention: **STIC**

Doctoral College **SICMA**

**Alexandre Bouënard**

Laboratoire de Recherche Informatique  
et ses Applications de Vannes et Lorient

# **Synthesis of Music Performances: Virtual Character Animation as a Controller of Sound Synthesis**

**Ph.D. Thesis defended on December 10th, 2009**  
for the examination committee composed of:

**René Caussé**

Senior Researcher, Institut de Recherche et Coordination  
Acoustique/Musique, France / President

**Christian Jacquemin**

Professor, Université Paris XI, France / Reviewer

**Ronan Boulic**

Senior Researcher, Ecole Polytechnique Fédérale de Lausanne,  
Switzerland / Reviewer

**Sofia Dahl**

Assistant Professor, Aalborg University, Denmark / Examiner

**Sylvie Gibet**

Professor, Université de Bretagne Sud, France / Advisor

**Marcelo M. Wanderley**

Associate Professor, McGill University, Canada / Advisor



" For seeing life is but a motion of limbs, the beginning whereof is in some principal part within, why may we not say that all automata (engines that move themselves by springs and wheels as doth a watch) have an artificial life? For what is the heart, but a spring; and the nerves, but so many strings; and the joints, but so many wheels, giving motion to the whole body? (...) Art goes yet further, imitating that rational and most excellent work of nature, man. "

T. Hobbes. *LEVIATHAN or the Matter, Form and Power of a Commonwealth Ecclesiastical and Civil*. London, 1651.



# Abstracts

## Résumé

Ces dernières années ont vu l'émergence de nombreuses interfaces musicales ayant pour objectif principal d'offrir de nouvelles expériences instrumentales. La spécification de telles interfaces met généralement en avant l'expertise des musiciens à appréhender des données sensorielles multiples et hétérogènes (visuelles, sonores et tactiles). Ces interfaces mettent ainsi en jeu le traitement de ces différentes données pour la conception de nouveaux modes d'interaction.

Cette thèse s'intéresse plus spécifiquement à l'*analyse*, la *modélisation* ainsi que la *synthèse* de situations instrumentales de percussion. Nous proposons ainsi un système permettant de synthétiser les retours visuel et sonore de performances de percussion, dans lesquelles un percussionniste virtuel contrôle des processus de synthèse sonore.

L'étape d'analyse montre l'importance du contrôle de l'extrémité de la mailloche par des percussionnistes experts jouant de la timbale. Cette analyse nécessite la capture préalable des gestes instrumentaux de différents percussionnistes. Elle conduit à l'extraction de paramètres à partir des trajectoires extrémité capturées pour diverses variations de jeu. Ces paramètres sont quantitativement évalués par leur capacité à représenter ces variations.

Le système de synthèse proposé dans ce travail met en oeuvre l'animation physique d'un percussionniste virtuel capable de contrôler des processus de synthèse sonore. L'animation physique met en jeu un nouveau mode de contrôle du modèle physique par la seule spécification de la trajectoire extrémité de la mailloche. Ce mode de contrôle est particulièrement pertinent au regard de l'importance du contrôle de la mailloche mis en évidence dans l'analyse précédente. L'approche physique adoptée est de plus utilisée pour permettre l'interaction du percussionniste virtuel avec un modèle physique de timbale.

En dernier lieu, le système proposé est utilisé dans une perspective de composition musicale. La construction de nouvelles situations instrumentales de percussion est réalisée grâce à la mise en oeuvre de partitions gestuelles. Celles-ci sont obtenues par l'assemblage et l'articulation d'unités gestuelles canoniques disponibles dans les données capturées. Cette approche est appliquée à la composition et la synthèse d'exercices de percussion, et évaluée qualitativement par un professeur de percussion.

**Mots clé :** Animation par Ordinateur, Informatique Musicale, Interaction Geste Instrumental-Son.

## Abstract

The last few decades have witnessed the emergence of a plethora of musical interfaces aiming at expanding musical performance experiences. The design of these interfaces generally emphasizes the expertise of musicians in managing heterogenous sensory informations (visual, sound and tactile). Such musical interfaces involve therefore the processing of these various sensory data for designing novel interaction modes.

This thesis addresses more specifically the *analysis*, *modeling* and *synthesis* of percussion performances. We propose a system that realizes the synthesis of the visual and sound feedback of percussion performances in which a virtual percussionist controls sound synthesis processes.

The analysis step of our work shows the importance of the fine control of mallet extremity trajectories by expert percussion performers playing timpani. It includes the collection of instrumental gesture data from several percussionists. We extract movement parameters from the recorded mallet extremity trajectories for different percussion playing variations. Such parameters are quantitatively evaluated with respect to their ability to represent the various playing variations under study.

We then propose a system for synthesizing timpani performances involving the physical modeling of a virtual percussionist that interacts with sound synthesis processes. The physical framework includes a novel scheme for controlling the motion of the virtual percussionist solely by the specification of trajectories of mallet tips. This control mode is shown to be consistent with the predominant control of mallet extremity presented in the previous analysis step. The physical approach is also used for allowing the virtual percussionist to interact with a physical model of a timpani.

Finally, the proposed system is used in a musical performance perspective. A composition process based on gesture scores is proposed to achieve the synthesis of novel percussion performances. Such gesture scores are obtained by the assembly and articulation of gesture units available in the recorded data. This compositional approach is applied to the synthesis of several percussion exercises, and is informally evaluated by a percussion professor.

**Key words:** Computer Animation, Computer Music, Instrumental Gesture-Sound Interaction.



# Acknowledgments

I would like to thank first Sylvie Gibet (Université de Bretagne Sud, VALORIA lab.) and Marcelo M. Wanderley (McGill University, IDMIL lab.) for giving me the opportunity to work in their respective research groups, as well as for all of their support and enthusiasm over these years. Let me highlight their courage, tenacity and constant rightness while accompanying me in this Ph.D. project.

I would like also to thank them for making this "backpack Ph.D." possible. I consider their trust as a privilege, letting me cross several times the Atlantic ocean between Vannes and Montreal for accomplishing this work. It has for sure changed both my professional and personal lives forever.

---

I am grateful to Christian Jacquemin (Université Paris XI) and Ronan Boulic (EPFL), for all of their time and guidance throughout the Ph.D. review process. I would like also to thank René Caussé (IRCAM) and Sofia Dahl (Aalborg University) who have accepted the task of examining this dissertation.

Many thanks also to all the members of the VALORIA lab. for making my several venues in Vannes really encouraging and interesting periods. Especially, I would like to thank Frédéric, Gersan, Gildas, Jean-François, Nicolas (VALORIA-mafia style rocks on!), Pierre-François and Salah for the daily discussions.

Thanks as well to Alban, Céline, Charly, Julien (for discussions, support and the camera shooting), Kahina (for sharing good moments in the office, for the Ph.D. party organization, and for the "table" keyword ;-◇), Kyle, Nicolas, Pierre, Roméo, Salma (for the Ph.D. party organization), Sébastien, Thibault, Vincent and Youen, for their general support when writing this Ph.D. dissertation.

Many thanks also to Sylviane, whose administrative efficiency has been more than helpful over these years, especially when I was abroad.

Special thanks to all Music Tech-ies for contributing to make each of my stay at the IDMIL lab. a stimulating environment for conducting my research. Among them, thanks to Avrum, Erika, Joe M., Joe T., Mark, Marlon, Michael, Rodolphe, Steve (especially for the daily discussions and his OSC-linux mastering knowledge), Vanessa and Vijay for expanding my horizon in music tech research through their studies. Thanks to Bertrand, Vincent F. and Vincent V. for the "French connection" ;-◇, many thanks to Erwin for helping me during the time-consuming task of gathering motion capture data.

Thanks also to Darryl for easing many technical tasks at the IDMIL.



I am also grateful to Fabrice Marandola (Schulich School of Music, McGill University) and Sofia Dahl (Aalborg University) for their expert insights about percussion performances. Thanks as well to all percussionists who have made possible such a study about propably one their most personal musical material, their instrumental gestures.

In a more general perspective, I would like to wish all the best for the next future to all the previously cited people and those I have unfortunately forgotten. We will for sure meet each other again even if we do not know where and when, but do we really need a determined timing and appointment?

---

Eventually, here come some more personal acknowledgments.

My unmeasured recognition goes to my dear family (Alain, Anne-Soizic, Joëlle, Julie and Pierre) for their every-day understanding and for coping with physical separation. I know these past few years have not been easy for you. Thanks for being part of my life, for welcoming me again and again, as truly as it was the first time!

I am also thankful to my family-in-law (Caro, J-Lo, Lise and Walid) for making each of my stay in Quebec a wonderful discovery, particularly in St-Narcisse. Each time I spent here with all of you was really a breath of fresh air, it has for sure contributed to the achievement of this work. Private joke for J-Lo: the "pousseux d'crayon" has finally graduated ;-◇.

I would like also to thank my dear friends: Alex, Anne, Boris, Flavie, Flo B., Flo, Jean-Louis, Nico, Marion, Maud, Pedro and Romain for their every-day understanding. Thanks as well to Audi, Brioche, Caro, Laureen and Nélène, who came to Montreal only for seeing my eyes, and also for the Quebec I am not fooled ;-◇. I am also grateful to Vincent and Julie who introduced me to Montreal's life.

To all of you, time and distance can be ennemies as you may know, could we just take it as an allied and simply keep on making fun of it?!

Last but not least, my warmest and unlimited thanks go to Annie, whose love, patience and encouragement have made all the difference.

---

A new life starts from these lines, thanks to Music for guiding it. I now hope I will be given the chance to write down many others.



# Table of Contents

<b>Abstracts</b>	<b>v</b>
<b>Acknowledgments</b>	<b>vii</b>
<b>Table of Contents</b>	<b>1</b>
<b>1 Introduction</b>	<b>5</b>
1.1 Interests and Challenges in Synthesizing Percussion Performances	5
1.2 Research Questions and Contributions . . . . .	7
1.3 Thesis Structure . . . . .	8
<b>2 State of the Art</b>	<b>11</b>
2.1 Computer Animation . . . . .	11
2.1.1 Virtual Character Models . . . . .	12
2.1.1.1 Kinematics-based Representation . . . . .	12
2.1.1.2 Physics-based Representation . . . . .	12
2.1.2 Motion Control of Virtual Characters . . . . .	14
2.1.2.1 Kinematics-based Methods . . . . .	15
2.1.2.1.1 Forward Kinematics . . . . .	15
2.1.2.1.2 Inverse Kinematics . . . . .	16
2.1.2.1.3 Conclusion . . . . .	20
2.1.2.2 Physics-based Methods . . . . .	21
2.1.2.2.1 Forward Dynamics . . . . .	22
2.1.2.2.2 Inverse Dynamics . . . . .	22
2.1.2.2.3 Conclusion . . . . .	27
2.1.2.3 Hybrid Methods . . . . .	27
2.1.2.3.1 Kinematics, Kinetics and Dynamics . . . . .	28
2.1.2.3.2 Motion-driven Physics-based Methods . . . . .	28
2.1.2.3.3 Conclusion . . . . .	29
2.1.3 Summary . . . . .	30
2.2 Computer Music . . . . .	31
2.2.1 Sound Synthesis Models . . . . .	31
2.2.1.1 Descriptive Sound Synthesis . . . . .	31
2.2.1.1.1 Sampling Models . . . . .	31
2.2.1.1.2 Spectral Models . . . . .	32
2.2.1.1.3 Conclusion . . . . .	33
2.2.1.2 Physics-based Sound Synthesis . . . . .	34
2.2.1.2.1 Digital Solution of the Wave Equation . . . . .	34
2.2.1.2.2 Modal and Mass-Spring Models . . . . .	38

2.2.1.2.3	Conclusion . . . . .	39
2.2.2	Gestural Control of Sound Synthesis Processes . . . . .	40
2.2.2.1	Acquisition, Mapping and Digital Instruments . . . . .	40
2.2.2.1.1	Direct and Indirect Acquisition . . . . .	41
2.2.2.1.2	Mapping . . . . .	41
2.2.2.1.3	Digital Music Instruments . . . . .	42
2.2.2.2	Gesture Analysis and Modeling . . . . .	43
2.2.2.3	Conclusion . . . . .	43
2.2.3	Summary . . . . .	44
2.3	Virtual Music Performances . . . . .	46
2.3.1	Computer Animation Contributions . . . . .	46
2.3.2	Computer Music Contributions . . . . .	47
2.3.2.1	Percussion Sound Synthesis . . . . .	47
2.3.2.2	Percussion Gestural Controllers . . . . .	47
2.3.2.3	Virtual Percussion Gesture Models . . . . .	48
2.3.3	Summary . . . . .	48
<b>3</b>	<b>Overview of the Approach</b>	<b>51</b>
3.1	Analysis . . . . .	51
3.2	Synthesis . . . . .	53
3.3	Evaluation . . . . .	54
<b>4</b>	<b>Analysis of Timpani Percussion Performances</b>	<b>55</b>
4.1	Timpani Basics . . . . .	55
4.1.1	Equipment . . . . .	56
4.1.2	Acoustics . . . . .	56
4.1.3	Playing Techniques . . . . .	58
4.2	Motion Capture Protocol and Database . . . . .	61
4.3	Analysis . . . . .	62
4.3.1	Segmentation of Motion Capture Data . . . . .	62
4.3.2	Methodology . . . . .	63
4.3.3	Percussion Grips . . . . .	64
4.3.4	Playing Modes . . . . .	68
4.3.5	Discussion . . . . .	68
4.3.5.1	Nature of the Parameters . . . . .	68
4.3.5.2	Classification Technique . . . . .	69
4.4	Conclusion . . . . .	71
<b>5</b>	<b>Synthesis of Timpani Percussion Performances</b>	<b>73</b>
5.1	Overview . . . . .	73
5.2	Physics-based Modeling and Motion Control . . . . .	75

5.2.1	Virtual Character Modeling . . . . .	76
5.2.1.1	Representation and Anthropometry . . . . .	76
5.2.1.2	Joints . . . . .	76
5.2.2	Motion Control . . . . .	77
5.2.2.1	<i>ID</i> Motion Control . . . . .	77
5.2.2.2	Hybrid Motion Control . . . . .	78
5.2.3	Results . . . . .	80
5.2.3.1	Qualitative Evaluation . . . . .	81
5.2.3.2	Quantitative Evaluation . . . . .	83
5.2.4	Conclusion . . . . .	85
5.3	Interaction between Motion and Sound Synthesis . . . . .	86
5.3.1	Asynchronous Client-Server Architecture . . . . .	86
5.3.2	Motion-Sound Physics Interaction . . . . .	87
5.3.3	Results . . . . .	88
5.4	Conclusion . . . . .	91
<b>6</b>	<b>Musical Application and Evaluation</b>	<b>93</b>
6.1	Gesture Edition and Composition . . . . .	93
6.2	Musical Evaluation of Virtual Percussion Performances . . . . .	96
6.2.1	General Comments . . . . .	98
6.2.2	Validation Exercises . . . . .	98
6.2.3	Extrapolation Exercises . . . . .	100
6.2.3.1	Playing Modes . . . . .	100
6.2.3.2	Tempo Variations . . . . .	101
6.2.3.3	Impact Location Variations . . . . .	102
6.3	Discussion: Advantages and Limitations . . . . .	103
6.3.1	Instrumental Gesture Simulation . . . . .	103
6.3.1.1	Advantages . . . . .	103
6.3.1.2	Limitations . . . . .	105
6.3.2	Motion Database and Gesture Edition . . . . .	107
6.3.2.1	Advantages . . . . .	108
6.3.2.2	Limitations . . . . .	108
6.4	Conclusion . . . . .	109
<b>7</b>	<b>Conclusion and Future Work</b>	<b>111</b>
7.1	Conclusion . . . . .	111
7.2	Future Work . . . . .	112
7.2.1	Analysis . . . . .	112
7.2.2	Synthesis . . . . .	113
7.2.3	Musical Applications . . . . .	114

---

<b>A</b>	<b>Sound Synthesis Models</b>	<b>117</b>
A.1	d'Alembert Equation . . . . .	117
A.2	d'Alembert Equation: Fourier's Solution . . . . .	118
A.3	Modal Decomposition and d'Alembert Equation . . . . .	120
A.4	Finite-Difference and Finite-Element Formulations . . . . .	122
A.5	Digital Waveguide Formulation . . . . .	126
A.6	Modal Synthesis Formulation . . . . .	128
<b>B</b>	<b>Motion Capture Protocol</b>	<b>131</b>
<b>C</b>	<b>Bibliography</b>	<b>135</b>
C.1	Computer Animation . . . . .	135
C.2	Computer Music . . . . .	150
C.3	Instrumental Performance Analysis . . . . .	163
<b>D</b>	<b>List of Figures</b>	<b>167</b>
<b>E</b>	<b>List of Tables</b>	<b>169</b>
<b>F</b>	<b>List of Equations</b>	<b>171</b>
<b>G</b>	<b>List of Algorithms</b>	<b>173</b>

# 1

## Introduction

People have always practiced music and built various musical instruments. Most of the history in the development of musical instruments can be seen as a continuous quest for new instrumental experiences. The advent of computers has rapidly made possible to expand sound possibilities, for instance by recording, sampling, transforming and playing back any sound that can be produced by musical instruments. Later, with the everyday use of computer technologies, the last decades have gone beyond the exploration of sound possibilities, namely by focusing on new ways of thinking interfaces and interaction models to control sound production processes. A plethora of software and hardware musical interfaces have then been proposed, with the common aim of expanding musical performance practice to territories previously unaccessible with traditional musical instruments. One of such territory at the heart of this dissertation is the synthesis of musical performances. Such virtual performances involve virtual characters, where both the gestures made by a virtual musician and the resulting sounds generated by its interaction with a virtual musical instrument are simulated.

### 1.1 Interests and Challenges in Synthesizing Percussion Performances

This dissertation addresses the synthesis of musical situations with a particular focus on percussion performances. These latter are unique among musical performances, both regarding the nature of percussion instruments and the way performers are playing it.

It is generally admitted that the act of percussion for making music dates back to the prehistoric era, when hand claps or body strokes were used to convey expressions [Ros00]. Percussion instruments are therefore of special interest due to the fact there is a tiny distinction between the notions of object and instrument. As every object has an inherent acoustic response to everyday actions such



as hitting, scraping or slaping, every object can be considered as a percussion instrument [Coo01]. For instance, a table could hardly been played as a string instrument, whereas people can easily strike or roll tools on it while feeling and hearing the resulting feedback.

This blurred frontier between everyday objects and percussion instruments may explain why a large amount of percussion instruments is nowadays available, and therefore why percussion performers are highly trained to play on a large collection of instruments [Ros00]. During instrumental situations, performers especially involve complex mechanisms for striking a percussion instrument according to a desired sound effect. Interestingly, percussion performers are not only experts in maintaining rhythm, but also in creating a wide range of timbres even from a single percussion object. Supporting evidence for this claim can for instance be heard and seen concerning John Bonham's drum outtakes, or Max Roach's "hi hat" solo. Percussion performers often have to learn to consider percussion instruments with their inherent and subtle timbre capabilities in addition to rhythmic patterns. Such consideration can lead to learning more complex and abstract mechanisms, such as the elaboration of an own understanding of gesture-sound mechanisms. Among them, a noticeable one's is referred to as sound projection, a mechanism by which percussion performers propose an interpretation of the resulting sounds. John Cage explains this phenomenon as the "feeling given by certain blocks of sounds (...) giving a sense of prolongation, a journey into space" [Pat08]. To that mean, percussionists develop an expert knowledge of how to shape their gestures according to the desired sound effects. This involves for instance an accurate control of preparatory gestures and beat impact forces, especially considering the short time duration of stroke impacts [Wag06].

For all these reasons, synthesizing percussion performances creates challenges in transposing these real-world phenomena into virtual situations. In fact, due to the complex mechanisms occuring during (real) percussion performances, synthesizing percussion performances necessitates the availability of motion data recorded for ensuring the realism of the synthesized percussion motion. Such data are also helpful for conducting a fine analysis of percussion motion under various playing conditions.

But what is exactly the interest of synthesizing percussion performances? The analysis of percussion performances may lead to understanding these underlying complex mechanisms. The use of such *a priori* knowledge can be helpful for, on the one hand, guiding the design of a system able to synthesize virtual percussion performances, and on the other hand for its evaluation. The interest of such

system is to provide a realistic modeling of percussion performances, as well as to offer a tool for creating novel performances that may go beyond this reality. Pushing the limits of the reality can involve for instance various virtual character models or different interaction schemes between the simulated motion and sound.

Furthermore, a system able to synthesize percussion performances can be used in many musical applications. Among them, one application could consist in providing a pedagogical tool dedicated to percussion professors and students for exploring synthesized percussion performances with both their visual and sounding feedback. Such a tool could also be used in a musical composition context, by offering a user interface where novel percussion performances can be synthesized given various inputs regarding percussion playing conditions at the gestural level.

## 1.2 Research Questions and Contributions

The modeling and motion control of a percussionist interacting with sound synthesis necessitates to answer the following research questions:

- a fine understanding of the motion involved in real music performances. In other words, what are the essential motion characteristics that have to be preserved when synthesizing music performances, especially considering interaction and also preparatory gestures?
- a complete modeling of a real performer, especially in the case of physics-based simulations of music performances. In other words, how can one move from simulations involving simple rigid bodies to simulations involving the actions of a physics-based model of a real performer?
- a correspondence of the motion control rules for putting the virtual performer into motion, with real motion data. In other words, how can one ensure that the motion control paradigms involved in the animation of a virtual performer are consistent with real motion data?
- an interaction scheme between the simulated gestures of a virtual performer with sound synthesis models. In other words, how can one explicitly (at the physics level) control sound processes by the actions of the virtual performer?

We propose a system that realizes the synthesis of the visual and sound feedback of percussion performances in which a virtual percussionist controls sound synthesis processes. The synthesis system is conducted by percussion motion data. It lays its foundations on the analysis of percussion movements, as well as

on its musical evaluation.

The analysis step of our work shows the importance of the fine control of mallet extremity trajectories by expert percussion performers playing timpani. It includes the collection of instrumental gesture data from several percussionists. We extract movement parameters from the recorded mallet extremity trajectories for different percussion playing techniques. Such parameters are quantitatively evaluated with respect to their ability to represent and discriminate the various playing techniques under study [BWG08, BWG09a, BWG09b].

We then propose a system for synthesizing timpani performances involving the physical modeling of a virtual percussionist that interacts with sound synthesis processes. The physical framework includes a novel scheme for controlling the motion of the virtual percussionist solely by the specification of trajectories of mallet extremities. This control mode is shown to be consistent with the predominant control of mallet extremity presented in the previous analysis step. The physical approach is also used for allowing the virtual percussionist to interact with a physical model of a timpani [BGW09a, BGW09b, BGW09c].

Finally, the proposed system is used in a musical performance perspective. A composition process based on gesture scores is proposed to achieve the synthesis of novel percussion performances. Such gesture scores are obtained by the assembly and articulation of gesture units available in the recorded data. This compositional approach is applied to the synthesis of several percussion exercises, and is informally evaluated by a percussion professor [BGW08, BWG09, BWGM09].

### 1.3 Thesis Structure

The remainder of this document is organized as follows.

Chapter 2 reviews related works, covering each module that is necessary for proposing virtual percussion performances involving animated virtual characters that can interact with sound synthesis processes. This includes previous works related to *Computer Animation*, *Computer Music*, as well as percussion-related models.

Chapter 3 presents an overview of the proposed system for synthesizing percussion performances.

Chapter 4 conducts an analysis of motion data recorded from real percussion

performances. This analysis leads to the extraction and evaluation of motion parameters for discriminating between various playing conditions.

Chapter 5 presents and evaluates a complete modeling of a virtual character, as well as a hybrid motion control system dedicated to the simulation of percussion movements. A drum model and the way the virtual percussionist can interact with it are also proposed.

Chapter 6 conducts experiments to evaluate the musical possibilities of the proposed system in synthesizing novel percussion performances from recorded motion clips.

Finally, Chapter 7 concludes this document, and proposes future perspectives for extending this work.



# 2

## State of the Art

In this section, we provide a state of the art on the synthesis of music performances, focusing on all the protagonists that act and are in interaction during an instrumental situation. It includes the modeling and animation of a virtual character, as well as its interaction with sound synthesis processes.

We draw an overview of the main virtual human models as well as the animation techniques that steer such representations to a desired motion (*Computer Animation*, section 2.1).

We focus on the concepts that have been elaborated for synthesizing sounds, and specifically how these models are related to instrumental gestures (*Computer Music*, section 2.2).

Although *Computer Animation* and *Computer Music* research fields have grown quite independantly along the years, we refer to works that can be considered at the frontier of these areas (section 2.3). It includes an overview of various models for synthesizing real/virtual music performances, which is the core subject of our contributions.

### 2.1 Computer Animation

Nowadays, the exploration of virtual environments by virtual characters is a widespread issue in many applications. To cite a few, it ranges from video games to the film industry, as well as human movement analysis and perception. Whatever application is targetted, a virtual character model and animation technique has to be chosen, depending on the final application.

The *Computer Animation* community commonly divides this problem into two stages: 1) provide a model (or representation) of the virtual character itself (subsection 2.1.1), and 2) develop techniques for controlling this representation

as wanted over time (subsection 2.1.2).

### 2.1.1 Virtual Character Models

Many virtual character models have been proposed in the literature, but each has obviously been highly inspired by the human anatomy. A common basis for modeling the human anatomy is to consider a set of virtual limbs connected by virtual joints. It generally involves a hierarchical tree structure (an example is provided in FIG. 2.2(b)) describing the dependencies between these virtual entities. Typically, two main representations are used in *Computer Animation*, the kinematics-based and the physics-based representations.

#### 2.1.1.1 Kinematics-based Representation

Virtual characters can be represented by a *kinematic* representation  $\mathbf{T}^{\mathbf{K}}$ , which gives mainly a *geometrical* interpretation of the human anatomy. Such a representation defines implicitly the limbs (or bones) of the virtual human by segments between paired joints. This hierarchical structure is defined by a root joint  $\mathbf{j}_r^{\mathbf{K}}$  (usually the pelvis) and a set of child joints  $\mathbf{j}^{\mathbf{K}}$ , EQU. (2.1).

The kinematic system  $\mathbf{T}^{\mathbf{K}}$  is then defined by the observation of its state over time, *i.e.* a kinematic configuration (or *pose*)  $\mathbf{q}^{\mathbf{K}}$ .  $\mathbf{q}^{\mathbf{K}}$  is composed of the root joint position  $\mathbf{r}$  and the angular state of each child joint  $\Theta^{\mathbf{K}}$ , EQU. (2.2). The latter may have several *degrees of freedom* (or *DoFs*), thus constraining the motion of the virtual character.

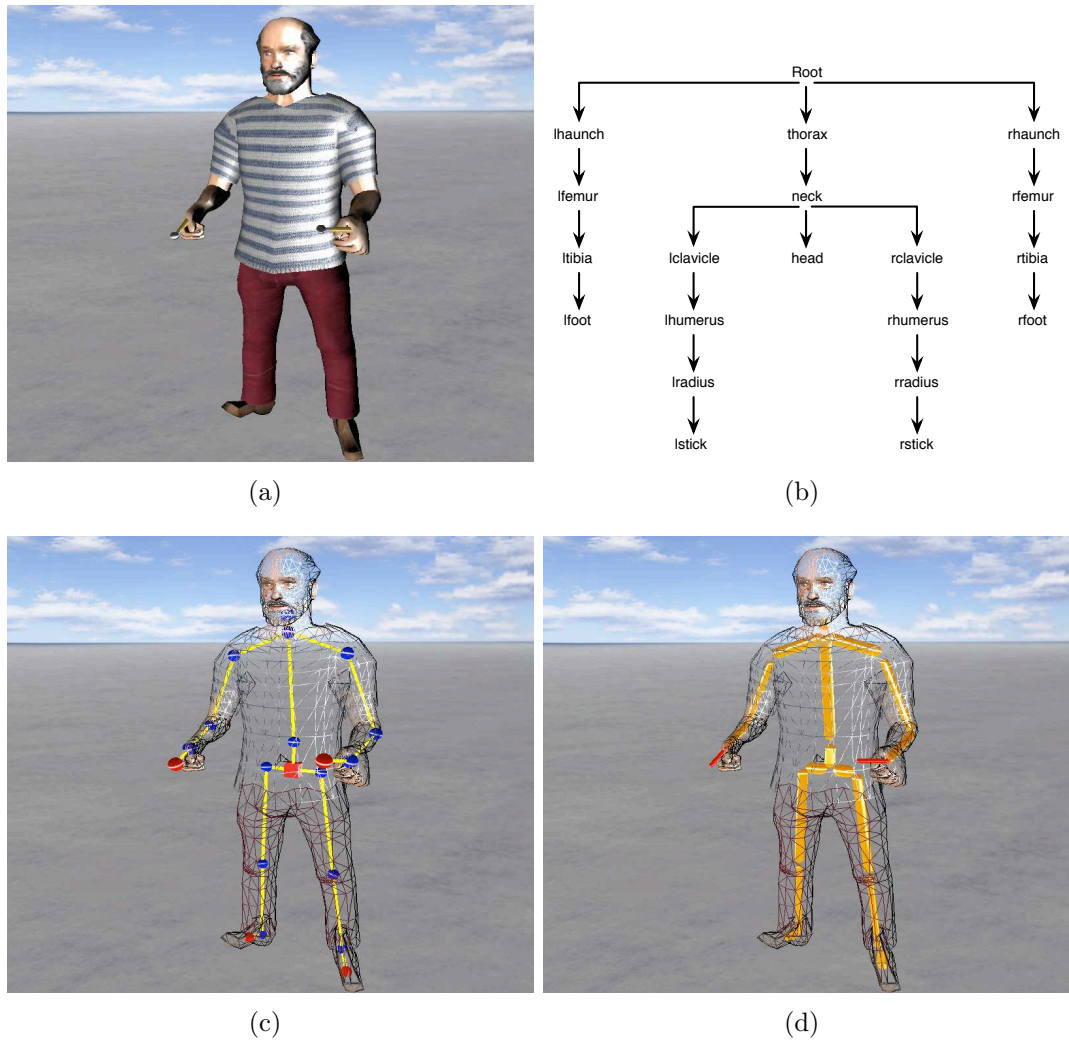
$$\mathbf{T}^{\mathbf{K}} = [\mathbf{j}_r^{\mathbf{K}}, \mathbf{j}^{\mathbf{K}} = \{\mathbf{j}_i^{\mathbf{K}}\}_{i \in [1..n]}] \quad (2.1)$$

$$\mathbf{q}^{\mathbf{K}} = [\mathbf{r}, \Theta^{\mathbf{K}} = \{\Theta_i^{\mathbf{K}}\}_{i \in [1..n]}] \quad (2.2)$$

FIG. 2.2(c) depicts an example of kinematic representation of a virtual character, with its root position (red box), set of joints (spheres) and implicit limbs (yellow lines). A standard kinematic representation for modeling virtual characters is commonly used for this type of approach [HA09], easing the comparison, sharing and reusability of representations.

#### 2.1.1.2 Physics-based Representation

The *physical* representation  $\mathbf{T}^{\mathbf{D}}$  of a virtual character provides a *dynamic* interpretation of the human anatomy. In this case, the bones are explicitly modeled by a set of rigid solids  $\mathbf{s}^{\mathbf{D}}$  articulated by mechanical joints  $\mathbf{j}^{\mathbf{D}}$ , EQU. (2.3). Every solid composing this physical representation  $\mathbf{s}^{\mathbf{D}}$  is parameterized by physical characteristics such as its mass  $m$ , density  $d$  and inertia tensor  $I$ . Mechanical joints



**FIG. 2.1.** Kinematics and physics representations of a virtual character: (a) final virtual character, (b) the hierarchical tree structure, (c) an example of an underlying kinematic representation with its root (red box), set of joints (spheres) and implicit limbs (yellow lines), (d) an example of an underlying physics representation with its explicit limbs (capsule-shaped rigid bodies).



$\mathbf{j}^D$  can also be characterized by several DoFs constraining the motion, as well as physical features influencing the relative behavior between solids at the joint level.

The observation of the state of  $\mathbf{T}^D$  over time leads to a dynamic configuration  $\mathbf{q}^D$ , composed of each rigid body linear-angular position  $\mathbf{r}^D$  as well as each mechanical joint angular position  $\Theta^D$ , EQU. (2.4).

$$\mathbf{T}^D = [\mathbf{s}^D = \{m_i^D, d_i^D, I_i^D\}_{i \in [1 \dots n]}, \mathbf{j}^D = \{k_{s,j}^D, k_{d,j}^D\}_{j \in [1 \dots m]}] \quad (2.3)$$

$$\mathbf{q}^D = [\mathbf{r}^D = \{r_i^D, \Theta_{r,i}^D\}_{i \in [1 \dots n]}, \Theta^D = \{\Theta_j^D\}_{j \in [1 \dots m]}] \quad (2.4)$$

FIG. 2.2(d) shows an example of a physic representation of a virtual character made of capsule-shaped rigid bodies. These models usually assume a uniform density distribution of the rigid bodies. Human body measures can be considered for initializing the physical properties of the rigid bodies composing such a representation [DG67, ST09].

### 2.1.2 Motion Control of Virtual Characters

The characteristics of these different virtual character representations previously presented intrinsically differ from the nature of human motion features they focus on. They differ as well as on the low/high level control properties and environment interactions that they can take into account.

The *kinematic* model can be considered as an *effect-centered* representation, offering a well formulated formalism for a high task control level. It however provides no explanation of the causes that are at the origin of the motion system  $\mathbf{T}^S$ , and hardly provides the opportunity to integrate interactions with its surrounding environment. Kinematics-based animation techniques will therefore use methods on the top of the geometrical observation  $\mathbf{q}^K$  of the human motion, in order to reproduce, adapt or modify this observation (section 2.1.2.1).

Conversely, the *physical* representation can be considered as a *cause-centered* model since it makes available a responsive model to the application of forces and torques on the system  $\mathbf{T}^D$ . It gives a quite low level control scheme, but facilitates its interaction with the surrounding environment. Physics-based animation techniques aim at building control paradigms to *produce* adequate forces and torques that are the cause of the human motion based on the state observation  $\mathbf{q}^D$  (section 2.1.2.2).

Finally, some contributions aim at taking advantage of each representation's assets, therefore mixing kinematics-based and physics-based animation techniques.

We will refer to this kind of formulations as hybrid methods (section 2.1.2.3).

It should be noted that this state of the art mainly focuses on the various *control* models that have been introduced in the *Computer Animation* community. The recall and derivation of equations is justified as a means of precisely identifying the nature (geometric, kinematic, dynamic) of the parameters that each model refers to.

### 2.1.2.1 Kinematics-based Methods

The animation of virtual characters has a huge history, which begins with hand-made (cartoon) productions. Two of the most common kinematics-based computer animation techniques come from this historical background: *keyframing* and *interpolation* (or *inbetweening*). At its initial stage, the process of producing hand-made animations was stamped by its organisational approach (Taylorism), involving firstly senior animators producing action-based drawings (*keyframes*) at targetted specific times of the final animation, and secondly junior animators in charge of linking these *keyframes* by *inbetween* drawings.

#### 2.1.2.1.1 Forward Kinematics

Inspired by the early works of Walt Disney Studio in the late 1920's, most of *Computer Animation's* work began therefore with adapting purely traditional 2D techniques to the new possibilities offered by computers [Cat78, Las87]. This has led to computer storytelling, keyframe animation (keyframing) [BW71] as well as the elaboration of various interpolation techniques [KB84, Sho85]. Keyframing involves the specification of key postures ( $\mathbf{q}^K$ ) so that the virtual character produces the desired task ( $\mathbf{E}^K$ ). This process is called *forward kinematics*, EQU. (2.5). Then an interpolation technique is used to automatically compute inbetween postures given these keyframes. The interpolation algorithm is of great importance, since it may critically influence the appearance (realism) of the final animation.

$$\mathbf{E}^K = \mathcal{K}(\mathbf{q}^K) \tag{2.5}$$

While such a process is quite natural for producing animations, it rapidly becomes a tedious and time-consuming task depending on the length of the final animation, and moreover depending on the complexity of the model to put into motion. Particularly in the case of virtual character animation, virtual humans can contain up to sixty DoFs (dimension of  $\Theta^K$ ), making this process challenging just for instance to place the hand at a desired position. That is the reason why algorithms have been developed for inferring the kinematic posture of a virtual

character, given only the task to be performed. Such a technique is called *inverse kinematics* (or *IK*).

### 2.1.2.1.2 Inverse Kinematics

The inverse kinematics problem, EQU. (2.6), provides a formal formulation for automatically computing the kinematics posture  $\mathbf{q}^{\mathbf{K}}$  of the virtual character given a desired task  $\mathbf{E}^{\mathbf{K}}$  to accomplish.  $\mathbf{q}^{\mathbf{K}}$  is typically characterized by all the DoFs defining the virtual character’s skeleton, and  $\mathbf{E}^{\mathbf{K}}$  defines the task, generally the linear and/or angular position of limbs’ end-effectors – such as hands or feet configuration, see red spheres on FIG. 2.2(c).

$$\mathbf{q}^{\mathbf{K}} = \mathcal{K}^{-1}(\mathbf{E}^{\mathbf{K}}) \quad (2.6)$$

The difficulty with this formulation mainly lies in the redundancy of the system  $\mathbf{T}^{\mathbf{K}}$  which makes the kinematics problem under-determined, as many different kinematics postures  $\mathbf{q}^{\mathbf{K}}$  can lead to the same desired task  $\mathbf{E}^{\mathbf{K}}$ . More specifically for the human arm, it has been shown that an analytical solution may be determined for an arm model with seven DoFs [Kor85]. Additional numerical methods can even take into account joint limits [TGB00].

However, when considering controlled systems of higher dimension, an analytical solution to the inverse kinematics problem is not envisageable. In this case, numerical methods are generally deployed, involving either optimization or techniques taking advantage of the availability of motion data.

#### 2.1.2.1.2.1 Optimization Methods

Among optimization methods, one of the most widespread solutions to the inverse kinematics problem is to apply a local linearization to EQU. (2.5) for finding the Jacobian matrix  $\mathcal{J}$  of the system to control [Whi69]. This process can be seen as relating the effect of small variations in the joint space to small variations in the task space. Jacobian-based inverse kinematics schemes then rely on the inversion of this Jacobian matrix  $\mathcal{J}$ , EQU. (2.7). This inversion condition is however not always realized due to the fact that in most of the cases  $\mathcal{J}$  is not squared (again due to the redundancy of the system) and possibly singular.

$$\Delta \mathbf{q}^{\mathbf{K}} = \mathcal{J}^{-1}(\mathbf{q}^{\mathbf{K}}) \cdot \Delta \mathbf{E}^{\mathbf{K}} \quad (2.7)$$

Alternatives to the inversion of  $\mathcal{J}$  have consequently been proposed, firstly to solve its invertibility, such as the transpose of the Jacobian matrix  $\mathcal{J}^T$  [WE84], its pseudo-inverse  $\mathcal{J}^\dagger$  [Gre59], or other methods introducing biological functions to solve this inverse problem [GM94]. Secondly, solutions have been proposed

to cope with its singularity [Wam86], leading to an adaptation of the pseudo-inverse  $\mathcal{J}_\lambda^\dagger$  parameterised by a damping factor ( $\lambda$ ) that can be automatically and dynamically computed [MK88].

$$\Delta \mathbf{q}^K = \mathcal{J}_\lambda^\dagger(q^K) \cdot \Delta \mathbf{E}^K + (I - \mathcal{J}^\dagger \cdot \mathcal{J}) \cdot \Delta \mathbf{z}^K \quad (2.8)$$

Other solutions propose the combination of such alternatives while exploiting the redundancy of  $\mathbf{T}^K$  by defining secondary tasks ( $\Delta \mathbf{z}^K$ ) to be performed. This has been firstly done in [Lié77], by expressing these secondary tasks on the null space of  $\mathcal{J}$ , EQU. (2.8).

It has the advantage of allowing the integration and respect of high-level control tasks such as object avoidance [SS91], prioritized physical and shape constraints [BB98, BB04, CB06] as well as ergonomic constraints [YDYY05].

Fundamentally different iterative approaches avoid the Jacobian matrix inversion, such as the Cyclic-Coordinate-Descent (CCD) method [WC91]. It is an heuristic formulation of the inverse kinematics problem, that can be compared to the linearization of EQU. (2.5). More recently a hierarchical CCD algorithm involving the independent treatment of joint sub-groups as well as task priorities has been proposed [KM05].

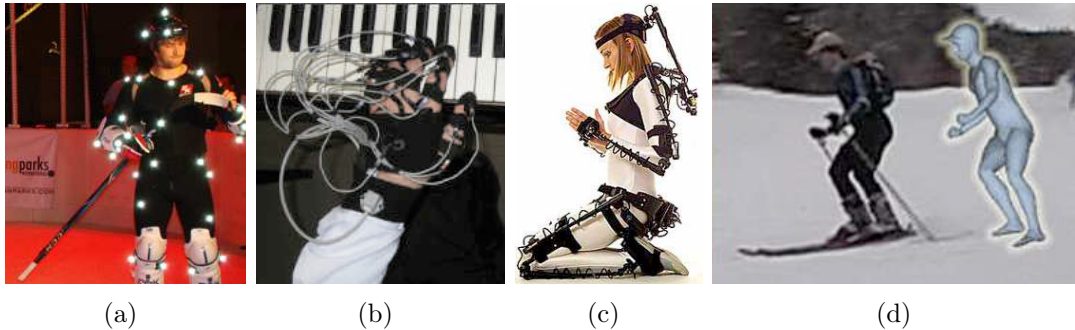
#### 2.1.2.1.2.2 Example-based Methods

Kinematics-based animation techniques are widely used both in the *Computer Animation* research community and in the industrial domain. One open question is how to specify the task to be accomplished by the virtual character, whatever a *forward* or *inverse* kinematics scheme is used. Methods then consist in synthesizing new movement sequences from the capture of examples of human motion. This includes the combination of *IK* formulations with motion data, learning approaches, as well as methods for retrieving, adapting, combining and retargetting existing motion data.

#### Motion capture solutions

A growing demand in human motion data has risen these past few years in industrial fields such as computer games and 3D animation films. This has led to the conception of various hardware for reliably capturing human motion. It includes mainly two classes of hardware: intrusive and non-intrusive systems.

Among intrusive methods, optical systems involve reflective markers put on real humans whose motion can be triangulated by infrared cameras, such as Vicon hardware [Vic09]. The motion of electromagnetic sensors can also be measured relatively to a magnetic reference, such as Polhemus solutions [Pol09]. Another



**FIG. 2.2.** Motion capture systems, from left to right: (a) infrared camera tracking (courtesy of [Vic09]), (b) electromagnetic tracking (courtesy of [Pol09, MKY<sup>+</sup>06]), (c) electromechanical sensors (courtesy of [Mot09]) and (d) video-based tracking (courtesy of [VAV<sup>+</sup>07]).

intrusive method is provided by electromechanical systems which are based on exoskeletons, such as systems provided by Meta Motion [Mot09]. As shown in FIG. 2.2, such hardware solutions can be quite intrusive and therefore possibly constraint the motion that is to be recorded.

On the contrary, non-intrusive methods do not rely on sensors directly in contact with captured human, and involve usually video-based processing methods, FIG. 2.3(d). Such techniques appear to be a promising approach for capturing the human motion in many capture condition [VAV<sup>+</sup>07, Qua09], in both indoor and outdoor environments.

### Motion-driven Inverse Kinematics

Early works among motion-driven inverse kinematics contributions aimed at inferring a kinematic posture target for achieving a desired task, while preserving the characteristics (synergies, styles) of a given motion.

It has been first proposed to interpolate between available motion clips that are similar in the kinematic posture space to a posture that show the desired task [WH97a]. The cost reduction of such an approach due to the required amount of motion clips has been addressed in [RSC01]. [KKK<sup>+</sup>03] then proposed a formulation for computing a weight matrix representing the style of a single motion example at the joint level. Such weight matrix combined with an *IK* process allows to synthesize new movements while preserving the style of the original motion.

Data reduction techniques have also been considered for representing motions in latent spaces. This generally involves the use of the Principal Component Analysis (*PCA*) [AM00]. It has then been shown that the respect of constraints

in such PCA space combined with a prioritized *IK* formulation can lead to the synthesis of new motions [CBT07]. Other solutions have been proposed for taking advantage of such latent spaces, for instance for modeling time-varying joint synergies by linear functions [RB09].

### Learning Methods

The basic approach of learning methods consists in learning the mapping between kinematic configurations and end-effector tasks.

Hidden markov models [BH00] as well as linear dynamic systems [LWS02] have been used for learning the style of human motion examples. Models based on radial basis functions [RSC01] and Gaussian latent variables [GMHP04] have also been used for learning kinematic configurations with the aim of solving the inverse kinematics problem. This problem has been solved by the determination of the Jacobian inverse matrix through the learning of local transformations [GM03]. The mapping between poses and style parameters can also be learnt by using Gaussian mixture models [WLZ06]. Physical constraints can be learnt from existing motion by clustering [LHP05], and user-defined constraints can be taken into account by combining these latter with a statistical model learnt from motion capture data [CH07]. Another solution is to use Monte Carlo casting alongside with a sequential particle filtering formulation [CA08]. Finally, another approach is presented in [AJG09] to model joint synergies from existing movement sequences for solving the *IK* problem.

### Motion Retrieval, Combination and Retargetting

The ever growing availability of human motion data [Uni09] has led to new research questions. Although such research directions are out of the scope of this thesis, we give some hints on research trends in extending the possibilities of motion capture databases, such as motion retrieval, motion combination and motion retargetting to cite a few.

Motion capture databases usually collect thousands of motion clips, creating challenges to store, access and process the data present in the database. Motion retrieval techniques usually work on two sub-problems. On the one hand it involves the identification of an adequate representation (indexing) of motion data so that high-dimension motion clips can be characterized in a low-dimension space. Secondly, reliable similarity measures have to be determined so that the comparison (retrieval) of motion chunks can be possible.

Regarding motion representations, geometric features have been proposed describing the dependencies between joints motion clips [MRC05, Lin06]. Another

representation is to involve more or less automatic methods for indexing motion clips with annotations [AFO03, BSP<sup>+</sup>04, ACD<sup>+</sup>09], thus characterizing differences in the semantic nature of a motion database. Other reduction methods represent motion in low-dimension spaces. Such techniques use for instance spectral transformations [AFS93], linear reductions methods such as PCA [FF05] or non-linear ones such as Isomap [XZ07]. Clustering can also be involved in such data reduction methods [LZWP03].

Among simple similarity measures, works have involved measures that cope with velocity and acceleration differences, and suggest to use rotation-invariant metrics on motion clips of the same length [KGP02]. Another method is Dynamic Time Warping (*DTW*) which gives generally good results [DTS<sup>+</sup>08] for finding an optimal alignment between motions. Aligning two motion sequences of different time lengths in *DTW* has been proposed in [KPZ<sup>+</sup>04].

Highly related to motion retrieval are motion combination techniques, which aim at creating new animations from an available set of motion chunks. Multivariate interpolation has been introduced [RCB98] through a verb (motion) /adverb (interpolation) metaphor. It typically uses a verb graph linking therefore motion clips with each other, in which allowed transitions can also be computed automatically [KG03]. Other techniques use motion graphs for modeling envisageable transitions between motion clips [KGP02], in which annotations [AFO03] and path planning [MK08] can be integrated.

Motion retargetting deals with the problem of adapting a given motion to other conditions or styles. Retargetting has been involved for adapting motion to different skeleton sizes [Gle98] as well as kinematic constraints [BW95]. Concerning style transformation, it has been observed that various frequency filtering bands can produce stylistic conditions. Emotional transformations between motion data have first been proposed in [ABC96] based on signal processing methods. Latter works have for instance combined style translations with inverse kinematics [GMHP04] and time warping [HPP05, HCGM06].

### 2.1.2.1.3 Conclusion

Kinematics-based animation techniques are shown to be well-suited methods when a high-level control over the task specification is needed. They give a formal framework for managing several tasks at the same time (such as secondary goals with Jacobian-based techniques) or for combining motion chunks (motion combination and retargetting).

As for inverse kinematics methods, an extensive evaluation of some of the previously cited techniques can be found in [UPBS08]. It shows namely that neither the Jacobian transpose nor CCD formulations give satisfactory results in

terms of motion quality, compared to other Jacobian-based methods.

Concerning learning techniques, one critical issue is that new motion are generally obtained by interpolating the learnt parameters, while no guaranty is provided to ensure that the resulting new movement sequences still respect the physical laws of motion.

In addition, the main drawback of methods dealing with motion retrieval and combination is that there is for the moment no other solution in this avenue apart from capturing all possible human gestures to build an extensive motion capture database that could be used for creating any animation sequence.

Despite the high realism in the resulting animations obtained by the previously detailed methods, their main drawback lies in the fact that they do not deal with dynamics. Therefore, they cannot take into account the possible physical interactions of virtual characters with their environment. It is true both for simply modeling physical interactions (for instance the gravitation field or collisions) as well as reusing motion capture in different contexts (even if some hybrid methods are available, see section 2.1.2.3).

An alternative to this problem is to take advantage of the physical representation of the virtual character, so that environmental interactions can be explicitly modelled.

### 2.1.2.2 Physics-based Methods

Animation techniques based on the physical representation of virtual humans are attractive since the synthesized motion implicitly respects the physical motion laws. The virtual character is put into motion by the application of forces ( $\mathbf{F}^D$ ) and torques ( $\boldsymbol{\tau}^D$ ) so that the physical representation reaches a desired configuration  $\mathbf{q}^D$ . This scheme is called the *forward dynamics* formulation, EQU. (2.9).

$$\mathbf{q}^D = \mathcal{D}(\mathbf{F}^D, \boldsymbol{\tau}^D) \quad (2.9)$$

Although there are several formulations of the motion laws (see the Lagrangian method for example [Bar96, Noc01]), they are strictly equivalent to Newton's motion laws, which relate the application of forces and torques, EQU. (2.10), to the linear ( $\boldsymbol{\Gamma}$ ) and angular ( $\boldsymbol{\Omega}$ ) state of a solid characterized by a mass  $m$  and an inertia tensor  $I$ .

$$\begin{aligned} \mathbf{F} &= m \cdot \boldsymbol{\Gamma} \\ \boldsymbol{\tau} &= I \cdot \dot{\boldsymbol{\Omega}} + \boldsymbol{\Omega} \cdot I \cdot \boldsymbol{\Omega} \end{aligned} \quad (2.10)$$



### 2.1.2.2.1 Forward Dynamics

Forward dynamics implies the explicit application of time-varying forces and torques to any solid composing the physical representation  $\mathbf{T}^D$  of the virtual character. The update of the physical configuration  $\mathbf{q}^D$  uses an integration of solids' linear acceleration and angular velocity over time. Such a framework therefore automatically handles external influences, such as the gravitation field as well as the collisions between solids. A comprehensive study of forward dynamics basics can be found in [Wil91].

The extension from the simulation of simple objects to fully articulated skeletons is far from straightforward. Early works involved the simulation of motion laws with numerical methods of high complexity. For example, the contribution from [WB85] proposed a solution with a  $O(n^3)$  complexity for an articulated chain composed of  $n$  DoFs. A recursive formulation associated to a tree structure for representing articulated bodies has then been proposed for treating such problem with a complexity reduction to  $O(n)$  [AG85].

These early works involved the motion simulation of rigid bodies by the independent specification of forces at the joint level, which is admitted to be far too simple, since it does not take into account the integration and simulation of the internal interactions between connected bodies [Wil86]. Classically, an alternative is to add springed-damped effects (or equivalent) to joints for circumventing this difficulty (see section 2.1.2.2.2), even if additional numerical instabilities then have to be counteracted [Gir91].

Despite the availability of accurate mechanical studies of human motion, such as human locomotion [CHt76, McM84], these contributions generally cannot be straightforwardly integrated into simulation frameworks. Analogously to forward kinematics, the main critical issue of forward dynamics is once again to rely on the animator's intuition for finding the adequate forces and torques to put the physical model into motion.

### 2.1.2.2.2 Inverse Dynamics

An alternative is to consider the inverse of the forward dynamics formulation, the so-called *inverse dynamics* (or *ID*) problem. It aims at automatically computing the needed forces and torques  $(\mathbf{F}^D, \boldsymbol{\tau}^D)$  to be applied on rigid bodies or mechanical joints composing the physical model, so that it reaches the desired configuration  $\mathbf{q}^D$ , EQU. (2.11).

$$(\mathbf{F}^D, \boldsymbol{\tau}^D) = \mathcal{D}^{-1}(\mathbf{q}^D) \tag{2.11}$$

Once the dynamic forces and torques are computed, a forward dynamics scheme is nevertheless used to put the physical model into motion. Classically, inverse dynamics methods are highly inspired from robotics, and imply either the respect of constraints along the simulation of motion laws, or the design of specific controllers. At the human body level, controllers can be considered as an *internal* representation of the muscle activity that occurs between the rigid bodies composing the virtual character. Whereas constraint-based methods can be qualified as an *external* formulation since they classically involve a global optimization scheme where no muscle model is provided.

#### 2.1.2.2.2.1 Constraint-based Methods

Early works have focused on the respect of geometric or kinematic constraints for finding the forces and torques to apply on articulated rigid bodies. Geometrical constraints (for example, a point belonging to an object fixed to a curve or to another object) are typically expressed linearly depending on forces and torques to be found [BB88]. It leads to the minimization of constraint equations over time. The translation and integration of these geometrical constraints into the minimization system can however conduct to an over-determined system that is difficult to solve. Another solution [IC87, Arn88, DAH90] is to compute the forces and torques given a set of kinematic constraints (such as desired linear and angular displacements) that are inserted into a formulation based on the Virtual Works principle of d'Alembert. Among these early works, one of the most complete integrations of kinematics constraints into physical simulation is certainly the work presented in [WK88]. It is based on the Lagrangian formulation of physics motion laws, in which initial geometrical and physical inputs are given. It then automatically computes the required forces to respect the initial kinematic constraints, and translates the system as a minimization problem. Such an optimization problem often considers the minimization of an energy function depending on the torques  $\tau^D$ .

Most of contributions have then focused on such optimization process, and specifically on its two main drawbacks: the control offered to users, and the computational cost of this approach.

Concerning the first drawback, enhancing the user control has first been addressed in [Coh92], giving the possibility to the user to interact with the iterative optimizations so that the convergence to an acceptable solution can be controlled. Users can also focus on time intervals, giving a subtle control over the overall simulation. The optimization process can also be controlled if desired transitions are required [RGBC96]. Further researchers have addressed the question of defining multi-objective functions that can cope with task priorities in the optimization process [AdSP07, JYL09].

Reducing the computational cost of such optimization-based methods has parallelly been addressed, mainly by considering low-dimensional representations of virtual characters [PSE<sup>+</sup>00], the motion itself [SHP04] or the underlying motion equations [BP08]. Simplified physical constraints also make possible the preservation of dynamic effects to a reduced cost, for example through the enforcement of momentum patterns [LP02]. Cost functions that can be optimized in linear time have also been under study [FP03] for decreasing this computational cost, highlighting a wide range of possible physical constraints such as bar and ground contact as well as flight phases.

Despite all these efforts, optimization-based methods are admitted to still have a high computational cost that moves away from real-time. The improvements detailed above for lowering the computational cost of such methods can also have an impact on the physical realism of the result, as attested in [vWvBE<sup>+</sup>09].

#### 2.1.2.2.2 Controller-based Methods

Contrary to constraint-based methods, controller-based methods do explicitly model the muscle activity between the rigid bodies composing the virtual character. We can therefore consider such controllers as an integral part of the dynamic model of the virtual character. Such contributions are motivated by early works on the the Hill-type muscle model [ZW90], whose spring-like effect is considered to be a key component of human postural stability:

" It turns out that for stable equilibrium the neuro-musculoskeletal system must possess the attributes of springs, and that for stability, these springs must exceed a certain critical stiffness. The passive, relaxed person is inherently unstable at many levels. (...) Certainly the central nervous system is partly responsible for this behaviour, but the primary mode of its influence can be considered to be adjusting muscle 'spring-like' behaviour. " [AW90].

An explicit modeling of the Hill muscle model has been presented in [LT06] for a complete simulation of neck dynamics, however such an option is currently banned for real-time animation since it necessitates a highly time-consuming pre-process (neural networks) for training the controller.

Most of works on controller-based methods have therefore focused on simplified formulations of the Hill muscle. It has led to the design of controllers for specific motor tasks, involving for instance two types of spring-like muscle models: the Proportional-Derivative (*PD*) and the Agonist-Antagonist (*AA*) formulations. Researchers have also tackled the issues of automatically adapting, tuning as well as composing such controllers.

### Proportional-Derivative Control

Precursor works have studied the design and adaptation of robotics-inspired controllers to the control of virtual characters [Rai86, RH91]. In recent years, most of contributions have focused on formulating motor commands applied to each mechanical joint in terms of proportional-derivative controllers. For each joint's degree of freedom, a PD controller is attached to paired rigid bodies whose relative motion is constrained by a spring-damped effect.

A PD controller is therefore traditionally composed of two terms, a proportional term parameterized by a stiffness coefficient  $k_s$  and a derivative term parameterized by a damping coefficient  $k_d$ . The proportional term models the muscle tension effect whereas the derivative term acts as the muscle relaxation. Given the angular state of the mechanical joint  $(\Theta^D, \dot{\Theta}^D)$ , and given a desired (target) joint configuration  $(\Theta_t^D, \dot{\Theta}_t^D)$ , a control torque  $\tau^D$  is computed so that its application on the linked rigid bodies put them into motion towards the desired configuration, EQU. (2.12).

$$\tau^D = k_s \cdot (\Theta_t^D - \Theta^D) + k_d \cdot (\dot{\Theta}_t^D - \dot{\Theta}^D) \quad (2.12)$$

A substantial amount of contributions have focused on the design of PD controllers for specific motor tasks. It includes the PD control of walking, jumping [HR91b], stairs climbing [HR91a], as well as running, cycling, vaulting and balancing [HWBO95]. PD control has also been used for simulating boxing fights [ZH02], swimming motion [YLS04], as well as breast motion [ZCCD04, DZS08].

Usually, torques computed by the PD control formulation are applied with respect to the relative frame formed between the linked rigid bodies. It has been shown however that expressing and applying the torques  $\tau^D$  towards the world frame leads to a better simulation stability [WJM06].

### Agonist-Antagonist Control

Contrary to the PD controller with its unique spring model, the Agonist-Antagonist control formulation integrates two spring-like behaviors that are intended to model the spring-like behavior of two opposite agonist-antagonist muscles. AA control is therefore composed of two tension proportional terms (parameterized by the coefficients  $k_{sL}$  and  $k_{sH}$ ) and a relaxation derivative term (parameterized by the damping coefficient  $k_d$ ). Note that this formulation only needs the angular joint limits  $(\Theta_L^D, \Theta_H^D)$  and its current angular state  $(\Theta^D, \dot{\Theta}^D)$ ,

and that no desired joint configuration is integrated into this formulation, EQU. (2.13) [Nef05].

$$\tau^D = k_{sL} \cdot (\Theta_L^D - \Theta^D) + k_{sH} \cdot (\Theta_H^D - \Theta^D) - k_d \cdot \dot{\Theta}^D \quad (2.13)$$

In fact, this approach is based on the *Equilibrium Point Control (EPC)* principle from Motor Control theory. The EPC principle argues that motor laws are not "programmed" but appear from the dynamic properties of the considered system under control [Fel66]. Human motion is then produced by transitionning between equilibrium points that are inherent to the dynamics of the human body.

The application of such principle is consequently equivalent to assume the existence of an equilibrium point configuration  $\Theta_{eq}^D$  that sums all external forces  $F_{ext}$  acting on a joint to zero, according to EQU. (2.14):

$$0 = k_{sL} \cdot (\Theta_L^D - \Theta_{eq}^D) + k_{sH} \cdot (\Theta_H^D - \Theta_{eq}^D) + F_{ext} \quad (2.14)$$

The motion of every joint composing the virtual character is then controlled by moving equilibrium points over time accordingly to desired configurations. This formulation has been used for example in [NF02] for human posture control and in [Kry05, KP06] for grasp movement control, though with a slightly different formulation. The *AA* formulation is surely the most accurate approximation of the Hill-type muscle model. It is nevertheless dedicated to the control of posture and implies ad-hoc methods for taking into account posture transitions.

### Automatic Tuning and Composition

The main drawback of the controller-based approach lies in the manual process of finding adequate parameters (stiffness and damping coefficients) so that the virtual character achieves a desired motion. Due to the time-consuming trial-and-error process for determining these coefficients, and due to the proliferation of many controllers for various motor tasks, researchers have focused on the scalability of such controllers.

Automatic methods for determining the internal coefficients of the presented spring-like controllers have been examined. Authors suggested to use neuromotor models [YCP03] based on motion perturbations to automatically compute the coefficients. A heuristic rule has also been presented in [ZH02] for determining adequate coefficients according to perturbation responses. The most achieved work in such contributions is based on the expression of joint composite inertia terms [ACSF07] that leads to the automatic computation of controller coefficients for upper-body motion. These composite inertia terms are shown to be of great

importance to take into account the influence of parent-child joint perturbations, which is a reason of the difficulty to find appropriate coefficients according to authors.

The first contribution addressing the composition of controllers has been developed in [vdPFV90]. It involves the introduction of controllers whose working area can be determined in a state space. Several state-space controllers can be concatenated, where the final state of a controller is the start space of another one. Such strategy for combining controllers depending on their final-start states has been successfully applied to locomotion controllers [WH97b, Woo98]. Different controllers can then be involved for controlling various phases in the same motion [LvdPF00, YLS04], namely by building finite-state machines. Finite-state machines have also been addressed in [FvdPT01], where the determination of pre and post-conditions defining possible transitions between controllers is automatically computed by a training/classification approach. The direct transition between controllers has also involved feedback-error learning policies for learning torque models [YLvdP07]. More recently, the optimization of controllers as regards to the states of a finite-state machine has been addressed [WFH09], mainly by invoking biomechanical properties of human locomotion. Another recent contribution uses state exploration and reinforcement learning to synthesize task-based control schemes that are shown to be more robust than the elementary controllers they are made of [CBvdP09].

#### **2.1.2.2.3 Conclusion**

Physics-based animation techniques show more and more compelling results for modeling the biomechanical causes that are at the origin of human motion. For instance, recent developments have shown promising results for modeling complex human systems such as a complete musculotendon model of the hand [SKP08].

The main difficulty for modeling and animating virtual characters by physical models remains however to find suitable kinematic postures to feed these models. For example, in EQU. (2.12) and (2.14), a knowledge of the targetted kinematic posture is needed. This problem is in a large extent inherent to both constraint-based and controller-based methods. Generally, this issue is addressed by hybrid methods based on available motion data, in an analogous manner to kinematics methods based on motion capture data.

#### **2.1.2.3 Hybrid Methods**

Despite many advances for giving more and more control over the chosen virtual character representation, animation techniques are still confronted to several

problems. On the one hand, kinematics-based techniques struggle with the lack of interactions of the virtual character with its environment. On the other hand, physics-based methods results show quite robotic motions and still need the specification of kinematic postures.

Kinematics and physics-based methods therefore tend to converge towards each other, firstly by the intensive use of motion capture data. Secondly, each type of method borrows tools from its counterpart, leading to what we call here *hybrid* methods.

#### 2.1.2.3.1 Kinematics, Kinetics and Dynamics

The use of dynamic features has appeared to be a promising compromise to overcome the interaction drawbacks of kinematics-based methods. The center of mass of the virtual character has been involved in early works [GM85] for balancing the character over its support polygon<sup>1</sup>. More formalized is the approach presented in [PB91a, PB91b], where the center of mass is considered as an end-effector to be controlled by an inverse kinematics approach. Works further developed in [BMT95, BMT96] improved this approach through a Jacobian-based method for controlling the center of mass (and balance) of the virtual character while respecting its mass distribution, a method which is referred to *inverse kinetics*. Such an inverse kinetics scheme has also been integrated along a CCD approach [KM05]. Expressive balance strategies have also been addressed in [NF04]. The inverse kinematics scheme can also be enhanced by momentum-based features, namely for controlling a kinematic virtual character and modifying pre-captured motion under push disturbances [KHL05].

#### 2.1.2.3.2 Motion-driven Physics-based Methods

Among physics-based methods that take advantage of motion capture data, constraint-based techniques have focused mainly on the modification of captured motion clips under physical motion laws. As for controller-based methods, most of works address the physics tracking of motion data.

##### 2.1.2.3.2.1 Constraint-based Motion Modification

Since early works in modifying motion capture data [Gle98], there has been a substantial interest in modifying motion capture data according to physics motion laws. This has been addressed first in [PW99] by considering a simplified physical model. As stated by authors, the result is nevertheless *not* physically realistic

---

<sup>1</sup>The support polygon represents the convex hull of feet positions. A balance criterium is then proposed to ensure that the projection of the center of mass onto the ground is inside the support polygon.

since no physics computations are processed on the full character model. Learning techniques have been therefore involved for automatically inferring the physical parameters of the model [LHP05]. The knowledge of these parameters leads to characterize the "style" of available motion capture data, and can produce other motions yet with the same captured style and a high computational cost. More interactive techniques then have been proposed for retiming motion capture data according to physical constraints [MPS06], as well as for editing and retiming acrobatics motion under momentum-based physical laws [MF07].

#### **2.1.2.3.2 Controller-based Motion Tracking**

Early works among hybrid tracking methods involved the physics control of figures via inverse dynamics from procedurally-generated kinematic data [KB96]. Then, the availability of motion capture data helped controller-based methods to specify the appropriate kinematic postures to be physically simulated. Most of early work among these techniques have therefore focused on tracking motion data [ZH99]. It includes first the design of controllers for specific motor tasks that can give responses to external disturbances, such as box punches [ZH02]. Such contributions mainly address the blending of kinematic postures (motion capture data) and simulated motion when needed [ZMCF05] for decreasing the computational cost of physics simulation. A more detailed organization of kinematic and dynamic controllers, as well as their transitions, has been proposed in [SPF03]. Motion retrieval techniques have also been investigated in such approaches so that the best suited kinematic posture can be selected, in order to drive the controllers at the right time [Man04, ZMM<sup>+</sup>07].

#### **2.1.2.3.3 Conclusion**

In conclusion, the recent development of hybrid methods attests of the convergence between kinematics and physics-based methods. This convergence implies mainly an intensive use of motion capture data as well as the mixing of techniques coming from both approaches.

Hybrid methods based on kinematic-based techniques succeed in elaborating control schemes that can take into account physics features such as the center of mass or the momentum of virtual characters. In this type of formalisms, it is however far from straightforward to include other interactions such as ground contacts or collisions . This confirms the limitations of providing a physical realism to the kinematics-based representation of virtual characters.

On the other side, the combination of physics-based techniques with motion capture data makes possible the modification of pre-recorded motion clips, as well as the tracking of motion capture data with controllers. In the former case,



this still induces a high computational cost that moves away from real-time. In the latter case, this does not offer an intuitive mode of control since the tracking is based on high-dimension kinematics postures (every joint positions and orientations).

### **2.1.3 Summary**

In this section, we have presented an overview of the modeling techniques of virtual characters, as well as the computer animation methods that are generally used for giving them life. This state of the art has reviewed kinematics-based techniques, physics-based techniques, as well as hybrid techniques taking advantage of the former two.

Regarding our goal to finely model the mechanisms that are at stake during percussion performances, it seems that physics-based techniques are the most adequate choice, since they offer a well-established formalism to physically control the motion of a virtual performer. Moreover, such techniques offer the opportunity to physically model the coupling between the virtual performer and a virtual instrument.

We will mainly focus on methods for motion tracking using controller-based techniques (section 2.1.2.3.2.2), since they combine our real-time requirements and the realism benefit of motion data. Two main drawbacks of such technique remain however, firstly the manual tuning of controllers' parameters, and secondly the tracking of kinematic postures of high dimension. While this thesis will not address the first drawback, one of our contributions is to cope with the second one by using a cascaded combination of inverse kinematics and inverse dynamics controllers.

## 2.2 Computer Music

Since its early beginnings, the main goal of *Computer Music* has been to develop a novel musical medium based on computer capabilities. It involves people coming from various research areas, such as performers, composers, scientists and engineers. As stated below, computers offer much more possibilities compared to acoustical instruments:

" Man's music has always been acoustically limited by the instruments on which he plays. These are mechanisms which have physical restrictions. We have made sound and music directly from numbers, surmounting conventional limitations of instruments. Thus the musical universe is now circumscribed only by man's perceptions and creativity. " [MPG62].

Two trends have then emerged from these early works, either considering computers as new instruments for synthesizing sounds, or as an help to music composition. In the first case, researchers focus on the modeling, analysis and synthesis of sound inner properties (section 2.2.1). In the latter case, it involves paradigms for controlling either natural or synthetic sounds for compositional purposes (section 2.2.2).

### 2.2.1 Sound Synthesis Models

Many sound synthesis models have been proposed since the advent of *Computer Music*. They can mainly be divided into two sub-classes, the models that *describe* sound properties (subsection 2.2.1.1), and those giving a physical interpretation of the *source* of sound properties (subsection 2.2.1.2).

#### 2.2.1.1 Descriptive Sound Synthesis

Among descriptive models, we will present here two of the main widespread models: *sampling* and *spectral* models, as a means of underlining the differences compared to physics-based models. For a more complete review about descriptive models, interested readers are referred for example to [Pol91, TVK98].

##### 2.2.1.1.1 Sampling Models

One of the most used synthesis models nowadays is sampling synthesis. Sampling synthesis is based on the playback of an extensive database of real pre-recorded sounds. Sampling necessitates therefore the recording of instrument sounds under a huge variety of playing conditions. This model has been used even before the advent of computer technologies, such as in the *Studio de Musique*

*Concrète* founded in the 1950's by P. Schaeffer. Since the launch of the *Fairlight CMI* [Cra09], most of current commercial synthesizers use sampling synthesis.

The main advantages of sampling synthesis are its straightforward implementation and its sounding fidelity since it is based on real sounds. But these advantages are however its main drawbacks. It necessitates a time-consuming process of recording sounds coming from an instrument under many playing variations. Moreover, the manipulation of sound samples requires a large amount of computer memory for storing all these pre-recorded sounds. An additional limitation is the uniqueness of the sound samples as regards to the recording conditions that prove their lack of flexibility towards environment changes.

Researchers have therefore developed techniques for decreasing the memory cost of such synthesis technique. This reduction can be significantly achieved by simply *looping* sound samples (or just one period of it), under the assumption that the tone of the considered instrument stays constant in amplitude and pitch during the steady-state part of the sound [Roa96].

The size of the sound sample database can also be decreased by collecting fewer sound samples than it would be needed, and process some *interpolation* between these key sound samples. For example, as stated in [Smi05], playing back a piano sound with a given key stroke velocity can be done by interpolating "the two sounds recorded at the nearest lower and higher key velocities".

The recording of all the possible tones of an instrument can also be sidestepped by involving digital audio effects methods, so that the pitch of an original sound can be transformed to another desired pitch. This can be done by *ad-hoc* methods (such as modifying the digital-to-analog clock frequency [Roa96]) or by more formalized transformations such as delay filters [LVKL96].

More generally, the memory cost for storing the sound sample database can be reduced by applying data reduction methods. This usually involves data compression techniques (such as downsampling sound samples) which may alter the quality of sound samples. Perception-based criteria can also be used for preventing the degradation of sound samples based on human perception properties [Roa96].

#### 2.2.1.1.2 Spectral Models

Contrary to sampling synthesis, spectral models represent the properties of sound waves in the spectral domain. One of the most used and secular methods in spectral synthesis is the *additive* model. This model considers a sound wave as the sum of sinusoidal components, just as pipe organs route air through a selected subset of pipes. This model has been used in early electronic music, such as the *Telharmonium* of Cahill [Wei95].

The additive synthesis model represents therefore the sound output  $S(t)$  as the sum of the effect of oscillators characterized by time-varying amplitudes  $A_i(t)$ , frequencies  $f_i(t)$  and phases  $\phi_i(t)$ , EQU. (2.15). The oscillator characteristics define the control functions of the additive synthesis scheme.

$$S(t) = \sum_{i=0}^N a_i(t) \cdot \sin(2\pi \cdot f_i(t) + \phi_i(t)) \quad (2.15)$$

These control features can be arbitrary fixed based on compositional matters, or can be obtained from the analysis of a real pre-recorded sound. Practically, the analysis stage involves a Fourier analysis, via the *Short-Time Fourier* transform for instance. A peak detection is usually processed on the Fourier spectrum for obtaining the amplitudes and frequency functions. As for phase functions, most of formulations do not take them into account in the model since its importance in the additive synthesis model depends highly on the context [Roa96]. Different variations of this spectral analysis scheme have also been proposed, such as the pitch-synchronous analysis [RM69] or the phase vocoder [FG66, Dol86].

Similarly to the sampling synthesis method, the main drawback of additive synthesis is its requirement of a large number of oscillators. Data reduction methods have been proposed to decrease the number of oscillators, such as the automatic approximation of amplitude-frequency functions by line segments [RM69, Str80], principal component analysis [Lau89], spectral interpolation [SDD90] or spectral modeling synthesis [SS90].

### 2.2.1.1.3 Conclusion

Descriptive methods involve mainly either time-domain or spectral-domain techniques, coupled with signal-processing methods for taking into account various sound phenomena.

Compared to sampling synthesis, spectral synthesis methods are more convenient as they provide a more accurate representation of sound properties regarding human perception characteristics [Moo03]. It also provides a reduced representation cost since, firstly time-domain signals are replaced by their spectral representations, and secondly human auditory limitations can be invoked such as the human ear frequency range and simultaneous-temporal masking [Pla05].

However, sampling and spectral models imply the use of a large amount of data (either sound samples or frequency-amplitudes functions) to finely represent sounds. It is consequently difficult to control such models for obtaining the desired sound effects when using such models.

### 2.2.1.2 Physics-based Sound Synthesis

Conversely, physical models of sound properties are generally considered to offer meaningful and intuitive control parameters over the mechanisms responsible of the sound production [RDFL94].

The first contribution in this area has been provided in [KL62] for modeling human vocal strings<sup>2</sup>. After roughly ten years of scarcity, physical models have been explored further, either focusing on digital solutions of the wave equation governing the acoustical mechanisms occurring in various instruments, or physics models of sounding objects themselves. In this section we derive these different models for highlighting the advantages and limitations of each solution. As an illustration, we will consider in each case the classical problem of the vibrating string.

#### 2.2.1.2.1 Digital Solution of the Wave Equation

Many methods have been proposed to provide a digital solution of the continuous formulation of the wave equation. They are usually divided into two categories, those providing a "*brute-force*" solution of the wave equation, and those based on a bank of delay lines called *digital waveguide*.

##### 2.2.1.2.1.1 Preliminaries on the Wave Equation

Before entering in the details of each model, we present the background inherited from early works about the problem of the vibrating string. This includes either solving a partial differential equation (*PDE*) or providing a modal decomposition of the string.

The vibration phenomenon of a string is classically expressed by the d'Alembert equation under adequate assumptions, EQU. (2.16) (see appendix A.1 for a full derivation). The vertical displacement  $z$  of the string depends on its longitudinal displacement  $x$  over time knowing the mass density  $\mu$  of the string and the applied tension  $T$ .

$$\frac{\partial^2 z}{\partial x^2}(x, t) - \frac{1}{c^2} \cdot \frac{\partial^2 z}{\partial t^2}(x, t) = 0 \text{ with } c^2 = \sqrt{\frac{T}{\mu}} \quad (2.16)$$

d'Alembert's solution to this equation is composed of two terms characterizing two travelling waves in opposite directions, EQU. (2.17). A more formalized explanation of this phenomenon is provided in appendix A.2 based on Fourier's solution.

---

<sup>2</sup>This first attempt turned rapidly into one of the most heard songs, since A. Clarke working with S. Kubrick on *2001: a Space Odyssey* introduced the song "A bicycle built for two" composed with this model [Woo91]

$$z(x, t) = z^+(ct - x) + z^-(ct + x) \quad (2.17)$$

The resulting wave of the string vibration is composed of two sinusoidal waves that propagate rigidly through the string with a velocity of  $c$ . The solution proposed by Fourier is more general, since it also introduces the representation of the travelling waves by a sum of *normal modes* that characterize the allowed deformations of the string.

An alternative to the PDE-based model is to consider the string as composed of a combination of mechanical elementary masses linked by springs and dampers. A justification of the use of such model is provided in appendix A.3, showing that a modal decomposition can analogously lead to the d'Alembert equation.

PDE-based and modal models are intrinsically different by nature since they provide either an *analytical continuous* or a *discretized* solution to the vibrating string problem. To a more general extent, researchers have focused on these continuous and discretized formulations for a wide range of instruments. In the case of the analytical formulation, it is however not always possible to provide a solution to the problem. Continuous problems have therefore also been studied through the discretization of the wave equation governing the considered systems.

#### 2.2.1.2.1.2 "Brute-force" Models

Two main methods are involved in brute-force models: the finite-difference and finite-element schemes. These two formulations proceed to a discretization of the wave equation both in space and time. A good introduction to the numerical analysis of partial derivative equations can be found in [Sai06]. Probably due to the required mathematical background, finite-element methods are most of the time eluded in the literature, so that we provide a complete derivation of these models applied to the simple vibrating string problem in appendix A.4.

The finite-difference scheme has been used since the beginning of physical models. An application to the vibrating string has been firstly proposed based on this scheme [Rui69, HR71]. The finite-difference formulation has also been used for modeling the acoustic response of other musical instruments with more complex (dissipative, dispersion) phenomena, such as xylophone bars [Bor95, CD97, DMC98] and drum membranes [Bou06].

The finite-element formulation has also been used in the simulation of many musical instruments, such as drum membranes [OB91] and bars [BCT<sup>+</sup>99, BSM99]. The finite-element method has also been integrated along with rigid-bodies simulation in [OCE01]. The *Modalys* framework [Mod09] from *IRCAM* [Ben03,

EBC05] makes also available such formulation.

The main advantage of the finite-difference scheme is its straightforward treatment of the discretized acoustical problem. However, its drawbacks are firstly in terms of simulation, since no control of the space-time error during the simulation is available, and secondly in terms of discretization, since such method cannot be applied to all geometries. On the contrary, from appendix A.4, finite-element methods provide a formal framework for controlling the error made both along space and time (based on a norm on a vectorial space), and its general discretization formulation (possibly non equally distributed) can be applied to all geometries.

However, both methods are admitted to have a high computational cost, preventing their use in real-time simulations, even if some contributions have explored a cost reduction. In the case of finite-difference, Field Programmable Gate Arrays (*FPGAs*) have for instance been used for parallelizing the numerical solution of the wave equation [MBW05].

### 2.2.1.2.1.3 Digital Waveguide

Another alternative based on partial derivative equations is the digital waveguide formulation. Instead of discretizing the PDE that governs the acoustical response of a musical instrument, it directly proceeds to the discretization of the *solution* of the considered PDE.

The first contribution in this area is the work from Karplus and Strong [SK83]. At the beginning, this work was more an extension of the descriptive *sampling* synthesis method than a physically-motivated solution. The aim was to get rid of the need of a huge amount of pre-recorded sound samples. Karplus and Strong proposed therefore an *ad-hoc* wavetable scheme based on delay lines initialized by random numbers rather than real sound samples (see appendix A.5 for the application to such a scheme to the vibration string problem).

A physical interpretation of the abstract solution of Karplus and Strong has then been proposed in [Smi87, Smi92]. Applied to the vibrating string problem, it involves the discretization of EQU. (2.17) given a time sampling period  $T_s$  and a space sampling step  $X_s$  such that  $X_s = c.T_s$ . The discretization leads to:

$$\begin{aligned} z(x, t) &\mapsto z(p.X_s, n.T_s) = z^+(n.c.T_s - p.X_s) + z^-(n.c.T_s + p.X_s) \\ z(x, t) &\mapsto z(p.X_s, n.T_s) = z^+((n - p).c.T_s) + z^-((n + p).c.T_s) \\ z(x, t) &\mapsto z(p, n) = z^+(n - p) + z^-(n + p) \end{aligned} \tag{2.18}$$

By removing the sampling constants, the discretization leads to EQU. (2.18). The discretized problem of the vibrating string described by EQU. (2.18) includes then the superposition of the effect of two nested delay lines corresponding to the two travelling waves in opposite directions, as explained in [Smi98] (see appendix A.5).

Further works on the digital waveguide approach have then explored improvements for taking into account other physical phenomena and instruments of larger (2D or 3D) dimensions. These problems are generally approached through the design of a system composed of delay lines excited by digital filters, as proposed by early works [MSW83], in a similar way to the design of *modifiers* in the work of Karplus and Strong (see appendix A.5). Digital waveguide methods use signal processing techniques for taking into account real physical effects. An open-source framework called the *Synthesis ToolKit (STK)* is available for sharing such contributions [CS09, CS99].

Regarding physical phenomena, the accounting of loss phenomena, such as amplitude attenuation, have for example been addressed in [VHKJ96, BRS97, VT98]. Dispersion effects, such as frequency-dependent delays, have been presented for instance in [RS96, RS99].

As stated in [Ban06], adding such digital filters may have dramatical effects on the simulation result, so that digital control filters have to be integrated for ensuring that the result belongs to an adequate frequency range. This includes for example the introduction of simple solutions such as first-order allpass filters [JS83], as well as more complex mechanisms such as fractional-delay filters [LVKL96].

Another difficulty arises from the complexity of nested digital filters that is the estimation of internal parameters of each digital filter included in a waveguide scheme [Smi83]. This parameter estimation generally involves the analysis of real pre-recorded instrument sounds [KEA<sup>+</sup>02].

The one-dimensional digital waveguide approach has been used for a wide range of musical instruments [Smi05]. For modeling instruments of higher dimensions than a string, researchers have studied the combination of one-dimensional digital waveguides. This includes the introduction of 2D [vDS93] and 3D [vDS96] waveguide meshes.

Digital waveguide models have also involved the modeling of the instrument bodies. It includes either post-processing digital filters for rendering the reverberation of sounds in body cavities [Gar87] or commuted synthesis techniques [KV93].



### 2.2.1.2.2 Modal and Mass-Spring Models

The main difference between modal or mass-spring models compared to the models based on the wave equation detailed in the previous section is the implicit or explicit spatial discretization of the acoustical object under study. To that mean, the models based on the wave equation proceed to an implicit spatial discretization since this discretization is invoked for discretizing the PDE both in space and time.

Conversely, modal and spring-mass models do not make any assumption of the PDE governing the acoustical system, and directly discretize the sounding object spatially, and then also in time for solving the mechanical problem. These two methods differ from the elementary objects they involve in their formulation, either  $2^{nd}$  order oscillators or masses connected with springs.

#### 2.2.1.2.2.1 Modal Synthesis

The modal formulation expresses the discretization of an acoustical system as a bank of  $2^{nd}$  order oscillators, each representing a mode of vibration of the sounding object. The modal formulation has been first introduced to sound synthesis in [Adr89, Adr91], these works are typically the fundamental basis of the *Modalys* framework [Mod09]. In this paragraph we will focus on a system represented by a network of  $N$  undamped oscillators, which can be characterized by a matricial equation EQU. (2.19) containing the governing equation of each oscillator.

$$\mathbf{M}.\ddot{\mathbf{z}}(t) + \mathbf{K}.\mathbf{z}(t) = 0 \quad (2.19)$$

The matrices  $\mathbf{M}$  and  $\mathbf{K}$  are referred to the mass and stiffness matrices. These are usually not diagonal, so that the system is characterized by  $N$  coupled oscillators. By introducing the notion of *modal shapes*  $\mathbf{S}$  (see appendix A.6), this system can be rewritten in the form of a system of  $N$  *uncoupled* oscillators representing the normal vibration modes  $\mathbf{q}$  of the system.  $\mathbf{q}$  represents then an orthogonal basis of the system, such that  $\mathbf{z}$  can be expressed as a linear composition of  $\mathbf{q}$ , EQU. (2.20).

$$\forall j \in [1, N] \quad z_j(t) = \sum_{i=0}^N s_{i,j} \cdot q_i(t) \Leftrightarrow \mathbf{z} = \mathbf{S}.\mathbf{q} \quad (2.20)$$

The interest of the modal formulation is that it does not make any assumption on the considered system. Such a formulation can moreover be shown to be equivalent to a continuous representation based on PDE. This affirmation is justified in appendix A.6. The modal method has inspired many works since its introduction to sound synthesis, mainly because one of the most important advantages of such a formulation is its compatibility with an analysis stage. This

last one may be used to determine the adequate basis (*normal modes*) necessary for describing a particular acoustical system. Analytical techniques have been developed for determining such features depending of the shapes of sounding objects [kvdDP96, kvdDP98]. These can also be obtained from measurements on real objects [PvdDJ<sup>+</sup>01, CvdDLH07].

#### 2.2.1.2.2.2 Mass-Spring Networks

An equivalent approach, the CORDIS-ANIMA framework [Cad79, Luc85, Raz86, CA09], is based on an analogous mechanical formulation, modeling any musical instrument as a network of mass-spring elements [FC91, CLF93]. This system has the advantage to treat the evolution of any sounding object at the sound sampling frequency (about 44 *kHz*). Such method is generally admitted to have a critical performance cost, although real-time simulation can be reached and combined to specific hardwares in particular cases [LFCC07]. CORDIS-ANIMA has particularly been the basis in a musical composition perspective, for example through the introduction of Genesis<sup>3</sup> [CC09].

#### 2.2.1.2.3 Conclusion

Physics-based methods focus either on the PDEs (or their solutions) governing the considered acoustical system, or on physics-based alternative which consist in providing a discretization of sounding objects into elementary structures (oscillators or mass-spring systems).

Hybrid methods have also been explored for combining the assets of both PDE-based methods with modal or mass-spring representations. Some works have for instance mixed a finite-element method to a modal synthesis formulation [OSG02], or a combination of modal and mass-spring representations [RL06] for generating sounds from rigid-body simulations. Acoustical features such as radiation factors [JBP06] as well as frequency-masking can also be used [vdDKP04] for enhancing the sound realism in virtual scenes.

The choice of a physics-based method is characterized by a compromise between real-time requirements and the techniques involved in the models. Finite-difference and finite-element methods typically fall short in real-time difficulties. Digital waveguides are quite low cost-expensive. However it generally involves the introduction of *ad-hoc* filters for taking into account physical phenomena that cannot be solved by a general PDE. The modal formulation expresses any musical instrument (and more generally any sounding object) as a finite network of oscillators. This method is admitted to be powerful for the main reason that no underlying assumption is made in such a model, compared to physical models based on PDEs. Such a formulation is characterized by a high computational

cost, however in practice the normal modes are usually precomputed so that the physical sound synthesis can run in real-time. Other speed-up improvements have also been proposed based on mode compression and truncation [vdDKP04, RL06], as well as spectral representations [BDT<sup>+</sup>08].

## 2.2.2 Gestural Control of Sound Synthesis Processes

Another research area in *Computer Music* is to explore how the previously detailed sound synthesis techniques can be involved in a musical composition process. This research direction addresses the *control* of sound synthesis processes. Such a question has been present since the beginnings of *Computer Music*:

" Today, scientists and musicians (...) are trying to make the computer play and sing more surprisingly and mellifluously. As a musical instrument, the computer has unlimited potentialities for uttering sounds. (...) Wonderful things would come out of that box if only we knew how to evoke them. " [Pie65]

It attests the fact that research works have provided many sound synthesis formalisms competing more and more in accuracy, but also forgot on this way the corporal (gestural) strategies that are involved to give them birth in instrumental performances. The last decades have therefore focused on solutions for proposing paradigms in order to recover this corporality. Such research direction is generally addressed as a means of *controlling* sound processes, either by involving the design of devices for capturing instrumental gestures, or by proposing models of these latter.

In this development, it should be noted that we will refer to *gestures* as the actions that can be performed by a user. We will not address a complete bibliographic review of gesture typologies for characterizing the differences between effective, accompanying or figurative gestures in the case of a music performance. Interested readers looking for an extensive review on these typologies are referred to [Gib87, Cad94, Wan01, Jen07].

### 2.2.2.1 Acquisition, Mapping and Digital Instruments

Capturing gesture characteristics are an important part of the design of devices for controlling sound synthesis processes. These devices are generally referred as digital music instruments (*DMIs*), including the design of a controller capable of acquiring and making a correspondance between gesture characteristics and sound synthesis processes. The nature of these characteristics will therefore influence the choice of sensor technology that will be incorporated to the device.

Once the input device is built and embeds the chosen sensor technologies, another issue arises then for specifying how the numerical conversion of the identified gestural characteristics should act on the inputs of a sound synthesis process. This issue is commonly known as the *mapping* problem.

#### 2.2.2.1.1 Direct and Indirect Acquisition

There are basically two ways of capturing gestural characteristics for the design of DMIs, either direct or indirect solutions.

Direct acquisition methods make use of sensors for monitoring the actions performed by users on the device. These sensors can be of different nature and capabilities for capturing various features. An extensive review on the use of these sensors applied to music performances is available in [MW06, Mal07]. These sensors can acquire both the external (action on the device) and internal (performer state) activities of the performer. To cite a few, external sensors can be magnetic or optical systems [Vic09, Pol09] for capturing the motion of performer's body parts, or less intrusive mechanisms that can be embedded to the device such as force-sensitive resistors, potentiometers or tactile-position sensors. Sensors can be used to measure the internal activity of the performer as biosignals, namely by capturing electromyogram (*EMG*) data. *EMG* data have been used for example for capturing conducting gestures [Nak00].

On the other hand, indirect acquisition methods proceed to the identification and extraction of gesture characteristics from sounds of acoustic instruments. Signal processing techniques are generally the basis of these methods for extracting sound features that characterize gesture features such as the position of the device, the performed fingering or the strike location on a drum membrane [WD04]. Such techniques have among other instruments been applied to the acoustic guitar [TDW03] and clarinet [Ego95].

#### 2.2.2.1.2 Mapping

Mapping can be defined as the gesture-sound relationship, that is how numerical data acquired from sensors are related to the inputs of sound synthesis processes. Although a mapping with basic one-to-one relations between gesture outputs and sound inputs should be the most straightforward and evident solution, it is now admitted that complex mappings are more suited since acoustical instruments are characterized with a non-linear and complex scheme for relating gesture and sound [HWK00].

Two main strategies have been therefore proposed for solving this mapping issue. The first solution is to explicitly express the mapping rules based on *a priori* knowledge on acoustical systems [HK00, vNWD04]. Another alternative is to automatically generate the gesture-sound mapping, for example by training artificial neural networks [CCH04] or processing a data reduction for keeping the

most statistically meaningful features [BRC02]. Generative methods for achieving the mapping is however quite arbitrary since such learning or statistical techniques do not ensure that the resulting statistical discrimination corresponds to gestural differences.

The mapping between gestural data and sound synthesis processes is therefore of paramount importance when designing a DMI, it may influence in particular the way performers can interact with the device. The design of the mapping can also be seen as an attribute of the composition process, since the way the mapping is expressed can influence the use of the gestural controller [Doc02].

### 2.2.2.1.3 Digital Music Instruments

The design of DMIs has been explored much before the advent of *Computer Music*, by building electric and/or electronic interfaces for generating sound from novel control surfaces that could be used by performers. One of the most well-known early achievements in this domain is the Theremin invented in the 1920s by Lev Termen [the29]. With both the growing availability and democratization of personal computers and the widespread use of the MIDI protocol [MID09], research works dealing with the design of DMIs have increased drastically [Cha97]. It has especially led to the appearance of digital music orchestras for exploring the performance possibilities of DMIs [PDH<sup>+</sup>09].

An extensive classification of DMIs is available in [MW06] according to their relationship to existing traditional acoustical/electric instruments. In this development we focus specifically both on *augmented* and *alternate* controllers.

DMIs based on augmented instruments extend existing instruments by adding sensors such as the sensors discussed in subsection 2.2.2.1.1. They provide extra control parameters to performers to those traditionally available with existing instruments. Augmentation solutions have been explored for a wide range of musical instruments. Augmented flutes have been elaborated for mapping the breath stream [dSWS05] and the position/orientation of the device [PQ08] to sound synthesis processes. This principle has also been applied to electric guitars, by including for example a touchpad to a bass guitar [BT01] or contact sensors for enhancing an electric guitar with percussive playing techniques [Läh09].

Conversely, alternate gestural controllers do not specifically refer to an existing traditional instrument, but rather elaborate novel musical interfaces from scratch or directly based on other objects. Such interfaces are usually based on hardware for sensing the actions of performers on the DMI. Many solutions have been proposed for capturing these actions, such as computer-vision methods [JKGB05], pressure [Buc09a] or capacitive sensors [MW07], as well as tablet devices [ZS06, DD07, ZS08]. Other alternate controllers also integrate haptic feedback so that performers can perceive the result of their actions on the devices. This ranges

from simple vibro-tactile feedback [RH00] to the simulation of the "feel" of the device from haptic devices [CLF90, SW08].

#### 2.2.2.2 Gesture Analysis and Modeling

Conversely to works detailed in the previous section, research works have been trying not only to capture instrumental gestures, but also analyse and propose models of gesture mechanisms that can take into account all their subtle variations as regards to musical variations.

This includes the study of the influence of dynamics and tempo on bow gestures [RFB06, RBWB08]. Other works have also studied anticipation phenomena that occur in instrumental gestures under musical nuances and tempo changes, this has been done for instance for bow [RB08], percussion [Dah00, Dah04] and piano [LP07] gestures. Mechanisms showing the influence of instrumental gestures on the audio-visual perception of instrumental performances have also put in evidence [WVM<sup>+</sup>05, DF07].

Most of models of instrumental gestures propose parameters that can take into account the direct interaction with sound processes. Such works differ mainly from their positioning towards the sound model. Some are proposing models for exciting directly the parameters of sound synthesis schemes, whereas some propose more high-level models by offering methods for generating these parameters.

Models dealing with the direct excitation of sound synthesis parameters have been proposed for instance for controlling bowed strings [Dem08] by the specification of attack parameters, or for drum models [Ava01, AR04] by the proposition of various contact models. As for higher-level models, some works have proposed the use of existing motion data. This can take the form of the optimization of a dynamic system by the addition of effort costs [RB08]. Another solution is to propose a system able to "follow" and recognize motion data by the training of an hidden markov model [BGS<sup>+</sup>07, BZS<sup>+</sup>09]. Finally, other contributions have proposed physical models for the whole gesture-sound system [Gib87]. In this work, the playing of a musical instrument can be conceived through the coupling of two physical models, the sound-producing gesture and the instrument. The entire gesture model can be identified using a slowly evolving mechanical model with force feedback. This has led to the proposition of a composition system through an analysis/synthesis approach [GF88].

#### 2.2.2.3 Conclusion

We have examined the two main streams of research works for controlling sound synthesis processes.

On the one hand, the control of sound synthesis processes can be achieved by the design of gestural controllers, involving gesture acquisition methods and exploring mapping schemes for relating the captured gesture characteristics to sound synthesis inputs. The most critical issue in this approach is the mapping between the extracted gestural data and the sound synthesis inputs.

On the other hand, research works in the control of sound synthesis have involved the direct modeling of gestures, by extending the current knowledge on instrumental gestures of various types, as well as providing models of interaction gestures that are strongly related to sound synthesis parameters.

### **2.2.3 Summary**

We have drawn in this section an overview both concerning the main sound synthesis models and the strategies for controlling sound synthesis processes. The sound synthesis models involve either descriptive methods for modeling the observation of sound properties, whereas physics-based techniques model the physics phenomena at the origin of these sound properties. We have also reviewed control paradigms of sound synthesis processes that include either the design of digital music instruments or models of instrumental gestures.

Regarding our goal to finely model the interaction between a virtual percussionist and sound synthesis processes, it seems that physics-based sound synthesis models are an adequate solution. Specifically, the modal formulation appears as the best choice for our real-time and complexity requirements. Concerning the mapping problem, another reason in favor of physics-based sound synthesis models is the possible explicit mapping based on the physics events that are available from the simulation of the virtual percussionist and the physics parameters of the chosen sound synthesis scheme. It should however be noted that the major difference between such a mapping and the one discussed previously lies in the fact that virtual gesture characteristics will be here extracted from the simulation of the virtual percussionist, compared to a mapping using gesture characteristics coming from natural gestures. This thesis will thus present a solution of the mapping problem in the particular case of the synthesis of virtual performances.

Furthermore, our work comes under the category of the analysis and modeling of instrumental gestures rather than in the design of control devices of sound processes. Our major contribution is to propose a model whose design is guided by a preliminar analysis of real mechanisms occurring in percussion performances. Our model is also able to synthesize whole movements, including preparation, interaction and retraction gestures, compared to related works focusing only on

interaction gestures.



## 2.3 Virtual Music Performances

In this section, we finally examine the contributions of both *Computer Animation* and *Computer Music* research fields to the synthesis of virtual music performances. We define "virtual music performances" as the *modeling* of all protagonists acting during a *real* music performance, *i.e.* modeling techniques for representing the performer, the instrument and the interaction between the two.

### 2.3.1 Computer Animation Contributions

Early works have explored the synthesis of virtual scenes including sounding objects with a motion-driven methodology [TH92, HGLG95]. These contributions however do not model a music performer explicitly, but rather consider virtual objects whose kinematic motion influences the sound rendering. Similar works have explored motion-driven methods for modifying pre-recorded sound clips [CBBJR03] or physics-based animation techniques for animating sounding rigid bodies [OCE01, OSG02].

Other contributions aim at animating virtual models from sound representations. The most explored solution includes works for animating virtual models from MIDI scores [Lyt90, Lyt01, Lyt09]. Such a strategy has also been proposed along with an explicit model of a kinematic representation of a virtual performer [WG97]. Other works have proposed either the animation of kinematic virtual models directly from sound clips [CBB<sup>+</sup>02], or the animation of a kinematic hand model from guitar tablatures [ES03].

Most of the time, the contributions from the *Computer Animation* research area elude both the questions of the fidelity of the synthesized motion and the interaction between virtual models with sound synthesis processes. Specifically, the relationship between virtual objects (or virtual performer models) and real instrumental gestures is generally not addressed, so that there is no way to say whether or not the synthesized gestures are similar to real-world mechanisms.

Similarly, apart from the case of MIDI-controlled animated models, the interaction between virtual models with sound processes is rarely explicated. It should moreover be noticed that animated models from MIDI scores consider the inverse mechanism of real instrumental performances. They indeed animate virtual models from the event-based structure of MIDI scores, whereas real instrumental performances show obviously that sounds result from the performer actions on sounding objects.

### 2.3.2 Computer Music Contributions

The *Computer Music* community has been prolific for providing models that can be used in virtual music performances. and especially for virtual percussion performances. It includes either percussion sound synthesis models, percussion gestural controllers and virtual percussion systems.

#### 2.3.2.1 Percussion Sound Synthesis

Most of sound synthesis models have been explored and applied to the specific case of percussion performances. These include finite-difference schemes [Bou06] accelerated with Field Programmable Gate Arrays (*FPGA*) [Bil05b, Bil05a] and the specification of a meaningful and explicit mapping strategies [COW07]. 2D and 3D waveguides meshes have also been applied to percussion sounds [FR95, ALF00].

More related to virtual performances is the contribution presented in [AR01, AR04, AC06] since these works have been combined to offer a virtual multimodal application concerned with the simulation of contact sounds. It nevertheless does not include a model of a virtual performer, but rather relies on haptic devices for exciting contact sound models with a 3D visualization feedback.

#### 2.3.2.2 Percussion Gestural Controllers

Several gestural controllers applied to percussion have been developed, such as the Lightning and Thunder controllers [Buc09b, Buc09a], the Radio Baton [BMS89] and the Korg Wavedrum [Rul95]. An extensive historical review of such early works is available in [Aim06].

Percussion controllers have been extensively used for controlling sound synthesis processes. This includes signal processing techniques combined with augmented instruments for enhancing European [Aim02, MRM02, Aim06] or Indian [KEDC03, KLD<sup>+</sup>04, Kap07] percussion instruments. A complementary approach to the growing number of gestural percussion instruments is to evaluate such devices towards the accuracy and the naturalness qualities they offer in a performance situation [CCW09].

Despite the quality and variety of such percussion controllers, these works are not directly related to virtual performances since they are mainly used for reacting to natural gestures performed by users. To our knowledge, the only work that can be related to virtual performances is the one developed by A. Kapur on Indian percussion instruments as we will see in the next section.

### 2.3.2.3 Virtual Percussion Gesture Models

Some systems have also been proposed for modeling virtual percussion performances, modeling the virtual performer and its interaction with sound synthesis processes. The most relevant works to that mean have been presented in [GM90], which specifically models a physics-based model of a human arm for hitting a vibrating percussion membrane. Other works have proposed a MIDI-driven system [HST96, SHV97] similarly to *Computer Animation* contributions. Another approach from A. Kapur's contributions models and simulates virtual percussion performances through robotics-inspired systems [Kap05, KTS<sup>+</sup>07, KSB<sup>+</sup>07]. We qualify these works as virtual since the resulting percussion performances are not produced by a human user, but rather involve robotic devices that collaborate to produce percussion performances. These contributions are at the frontier of virtual percussion performances, percussion controllers and automatic composition since it involves augmented percussion devices that are excited by robots and whose behaviors are handled through agent-based formulations.

Similarly to the contributions from the *Computer Animation* community, the most critical issue in these works is the lack of evaluation towards real data from percussion performances. An analysis-oriented evaluation of such contributions could for example take advantage of existing work dealing with the study of percussion performances such as the works presented in [Dah04, Dah05, Dah06].

### 2.3.3 Summary

In this section, we have described research works at the frontier of the *Computer Animation* and *Computer Music* communities, proposing systems for modeling virtual music performances. Works related to *Computer Animation* mostly focus on the simulation of the motion of virtual models, which rarely propose a complete model of a real performer and whose most critical issue is therefore the lack of correspondance to real-world mechanisms occurring in music performances. Works related to *Computer Music* have also proposed models to music performances, and especially in the case of percussion, but the same question of the relation towards real-world mechanisms arises.

This thesis presents one solution to these issues. We propose a complete model of a virtual percussionist whose motion can be easily related to real motion data in order to evaluate its fidelity towards real percussion performances. Moreover, a physics-based approach is adopted for modeling the virtual percussionist, for controlling its motion, as well as for specifying its interaction with a drum model. The design of the proposed system is especially guided by a preliminar analysis

study of real percussion performances.



# 3

## Overview of the Approach

In playing a music instrument, the musician establishes a more or less continuous interaction with the instrument. This interaction is based on complex mechanisms, allowing the fine-tuning of the sound-producing gestures via sensorimotor loops, including audio, visual and gestural feedbacks. Moreover, these sensory informations are directly influenced by the semantic information contained in the musical phrases, and may change the motor commands that produce the gesture. Virtual musical instruments controlled by natural gestures try to realistically reproduce this sensorimotor situation with the aim of approaching real instrumental situations.

Our work exploits the fact that there is a lack of knowledge on significant gestural characteristics for controlling virtual instruments producing sound. Especially when considering percussion gestures, preparatory and retraction gestures are considered as key components of music playing [Dah04]. In this thesis, we propose to go further and introduce a complete modeling of gesture, by designing a human-like character endowed with realistic and expressive behaviors. We particularly focus on the animation of virtual percussionists interacting with sound synthesis processes. Our approach is presented in FIG. 3.1, showing the analysis-synthesis methodology adopted for synthesizing percussion performances.

### 3.1 Analysis

One key issue of our work is therefore to finely analyze and understand (see the "Analysis" part in FIG. 3.1) the underlying processes involved in the control of the virtual performer, and the way it is related to motion captured data recorded during real performances.

In order to achieve these objectives, we highlight in an analysis process the significant dimensions of gesture that are involved when playing various percussion playing techniques. In particular we show that the mallet extremity trajectories

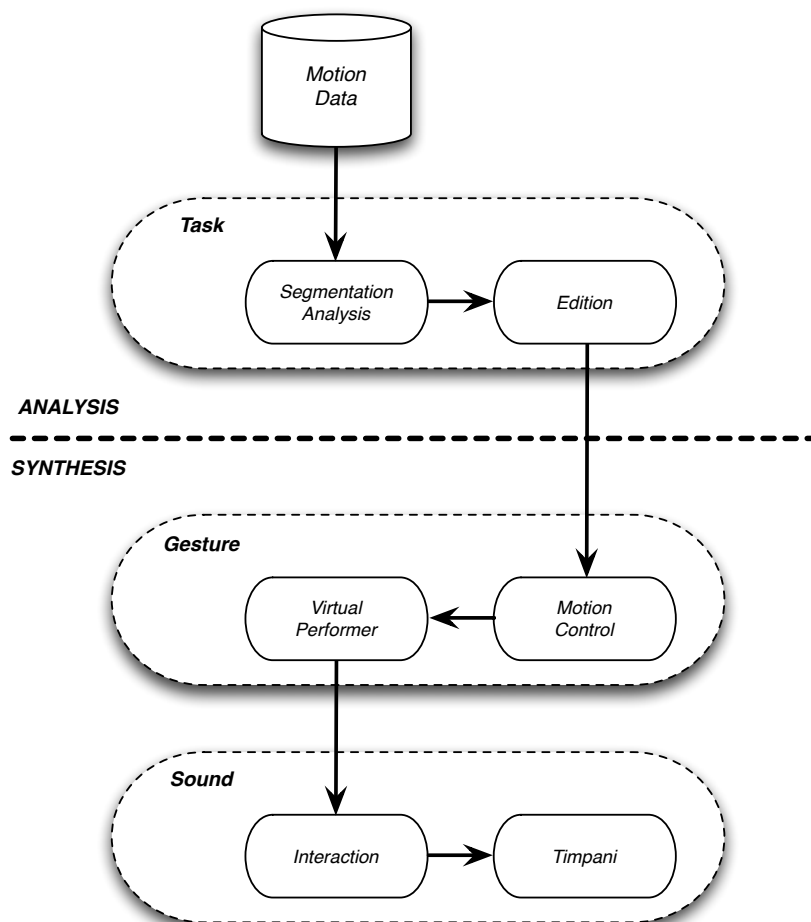


FIG. 3.1. Overview of the approach.

may significantly characterize percussion grips as well as playing modes.

Compared to *computer music* methods for *analyzing* instrumental gestures (see section 2.2.2.2), our work differs from the fact that we analyse the influence of various percussion playing techniques on instrumental gestures. We therefore hypothesize that these playing techniques reflect all or part of the expressivity contained in motion. They may be related for instance to the execution speed of movements, different dynamics or types of attacks. We also identify motion parameters for finely representing these playing techniques.

Such gesture characteristics may then be used for obtaining a segmented and reduced-dimension representation of motion. The resulting motion chunks may be the basis of an edition and assembling process, in which elementary gesture units are arranged and submitted to the synthesis system for controlling the motion of the virtual performer.

## 3.2 Synthesis

In order to reproduce the main characteristics of the control exerted on a real instrument, and more specifically the efforts involved in the interaction, we adopt a physics-based approach for modeling both the virtual performer and the sound-synthesis process (see the "Synthesis" part in FIG. 3.1).

Moreover we provide a new motion control scheme dedicated to percussion gestures for controlling the physics model of the virtual performer.

Among physics-based *computer animation* techniques, our solution comes under the category of *controller-based hybrid* methods (section 2.1.2.3). The main difference compared to these related works lies in the fact that our contribution leads to the control of the virtual performer only by specifying end-effector trajectories, *i.e.* mallet extremity trajectories. This control mode is also shown to be consistent as regards to the previous analysis results.

Compared to *computer music* methods for *modeling* instrumental gestures (see section 2.2.2.2), our framework allows to represent not only what occurs during the mechanical interaction between the instrumentist and the instrument, but also the whole gesture production system. It includes a physics-based solution for taking into account the preparatory and the retraction phases that precede and follow the interaction process. Furthermore, our work sensibly differs from those that use the gestural signals directly responsible for the sound production, as we build the physical model that produces these gestural signals.

The physics modeling approach adopted in this work is then used for allowing

---



to specify the interaction between the physics model of the virtual percussionist and a physics model of a timpani membrane. The output of the gesture synthesis are sent to the timpani model to produce the corresponding sounds, thanks to the identification of kinematic and dynamic features of the gesture attacks. A system architecture is also presented for allowing the real-time interaction between motion and sound synthesis.

### **3.3 Evaluation**

We then proceed to an extensive evaluation of both the analysis and synthesis modules presented in FIG. 3.1.

As for the analysis part of our work, we provide a quantitative evaluation of the extracted motion parameters that are used for representing the percussion playing techniques under study. Compared to works related to the analysis of percussion gestures (section 2.2.2.2), our contribution proposes a set of motion parameters that can clearly and distinctly classify the various playing techniques.

Regarding the synthesis system which is proposed in this thesis, we also provide an extensive evaluation of our control mode. It involves its comparison to classical methods that are traditionally used for physically controlling a virtual character (section 2.1.2.2.2.2). We focus especially on the accuracy response of these control modes, particularly for the critical issue in percussion gestures of finely controlling mallet extremity trajectories.

Finally, we qualitatively evaluate our global framework for synthesizing percussion performances in a musical perspective. This involves the synthesis of percussion exercises that were proposed by a percussion professor. The same percussion expert has then been asked to judge the overall realism and proficiency of the virtual performer to execute these percussion exercises.

# 4

## Analysis of Timpani Percussion Performances

In this chapter, we present the explicit extraction of characteristics of natural percussion gestures. The analysis of pre-recorded percussion (timpani) gestures leads to the identification of significant parameters characterizing percussion playing techniques. The presented parameterization is also evaluated, showing that such parameters are consistent for discriminating the percussion playing techniques under study.

The chapter is organized as follows. We initially introduce a general description of timpani playing techniques in section 4.1, as well as the building and pre-processing of a motion capture database of timpani percussion gestures in section 4.2. We then present the extraction of the parameters that are used for discriminating percussion playing techniques, and we evaluate such a parameterization in section 4.3. Finally, section 4.4 finally concludes this analysis work on percussion playing techniques.

### 4.1 Timpani Basics

There are many classifications of percussion instruments, one of the most established typologies is based on physical characteristics of instruments and the way by which they produce sound. According to this classification, timpani are considered as membranophones, "producing sound when the membrane or head is put into motion" [Coo97]. Timpani are also called "kettledrums", and are considered as the first percussion instrument introduced in the composition of a classical orchestra more than 400 years ago [Bla95]. The dramatic resounding and powerful effects that can emerge from the timpani makes this instrument one of most important percussion instrument in an orchestra.

In this subsection, we depict the basics for playing timpani, including the



**FIG. 4.1.** Timpani equipment: bowl, membrane and mallets.

**TAB. 4.1.** Bowl diameters and corresponding pitch intervals.

<i>Bowl diameter</i>	23"	26"	29"	32"
<i>Pitch interval</i>	$D3 - A3$	$B^{b2} - F3$	$F2 - C3$	$D2 - A2$

presentation of timpani-related equipments, acoustics properties of timpani sound properties, as well as its main playing techniques.

### 4.1.1 Equipment

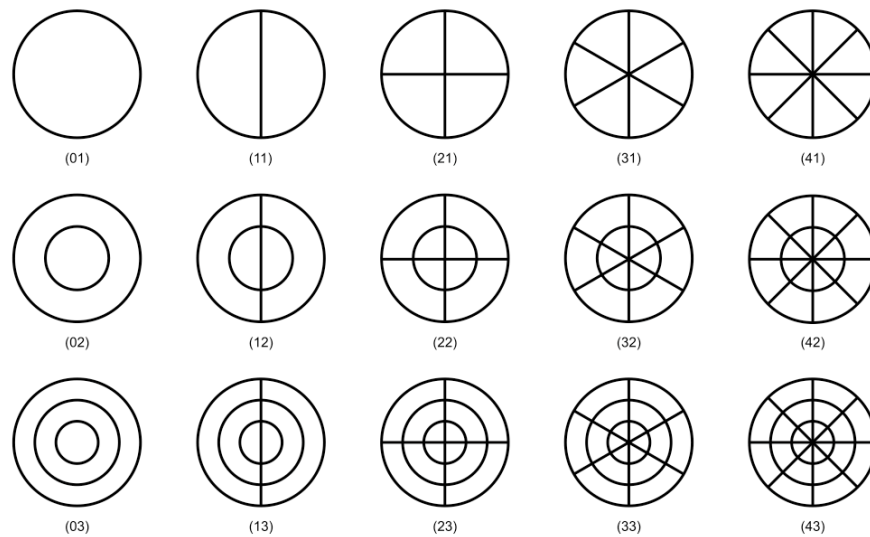
The equipment related to timpani is mainly composed of a bowl, a head and mallets as shown in FIG. 4.1. In general, timpanists have to cope with several timpani (usually four) with bowls varying in size. The diameter of timpani membranes usually range from 23" to 32" [Pet84] and players should properly tune each timpani to its corresponding pitch interval (TAB. 4.1).

As for timpani mallets, they consist of a shaft and a head. They are designed in various lengths, weights, thicknesses and materials [Coo97] and their choice is of great importance [Noa84], namely depending on the percussionist level and on the task to be performed in the instrumental situation (low or fast rolls).

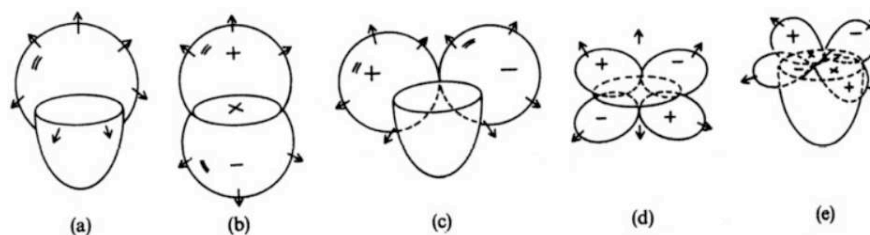
### 4.1.2 Acoustics

The acoustics of timpani percussion instruments has extensively been under study, mostly due to the timpani oddity known as the "missing fundamental" problem. The "missing fundamental" issue is well known among timpanists, this refers to

the fact that a strike in the membrane center result in a muffled sound, so that the fundamental of timpani is considered to be the sound resulting from the '11' vibration mode rather than the '01' vibration mode (FIG. 4.2). Other vibrating modes correspond to harmonics, such as the '21' and '31' modes which are respectively related to a fifth and a major seventh. It has been moreover shown that 'x1' vibration modes characterize predominant harmonic partials [Ros82]. The "missing fundamental" issue can therefore be seen as a vibrating mode shift as the fundamental is shifted from the '01' mode to the '11'.



**FIG. 4.2.** Vibrating (x, y) modes of a timpani membrane, vertically (x) and horizontally (y)(courtesy of [RF91]).



**FIG. 4.3.** Radiation poles of a timpani membrane: (a) monopole radiation of a baffled membrane in its (0, 1) mode, (b) dipole radiation of an unbaffled membrane in its (0, 1) mode, (c) dipole radiation of a baffled membrane in its (1, 1) mode, (d) quadrupole radiation of an unbaffled membrane in its (1, 1) mode, (d) quadrupole radiation of a baffled membrane in its (2, 1) mode (courtesy of [RF91]).

The "missing fundamental" issue has also been explained in radiation terms. The most predominant phenomenon responsible of the vibrating mode shift is the air loading effect [RF91] occurring in the timpani bowl. The radiated energy after a strike corresponding to the '01' is characterized by a monopole radiation scheme, so that the energy is radiated much more rapidly compared to other mode excitations (FIG. 4.3). Conversely other modes are radiating according to di- or quadrupoles that characterize lower radiation velocities.

The "missing fundamental" phenomenon is the reason why timpanists perform rarely a beat attack at the center of the timpani membrane. This observation is taken into account in our protocol for capturing timpani playing techniques, as we will see in the next section.

### 4.1.3 Playing Techniques

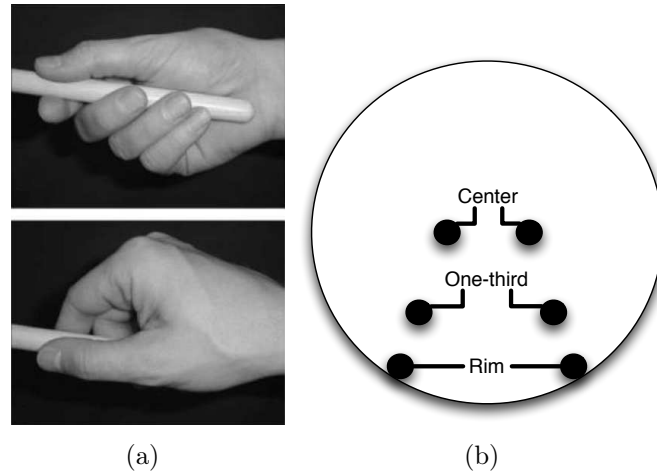
Timpani playing is characterized by a wide range of playing techniques, among them mallet grips as well as beat impact locations on the drum membrane are common features for differencing these techniques.

There are two main strategies for holding mallets, FIG. 4.5(a): the *French* grip (also called "thumbs-up") and the *German* grip (or "matched" grip). These techniques imply different positions of the hand (vertical palm with the *French* grip, horizontal with the *German* grip), thus different motions of the wrist and of the fingers.

Moreover, the use of the arms and forearms differ: broadly speaking, the *French* grip asks for bigger amplitude in the use of the arms and forearms. It should be noted however that many timpanists use other techniques combining features from the two main techniques described here, or with a position of the wrist placing the palm at an angle of 45 degrees with the head of the timpani.

Moreover, players commonly use three distinct locations of impacts (FIG. 4.5(b)): *One-third* of membrane radius, *Center* and *Rim* of the membrane. The most used is definitely the *One-third* location, producing a full sound with a lot of resonance. The *Center* location is characterized by a sharp and muffled attack with barely any resonance. And finally the *Rim* location is characterized by a metallic sound and a resonance bringing mostly out high frequencies. These *Center* and *Rim* beat impact locations are used less often and have been adopted by composers after Eliot Carter's "Eight Pieces for Four Timpani" [Car68].

Various playing modes can also be identified in timpani percussion performances. This includes different gesture variations, such as *legato*, *tenuto*, *accent*,



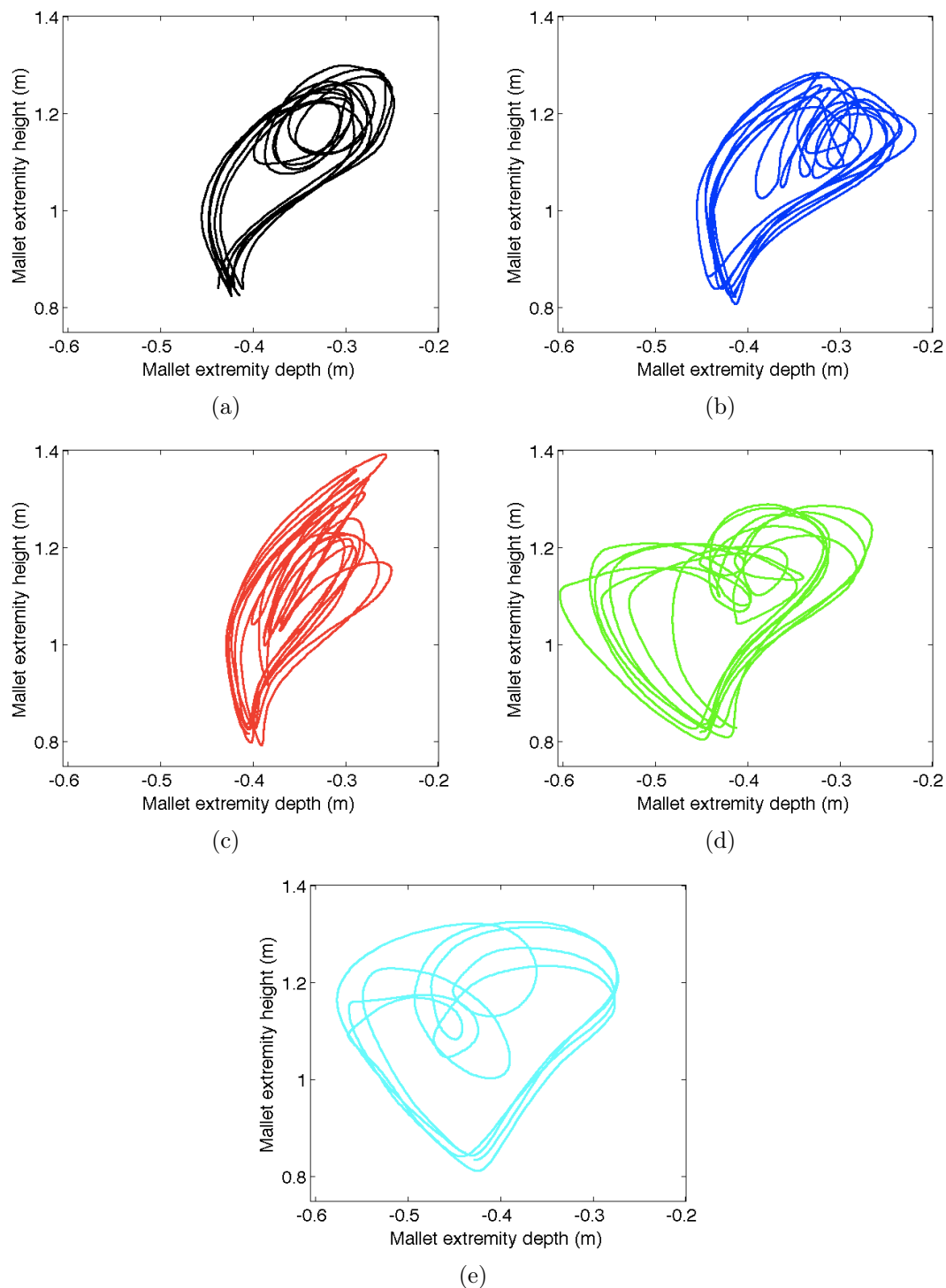
**FIG. 4.4.** (a) *French* (top) and *German* (bottom) mallet grips. (b) Top view of the drum membrane: beat impact locations.

*vertical accent* and *staccato*. FIG. 4.5 depicts examples of these playing modes for the *French* grip, through different profiles of mallet trajectories in the sagittal plane.

A first observation leads to the identification of "air-beats" between two beat attacks, characterized by internal "loops" in the curves. These air-beats can be interpreted as a strategy from performers for keeping a regular tempo and counting the number of left and right beat attacks. It should be noted that these air-beats may not be present for recorded data at a higher tempo.

These playing modes show also different gesture dynamics, both in the longitudinal and vertical directions. While *legato* and *tenuto* playing modes are quite similar, *accent* show a more stiff movement characterized by a bigger vertical amplitude, a lower longitudinal displacement, and therefore more constrained air-beats. Moreover, *vertical accent* and *staccato* shows a wider range of motion in the longitudinal direction, especially with a front projection of the mallet during the retraction phase (after the beat impact).

At a broader scale, whatever technique the timpanist is using, a compromise is generally found between the combined use of 1) the fingers, 2) the tightness of the grip on the shaft of the mallet, location of the grip on the mallet (depending of the length, diameter of the shaft and weight of the head of the mallet), 3) the position of the hand, 4) the amplitude and speed of the wrist, 5) the amplitude and speed of the forearms and arms, and eventually 6) the possibility or not to change the body's center of gravity (timpanist sitting down or standing up, the second position allowing the performer to use his whole body to control the weight, duration and velocity of the impact).



**FIG. 4.5.** Mallet extremity trajectories in the sagittal plane for the various playing modes: (a) *legato*, (b) *tenuto*, (c) *accent*, (d) *vertical accent*, (e) *staccato*. The same color typology will be used further for representing these playing modes.



FIG. 4.6. Timpani performer and the motion capture setup.

## 4.2 Motion Capture Protocol and Database

We captured the motion of several timpani performers by using a Vicon 460 system [Vic09] based on Infra-Red camera tracking, as well as a standard DV camera that allow both the synchronization and retrieval of gesture and sound. Percussionists used a lycra suit fitted with markers placed according to the marker position of Vicon’s *Plug-in Gait*. In order to retrieve beat impacts, markers have also been placed on the mallets. It should be noted that the placement of markers on mallets can have an impact on the recorded performance since it can slightly change their balance. The balance between left and right mallets can also be altered since different placements have been used, as a mean of recognizing left and right mallets during the post-processing phase of motion data. A careful choice of the size of the timpani has been done as regards to capture conditions. A 23” timpani has therefore been used in order to minimize the occlusion of markers by the timpani bowl.

Three percussionists (c.f. FIG. 4.6) were asked to perform a pre-defined capture protocol consisting of a single stroke roll for each gesture (*legato*, *tenuto*, *accent*, *vertical accent* and *staccato*). The capture protocol given to performers is provided in appendix B, as well as the corresponding ethical form.

For each gesture, performers were asked to change the location of the beat impact according to the three locations described in subsection 4.1. In total, fifteen examples of timpani exercises were performed for each percussionist (each with five beats per hand).

TAB. 4.2 presents a summary of the playing characteristics for each performer. The main differences across performers lie in their degree of expertise (*B.A. stu-*



**TAB. 4.2.** Subjects playing characteristics.

<i>Features / Subject</i>	<i>Expertise</i>	<i>Grip</i>	<i>Hand</i>	<i>Gender</i>
$S_1$	Prof.	French	Right	M
$S_2$	B.A.	German	Left	M
$S_3$	M.A.	German	Right	F

dent, *M.A. student* or *Prof.*), the grip strategy used (*French* or *German*), their dominant (*Left* or *Right*) hand, and gender.

One of the main issues using such hardware solutions is the choice of the sampling rate used for the capture of percussive gestures (because of the short duration of the beat impact [Dah97, Wag06]).

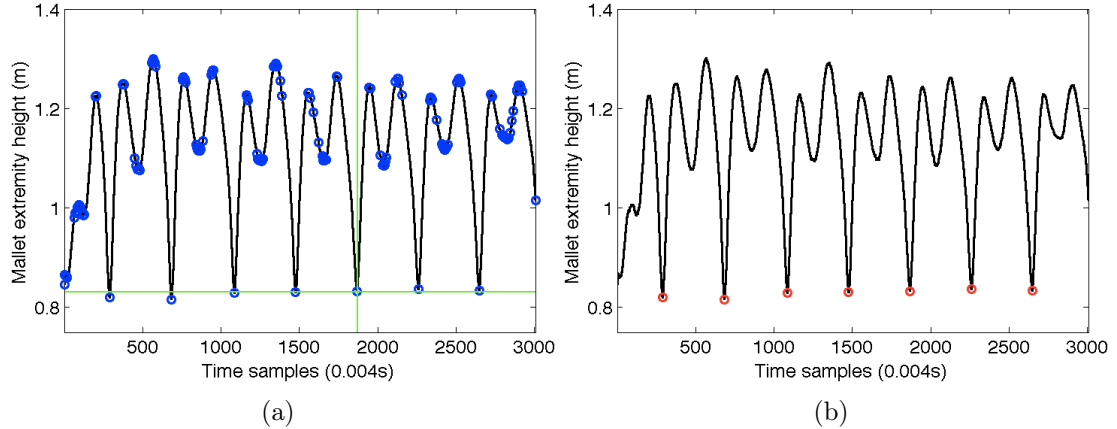
With a high sampling rate (500 Hz and above), one can expect to more accurately retrieve beat attacks, but the spatial range is significantly reduced so that it may be difficult to capture the whole body of the performer. For this project, a compromise was chosen by setting the cameras at 250 Hz, allowing both full-body capture as well as a reasonably high sampling rate for capturing reliably mallet beat impacts.

## 4.3 Analysis

We present here the analysis of timpani performance gestures collected in our database. This analysis focuses on the study of the trace of the percussionist action during musical performance: mallet extremity trajectories. As it can be intuitively stated [BWG08], percussionists more specifically control the motion of mallets over time. The aim of the following analysis is to provide a quantitative analysis of this hypothesis, as well as giving relevant parameters that are of particular interest for the modeling and synthesis part of our work (chapter 5). In order to highlight this assumption, we conduct below a study which aims at putting in evidence the effect of grips and playing modes on preparatory trajectories of mallet extremity.

### 4.3.1 Segmentation of Motion Capture Data

A refinement to the raw data contained in the motion capture database has been achieved as a first step of analyzing the effect of grips and playing modes on timpani preparatory gestures. As originally underlined in [Gib87, Ram91], these gestural units may be obtained by the physical activity of the performer-instrument



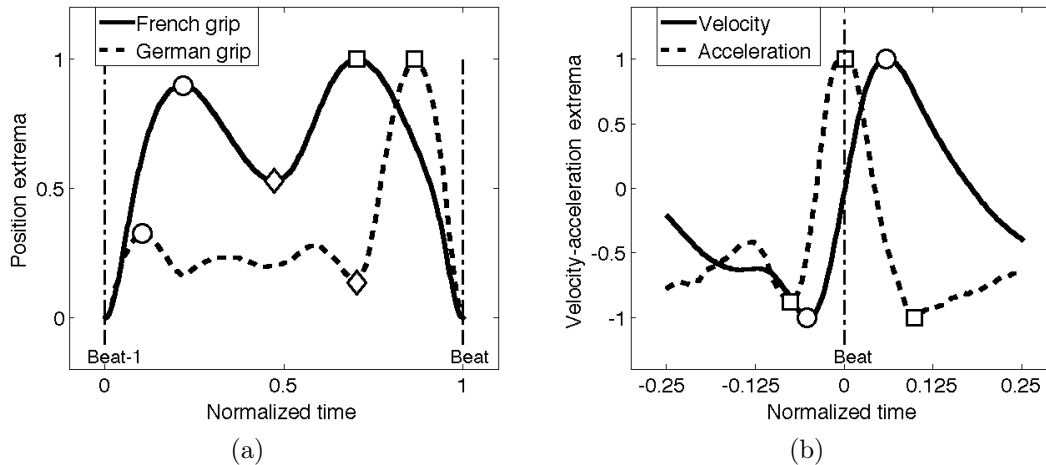
**FIG. 4.7.** Semi-automatic detection of beat impacts: (a) detection of local extrema (blue circles) and selection (green cross) of points of interest, (b) resulting beat impacts (red circles).

interaction (in our case beat impacts during percussion performances). We segment each percussion sequence that has been initially recorded into *beat-to-beat* gesture units. These latter have been identified by processing a detection of beat impacts on mallet height trajectories. FIG. 4.7 depicts a semi-automatic process for, firstly detecting local extrema on mallet trajectories that are potentially beat impacts, and secondly select manually the points of interest. Once *beat-to-beat* gesture units are available for the playing techniques under study (grips and playing modes), it is possible to conduct a fine analysis of the effect of the playing techniques on mallet trajectories.

### 4.3.2 Methodology

The parameters used to characterize mallet trajectories are local extrema extracted from position trajectories during *beat-to-beat preparatory* phases (FIG. 4.9(a)), as well as local extrema from velocity and acceleration *beat-centered* profiles (FIG. 4.9(b)). Each playing mode for a particular grip is represented by a vector of sixteen dimensions, which correspond to the eight amplitude parameters presented in FIG. 4.8 with their corresponding temporal occurrence. These parameters are extracted automatically, therefore building a low-dimension representation of grips and playing modes. FIG. 4.9 depict examples of position, velocity as well as acceleration trajectories for the various playing modes for both *French* and *German* grips.

The evaluation of such a parameterization is conducted by a quantitative analysis based on a classification/recognition scheme. The relevance of these pa-

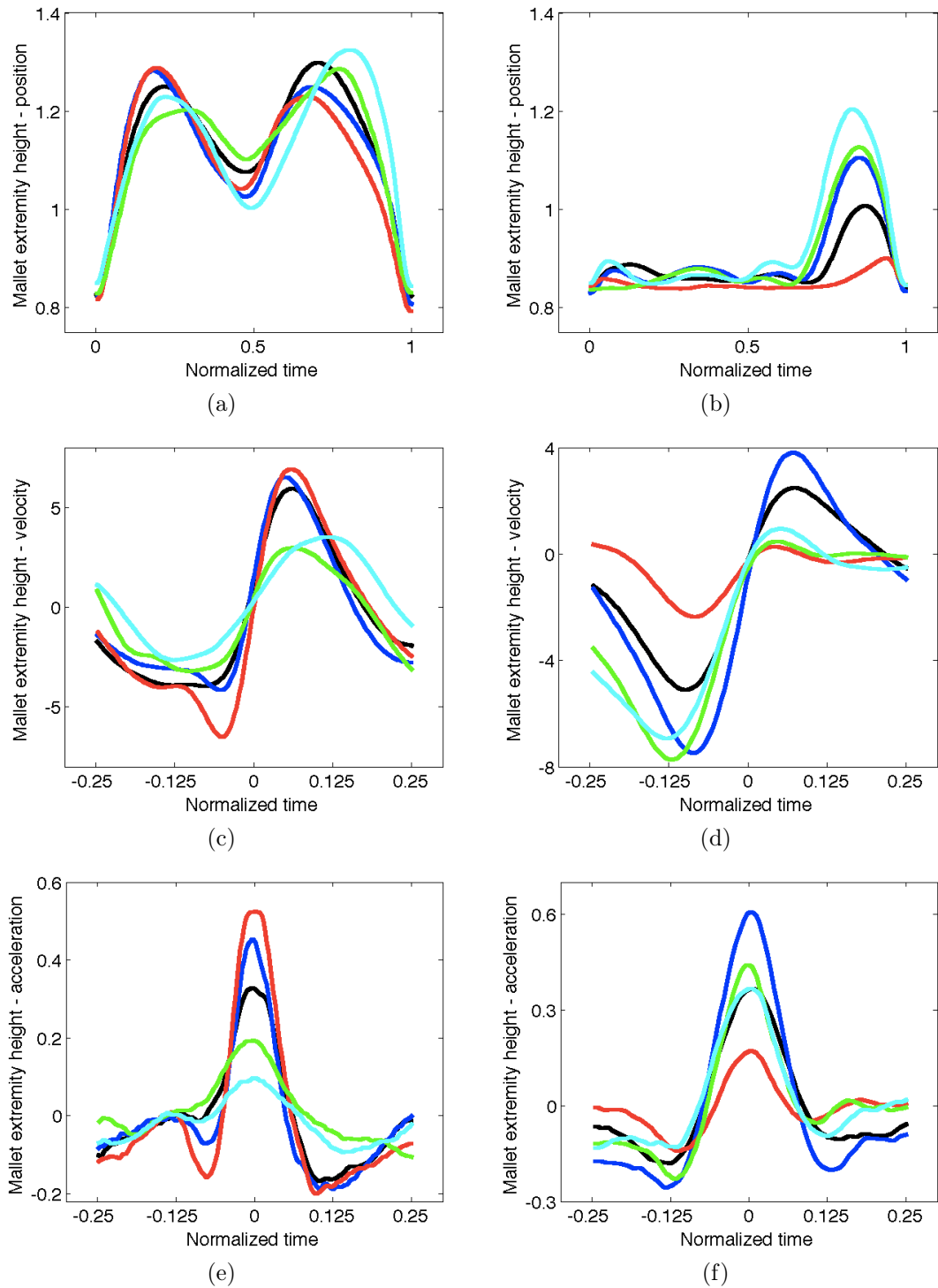


**FIG. 4.8.** Extrema parameters used in the classification/recognition process for characterizing percussion gestures, extracted from: (a) mallet height position during *beat-to-beat preparatory* phases (marker typology:  $E_1 = \circ$ ,  $E_2 = \diamond$  and  $E_3 = \square$ ) and (b) velocity and acceleration *beat-centered* profiles. These parameters are normalized for display purposes.

rameters is measured using the Support Vector Machine (*SVM*) classification method, with the use of Radial Basis Functions (*RBF*) as kernel functions. The scope of this evaluation concerns firstly parameters related to percussion grips (*French* or *German*), and secondly playing modes (*legato*, *tenuto*, *accent*, *vertical accent* and *staccato*). For each case, a combination of parameters is chosen and forms a refined data set of the motion capture database, on which the classification/recognition will work. This refined data set is then divided into two sub-sets, an excerpt of it is randomly extracted to represent determined classes that train a classifier, whereas the remaining data represent queries submitted to the classifier. The relevance of the selected parameters is then estimated according to their recognition success by the classifier. In this quantitative study, the determined classes to recognize typically correspond to two classes for percussion grips, and five classes for playing modes.

### 4.3.3 Percussion Grips

First of all, we focus on the analysis of the influence of *French* and *German* grips on mallet trajectories. Quantitative features (TAB. 4.3) processed on the vertical component of the tip of the mallet show that *French* grip-related data performs the same timpani gesture with much more amplitude. The mean of the mallet extremity height is about twenty centimeters higher than its *German* grip coun-



**FIG. 4.9.** Position, velocity and acceleration profiles, respectively (a), (c) and (e) for the *French* grip, and (b), (d), (f) for the *German* grip.

**TAB. 4.3.** Statistical features computed on mallet trajectories and extracted extrema: vertical Mean ( $M$ ), Variance ( $Var$ ) and Range of Motion ( $RoM$ ), as well as the average temporal (in percentage of gesture’s duration) and spatial characterization of extracted extrema presented in FIG. 4.8.

<i>Stats / Grips</i>	$M$ [m]	$Var$ [m]	$RoM$ [m]	$E_1$ [%/m]	$E_2$ [%/m]	$E_3$ [%/m]
<i>French</i>	1.1	0.1	0.04	0.21 / 1.3	0.47 / 1.1	0.68 / 1.3
<i>German</i>	0.9	0.05	0.02	0.12 / 0.9	0.68 / 0.8	0.85 / 1.0

terpart, with a variance twice higher. This fact is strengthened by the vertical range of motion ( $RoM$ ) of the tip of the stick for *French* grip-related data that is about twice higher than for *German* grip data. The mean of mallet extremity height for *German* grip data shows moreover that the tip of the stick is in average closer to the timpani membrane.

Specific local extrema can also be observed during the preparatory gesture of beat attacks. FIG. 4.8 presents an example of the vertical component of the preparatory gesture between two beat attacks, and the identification of three characteristic extrema denoted  $E_1$ ,  $E_2$  and  $E_3$ . These extrema are temporally (temporal apparition in percentage of gesture’s duration) and spatially characterized in TAB. 4.3.

Vertical extrema  $E_1$ ,  $E_2$  and  $E_3$  are temporally equi-distributed for the *French* grip-related data showing a continuous preparatory gesture, whereas local extrema for *German* grip-related data denote three discontinuous parts. In this latter case,  $E_1$  corresponds to the reaction to the previous beat attack, between  $E_1$  and  $E_2$  the tip of the mallet seems to seek a rest position (during more than the half of the whole movement duration) just above the timpani membrane, while  $E_2$  and  $E_3$  correspond to the amplitude that *German* grip-related data gives for the following beat attack. TAB. 4.3 quantifies also the effect of the *French* and *German* grips on the vertical amplitude of the extrema.

*French* and *German* grips influence the spatial and temporal characteristics of the extracted extrema from height trajectories of the tip of the mallets. For evaluating the relevance of such parameters to distinguish these percussion grips, we chose to use a classification/recognition process. The training set is randomly composed of only  $1/8^{th}$  of the total number of available data, and the query set is composed of the remaining data.

The high recognition rates of these extrema (superior to 97% in average), as

**TAB. 4.4.** SVM recognition of timpani grips (in percentage of success) using the mallet position extrema presented in FIG. 4.8.

<i>Training / Test</i>	<i>French</i>	<i>German</i>
<i>French</i>	<b>98.2</b>	1.8
<i>German</i>	2.7	<b>97.3</b>

**TAB. 4.5.** SVM recognition of *French* grip playing modes (in percentage of success) using the combination of mallet velocity and acceleration extrema presented in FIG. 4.8.

<i>Training / Test</i>	<i>Legato</i>	<i>Tenuto</i>	<i>Accent</i>	<i>V. Accent</i>	<i>Staccato</i>
<i>Legato</i>	<b>96.2</b>	3.1	0	0.7	0
<i>Tenuto</i>	2.1	<b>92.6</b>	3.2	2.1	0
<i>Accent</i>	2.4	0	<b>94.7</b>	2.9	0
<i>V. Accent</i>	0	0	0	<b>93.4</b>	6.6
<i>Staccato</i>	0	0	0	3.3	<b>96.7</b>

**TAB. 4.6.** SVM recognition of *German* grip playing modes (in percentage of success) using the combination of mallet position and acceleration extrema presented in FIG. 4.8.

<i>Training / Test</i>	<i>Legato</i>	<i>Tenuto</i>	<i>Accent</i>	<i>V. Accent</i>	<i>Staccato</i>
<i>Legato</i>	<b>92.4</b>	5.1	0	2.5	0
<i>Tenuto</i>	3.3	<b>93.1</b>	0	1.1	2.5
<i>Accent</i>	2.9	0	<b>94.3</b>	2.8	0
<i>V. Accent</i>	1.7	0	1.6	<b>91.8</b>	4.9
<i>Staccato</i>	1.1	0	0	5.4	<b>93.5</b>

shown by the confusion matrix in TAB. 4.4, indicate that such a parameterization is well-suited for characterizing the effect of percussion grips on the height trajectories of mallets.

### 4.3.4 Playing Modes

Following the same methodology, the considered set of extracted parameters is enhanced for taking into account more timpani playing techniques, namely the different playing modes available in the motion capture database (*legato*, *tenuto*, *accent*, *vertical accent* and *staccato*) for each percussion grip sub-group. These additional parameters are composed of the previously presented mallet height extrema (FIG. 4.9(a)), as well as local extrema extracted from mallet height velocity and acceleration on *beat-centered* profiles (FIG. 4.9(b)). Beat-centered profiles are the truncation of motion to a window of 120 milli-second, 60 milli-second before and after the beat impact occurs.

In both situations for classifying playing modes inside percussion grips, the training set is randomly composed of only  $1/4^{th}$  of the total number of available data, and the query set is composed of the remaining data.

The classification of playing modes related to the *French* grip is achieved by considering the combination of *velocity* and *acceleration* extrema. The results obtained with such a parameterization are presented by the confusion matrix in TAB. 4.5, with an average recognition rate superior to 94%.

Concerning playing modes related to the *German* grip, the classification is achieved by combining both *position* and *acceleration* extrema. The results obtained with such a parameterization are presented by the confusion matrix in TAB. 4.6, with an average recognition rate superior to 93%.

### 4.3.5 Discussion

We discuss in this subsection both the nature of the parameters extracted for separating grips and playing modes, as well as the classification technique used for obtaining such results.

#### 4.3.5.1 Nature of the Parameters

Regarding the nature of the spatio-temporal parameters highlighted through this analysis, we think that they are relevant according to the percussion task, since both the height and the timing of gestures are highly controlled during percussion

performances. The introduction of velocity and acceleration characteristics for discriminating between playing modes among *French* and *German* grips-related data can be interpreted as the parameterization of the dynamics intrinsically related to each playing mode.

More interesting is the use of different parameters' combinations (*velocity/acceleration* for *French* grip playing modes, and *position/acceleration* for *German* grip playing modes). This is to be related to the statistical features presented in subsection 4.3.3.

TAB. 4.3 shows indeed a mallet range of motion much more constrained for the *German* grip, attesting the importance of position parameters for discriminating playing modes in this case. On the contrary, the continuous and equi-distributed preparatory gesture shown for the *French* grip underlines a less stiff constraint on mallet position, so that the way velocity is involved is preponderant for discriminating playing modes in this particular situation. Acceleration parameters are both grip playing modes important features, as they may be related to the effort of the strike.

#### 4.3.5.2 Classification Technique

Another issue to be discussed about the results presented in this chapter is the classification technique used in the methodology of our work. The classification/recognition technique chosen in this work is the *SVM* method with *RBF* kernel functions. The reason of this choice is that the *SVM* technique gives really good results in the discrimination of mallet grips and playing modes. As an illustration of this affirmation, we have conducted the same classification/recognition methodology as presented in subsection 4.3.2 with replacing the *SVM* technique by the K-Nearer-Neighbours (*KNN*) method. TAB. 4.7, 4.8 and 4.9 respectively show the recognition results for the mallet grips, the *French* and *German* playing modes when using the *KNN* technique.

Apart from the results concerning the mallet grips, these results attest lower recognition results compared with the results obtained with the *SVM* technique (TAB. 4.4, 4.5 and 4.6).

Although the explanation of this discrepancy in the recognition quality is out of the scope of our thesis work, a qualitative and intuitive analysis leads us to hypothesize that the classes to be classified and recognized in the case of mallet grips are linearly separable (high recognition rates with the *KNN* technique), whereas those concerning playing modes are not (low recognition rates with the *KNN* method).



Moreover, when detailing the applied methodology in this work, we have precised that we used *RBF* kernel functions with their default parameters, so that no optimization process is involved in the recognition analysis developed in this work. We can therefore expect that an optimization process taking into account an identification of the *RBF* kernel functions would lead to better results than those presented in TAB. 4.4, 4.5 and 4.6.

**TAB. 4.7.** KNN recognition of timpani grips (in percentage of success) using the mallet position extrema presented in FIG. 4.8.

<i>Training / Test</i>	<i>French</i>	<i>German</i>
<i>French</i>	<b>96.4</b>	3.6
<i>German</i>	3.7	<b>96.3</b>

**TAB. 4.8.** KNN recognition of *French* grip playing modes (in percentage of success) using the combination of mallet velocity and acceleration extrema presented in FIG. 4.8.

<i>Training / Test</i>	<i>Legato</i>	<i>Tenuto</i>	<i>Accent</i>	<i>V. Accent</i>	<i>Staccato</i>
<i>Legato</i>	<b>65.5</b>	19.6	11	3.9	0
<i>Tenuto</i>	24.3	<b>59.3</b>	16.4	0	0
<i>Accent</i>	10.8	21.8	<b>67.4</b>	0	0
<i>V. Accent</i>	5.2	10.1	0	<b>65.5</b>	19.2
<i>Staccato</i>	0.2	0	0	20.8	<b>69</b>

**TAB. 4.9.** KNN recognition of *German* grip playing modes (in percentage of success) using the combination of mallet position and acceleration extrema presented in FIG. 4.8.

<i>Training / Test</i>	<i>Legato</i>	<i>Tenuto</i>	<i>Accent</i>	<i>V. Accent</i>	<i>Staccato</i>
<i>Legato</i>	<b>64.7</b>	20	12.6	0.6	2.1
<i>Tenuto</i>	14.9	<b>62.4</b>	16.1	1	5.6
<i>Accent</i>	20.2	14.5	<b>60.7</b>	4.6	0
<i>V. Accent</i>	6.3	5.7	6.7	<b>61.9</b>	18.4
<i>Staccato</i>	5.1	6.5	0	23	<b>65.4</b>

## 4.4 Conclusion

In this chapter, we proposed a methodology for capturing and analysing timpani performances. It has led to the collection of motion capture data for several timpani performers, allowing the study of percussion playing percussion techniques such as mallet grips (*French* and *German*), various playing modes (*legato*, *tenuto*, *accent*, *vertical accent* and *staccato*) as well as different beat impact locations (*one-third*, *center* and *rim*).

This work specifically focused on percussion grips and playing modes, as a mean of justifying the interest and highlighting the importance of the control of mallet extremity trajectories during timpani performances. To that mean, the adopted analysis methodology has accomplished the evaluation of a parameterization of percussion playing techniques. The evaluation was conducted via the the discrimination success of playing techniques under study by training a classifier with the extracted parameters.

Both for the discrimination of percussion grips and playing modes, we presented the extraction of a set of parameters solely regarding mallet extremity trajectories. The evaluation of such a parameterization is proved to be consistent, as attested by the high recognition rates obtained trough this classification/recognition scheme.



# 5

## Synthesis of Timpani Percussion Performances

In this chapter, we propose a physically-based framework, in which a virtual character dynamically interacts with a physically simulated percussive instrument<sup>1</sup>. In particular, we present and evaluate a novel motion control paradigm for controlling a physical model of a virtual character, which is consistent with the analysis work presented in the previous chapter. We also propose an interaction scheme for controlling sound synthesis processes from this physically-based simulation framework for synthesizing virtual percussion performances.

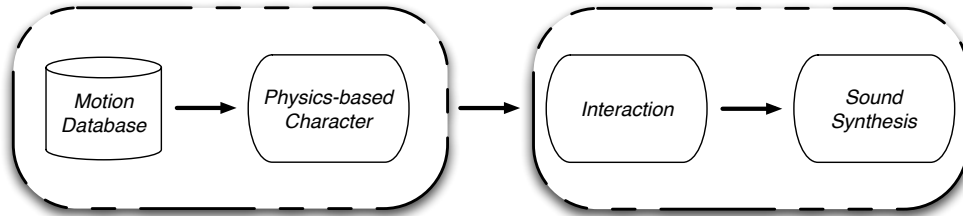
The chapter is organized as follows. We present an overview of the physical mechanisms occurring in our framework during the synthesis of virtual percussion performance in section 5.1. The physics kernel including the physics modeling of the virtual percussionist as well as the novel motion control paradigm is presented and evaluated in section 5.2. Section 5.3 addresses the control of sound synthesis processes by the physical simulation of the virtual percussionist. Finally, section 5.4 concludes this work on the synthesis of virtual percussion performances.

### 5.1 Overview

The architecture of our system is presented in FIG. 5.1 and is made of two modules, allowing the physics simulation of percussion gestures to control sound synthesis processes. The first module includes a motion capture database used for driving a motion control policy model of a physics-based character. The second module expresses the interaction between the percussion motion simulation and the physics-based sound synthesis.

---

<sup>1</sup>The results presented in this chapter are the basis of the simulation platform documented at <http://www-valoria.univ-ubs.fr/Alexandre.Bouenard/index.php/Research/PhDThesis>



**FIG. 5.1.** System architecture and multimodal outputs. The *Physics-based Motion Control and Synthesis* step involves a *Motion Capture Database* and results in the *Visual Feedback*. The *Interaction* expresses the mapping between the percussion motion simulation and the *Sound Synthesis*, which results in the *Sound Feedback*.

Concerning the motion control, the specificity of our contribution lies in the integration and the possible collaboration between inverse kinematics (*IK*) and inverse dynamics (*ID*) controllers, rather than handling strategies for transitioning between kinematic and dynamic controllers [SPF03, Man04, ZMCF05]. We can also find in [ZH99] the use of *IK* as a pre-process for modifying the original captured motion and simulating it on a different character anthropometry. We rather use *IK* as a basis of our hybrid control method for specifying the control of a dynamic character from end-effector trajectories. This hybrid collaboration is particularly consistent for the synthesis of such ballistic motion that is percussion performance, which is not taken into account in related contributions [ZH99] [BGW08].

Concerning the sound synthesis, our work involves physical models based on modal synthesis in a similar manner to recent works [RL06, BDT<sup>+</sup>08]. However we point out to the difficulty of relating these sound synthesis schemes to their cause in the context of music performance, i.e. the instrumental gesture of a virtual performer. Here the focus is on the simulation of instrumental (percussion) gestures, as well as on the interaction between motion and sound synthesis.

In addition, we propose an architecture to allow the synchronization of different modalities and heterogenous data types (motion capture, motion simulation, sound control parameters), as well as to enable the users to explore various interaction schemes between motion and sound simulations. Previous works also involved a motion-driven approach for the synchronized generation of soundtracks from animations [TH92]. But as recalled in [vdDKP01], despite many recent improvements, most of computer animation and simulation frameworks are still reluctant to integrate such synchronization method, moving therefore away from the interactive realism and presence that sound can give to computer animation

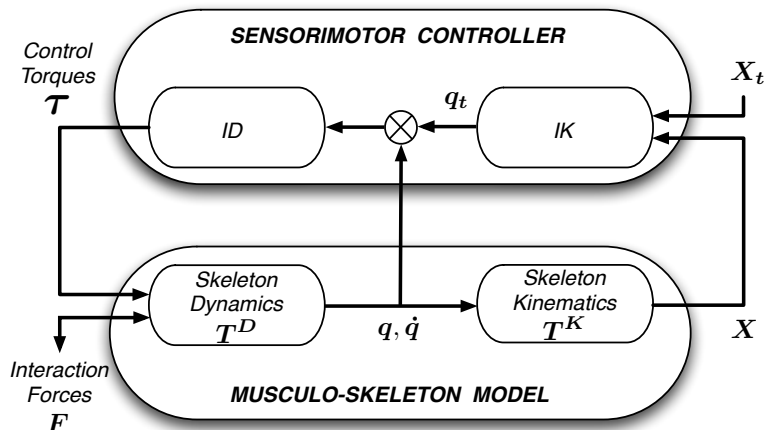


FIG. 5.2. Physics-based modeling and control.

applications. Our goal is to show how a physics modeling of the motion-sound interaction, combined with a carefully-designed system architecture can be of interest to that mean. Such a contribution has been proposed for haptic rendering systems [AC06, SW08], but to our knowledge it has not yet been exploited for the simulation and interaction of instrumental gestures with physics-based sound synthesis.

## 5.2 Physics-based Modeling and Motion Control

The approach that we adopt for physically controlling percussion gestures from motion capture data is described in FIG. 5.2.

It involves the physical modeling of a musculo-skeleton virtual percussionist from pre-recorded percussion performances, by merging the two kinematics ( $\mathbf{T}^K$ , EQU. (2.1)) and dynamics ( $\mathbf{T}^D$ , EQU. (2.3)) representations presented in section 2.1.1.

Our novel motion control paradigm takes advantage of this double and complementary representation by the definition of a sensorimotor controller that tracks motion capture data according to two control modes. The first control scheme uses simple *ID* controllers similarly to previous contributions [Rai86, HWBO95, ZH02], whereas the second mode involves a novel hybrid motion control scheme including a cascaded process of *IK* and *ID* controllers. This latter allows the control of the physical model of a virtual percussionist solely by specifying mallet extremity trajectories ( $\mathbf{X}_t$ ).

## 5.2.1 Virtual Character Modeling

### 5.2.1.1 Representation and Anthropometry

The musculo-skeleton model of the virtual performer is composed of two skeleton layers, a dynamics layer ( $\mathbf{T}^D$ ) and a kinematics layer ( $\mathbf{T}^K$ ). It should be noted that this representation is not generic to any human anthropometry but specific to the simplified anthropometry initially recorded (cf. section 4.2).

The dynamics skeleton  $\mathbf{T}^D$  models the physical properties of the virtual character, and is composed of rigid bodies articulated by mechanical joints. Motion capture data of percussion performances are used to physically model and parameterize both the anthropometry and the mechanical joints of the virtual character, making a direct correspondence between the real performer and the virtual character. The physical properties of each rigid body composing the virtual character, such as mass, size of the limbs, density and inertia matrix are consistent with the anthropometry extracted from motion capture data, by using anthropometric tables [DG67].

The kinematics layer  $\mathbf{T}^K$  of the virtual performer is composed of kinematics linear-angular position and velocity ( $\mathbf{q}, \dot{\mathbf{q}}$ ) of joints articulating the rigid body composing  $\mathbf{T}^D$ . State features  $\mathbf{q}$  and  $\dot{\mathbf{q}}$  are namely obtained by the integration of the motion equations over time (EQU. (2.10) for instance).

### 5.2.1.2 Joints

In addition, each mechanical joint has three rotational degrees of freedom, restricted to the angular limits of the human body, in order to avoid non realistic motion. Two formulations are available to extract the "lower" and "upper" limits of the angular joints, based on a statistical analysis of the motion capture data.

The first formulation uses the Euler angles representation, and computes basic statistical features to characterize joint limits. The singularity of this representation is however frequent (*gimbal lock*), therefore we propose another formulation based on the quaternionic representation.

Following the approach of [Joh03], we compute joint limits in the quaternion space<sup>2</sup>. The rotational (quaternionic) trajectory  $\mathbf{Q}$  of a joint over time is transposed in the tangential space of its quaternionic mean  $\bar{q}$ , in which an SVD decomposition is applied. The resulting eigen values and axes ( $\boldsymbol{\lambda}, \mathbf{e}$ ) are used for

---

<sup>2</sup>Thanks go to the developers of the SMR library from the VALORIA-SAMSARA lab, and particularly to N. Courty for helpful hints about the quaternion representation.

representing the motion distribution and for computing the joint limits by the specification of a bounding ellipsoid (ALG. 5.1).

---

**ALG. 5.1:** Formulating joint limits in the quaternion space

---

	In $\mathbf{Q} = \{q_i, i \in [[1 \dots n]]\}$	Quaternionic time serie
Out	$\{\boldsymbol{\lambda}, \mathbf{e}\}$	Eigen values and axes
<p><i>** Mean of <math>\mathbf{Q}</math> **</i></p> <p><math>\bar{q} = \text{mean}(\mathbf{Q})</math></p> <p><i>** Transposition of the motion <math>\mathbf{Q}</math> in the tangent space of <math>\bar{q}</math> **</i></p> <p><math>\mathbf{Q}^* = \log(\bar{q} * \mathbf{Q})</math></p> <p><i>** Eigen values and axes **</i></p> <p><math>\{\boldsymbol{\lambda}, \mathbf{e}\} = \text{SVD}(\mathbf{Q}^*)</math></p>		

---

## 5.2.2 Motion Control

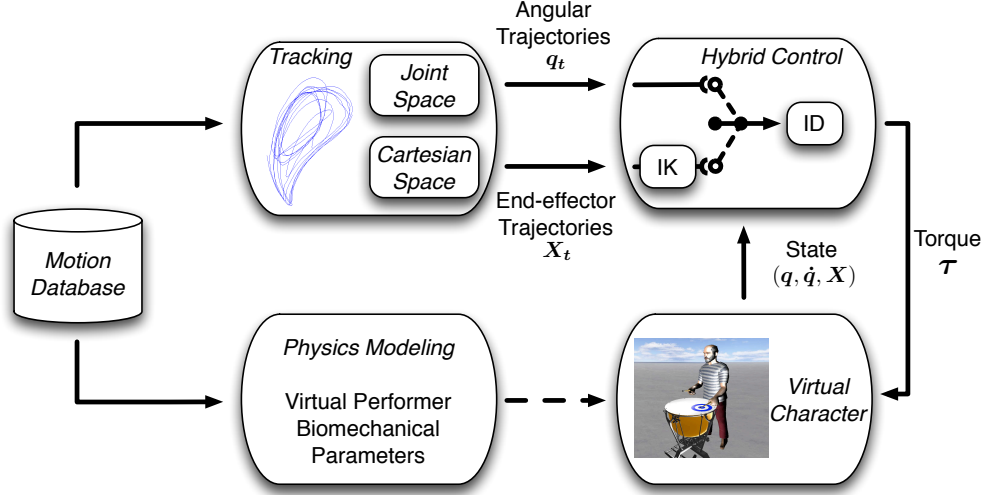
Our approach to dynamic character control uses percussion gestures from a motion capture database, allowing to take into account all the variability and expressiveness of real percussion gestures. The variability can be due to various percussion performances using different mallet grips, various beat impact locations and several musical playing modes (see section 4.2).

We propose two ways for achieving the motion control of the musculo-skeleton model (see FIG. 5.3), either by tracking motion capture data in the joint space (angular trajectories), or a novel hybrid motion control scheme for tracking end-effector trajectories in the 3D cartesian space. Tracking motion capture data in joint space requires *ID* control, whereas tracking in the end-effector space requires both a cascaded collaboration of *IK* and *ID* controllers. In this latter case, the two inversion processes are strongly linked.

### 5.2.2.1 *ID* Motion Control

This control mode is related to motion capture data tracking by using *ID* controllers. Angular trajectories ( $\Theta_t$ ) are first extracted from motion capture data, and used to drive the fully dynamically controlled virtual character. We use traditional proportional-derivative feedback controllers, modeled as damped springs





**FIG. 5.3.** Hybrid physics-based motion control and synthesis. The *Motion Capture Database* is used for the physics modeling of the *Virtual Character*, and for expressing two levels of *Tracking*. These two levels allow the physics-based motion capture tracking, either in the *Joint Space* from angular trajectories  $\mathbf{q}_t$ , or in the *Cartesian Space* from end-effector trajectories  $\mathbf{X}_t$ . The hybrid control involves the combination of inverse kinematics (*IK*) and inverse dynamics controllers (*ID*).

and parameterized by manually-tuned damping and stiffness coefficients ( $k_d, k_s$ ). Knowing the current state of the mechanical joint ( $\mathbf{q}, \dot{\mathbf{q}}$ ) and the joint target ( $\mathbf{q}_t$ ) to be reached, the torque ( $\boldsymbol{\tau}$ ) is computed and exerted on the articulated rigid bodies, accordingly to EQU. (5.1).

$$\boldsymbol{\tau} = k_s \cdot (\mathbf{q}_t - \mathbf{q}) - k_d \cdot \dot{\mathbf{q}} \quad (5.1)$$

### 5.2.2.2 Hybrid Motion Control

A more intuitive physics control of the virtual character is proposed by combining *IK* and *ID* controllers. Instead of directly tracking angular trajectories from the motion capture database, this tracking mode consists in extracting end-effector positions in the 3D Cartesian space. From these Cartesian targets, an *IK* method computes EQU. (5.2) the kinematic postures (joint vector  $\mathbf{q}_t = \{q_1, \dots, q_n\}$ ), which are used as the desired input of the *ID* controllers (ALG. 5.2), thus providing the required torques to control the physical character (FIG. 5.3).

$$\begin{aligned} \Delta \mathbf{q}_t &= \mathcal{J}^{-1}(\mathbf{q}) \cdot (\mathbf{X}_t - \mathbf{X}), \quad \mathbf{q}_t = \mathbf{q} + \Delta \mathbf{q}_t \\ \boldsymbol{\tau} &= k_s \cdot (\mathbf{q}_t - \mathbf{q}) - k_d \cdot \dot{\mathbf{q}} \end{aligned} \quad (5.2)$$

**ALG. 5.2:** Hybrid motion control combining *IK* and *ID* controllers

In	$AC = \{\mathbf{T}^K, \mathbf{T}^D\}$	Articulated chain
	$F_T$	Task frequency
	$F_S$	Simulation frequency
	$\mathbf{q}, \dot{\mathbf{q}}$	Current kinematic posture of $T^K$
	$\mathbf{X}$	Current end-effector position of $T^K$
	$\mathbf{X}_t$	Targetted end-effector position of $T^K$
	$\mathcal{J}$	Jacobian matrix
Out	$\mathbf{q}_t$	Targetted kinematic posture of $T^K$
	$\boldsymbol{\tau}$	Resulting torques to be exerted on $T^D$

```

while Read task at  $F_T$  do
   $\mathbf{X}_t \leftarrow$  New mallet extremity

  numIKID = 0

  for numIKID <  $F_S/F_T$  do
    ** Inverse Kinematics **
     $\mathcal{J} \leftarrow$  ComputeJacobian( $\mathbf{T}^K, \mathbf{q}$ )
     $\mathbf{q}_t \leftarrow$  InverseKinematics( $\mathcal{J}, \mathbf{X}, \mathbf{X}_t, \mathbf{q}$ )

    ** Inverse Dynamics **
     $\boldsymbol{\tau} \leftarrow$  InverseDynamics( $\mathbf{q}, \dot{\mathbf{q}}, \mathbf{q}_t$ )

    ** Update **
    ApplyTorques( $\mathbf{T}^D, \boldsymbol{\tau}$ )
    Update( $\mathbf{T}^K, \mathbf{T}^D$ )
  end
end

```

$\mathcal{J}$  represents the Jacobian of the system to be controlled,  $\mathbf{X}$  and  $\mathbf{X}_t$  represent respectively the current and target end-effector positions in the Cartesian space. The modularity of the presented approach for combining *IK* and *ID* controllers allows to consider any method, so that our system supports at the moment several *IK* formulations (transpose, pseudo-inverse and damped-least-squares, see section 2.1.2.1.2) that can be combined with simple *ID* formulations. One may equally use other *IK* techniques, such as learning techniques described in [GM03, GM07].

ALG. 5.2 shows the principle of the hybrid motion control, composed of two nested loops. The first one concerns the task specification (mallet extremity position). Between two task updates, the second loop involves the cooperation of the *IK* and *ID* formulations. For example, for a task frequency of 250 Hz and a simulation frequency of 10000 Hz, the second simulation loop is called  $10000/250=40$  times, which can be sufficient so that the *IK* algorithm converges, and consequently ending with the *ID* convergence.

Using such an *IK* formulation necessitates the computation of the Jacobian matrix, and therefore we defined an equivalent representation of the articulated chain, both in the kinematics and dynamics spaces. The main difficulty with the coupling of both kinematics and dynamics controllers is that the convergence of the *IK* algorithm is added to the difficulty of tuning the parameters of the *PD* dynamic controllers. But this approach enables the manipulation of motion capture data in the 3D Cartesian space (configuration  $\mathbf{X}_t$ ) instead of the angular space ( $\mathbf{q}_t$ ), which is more consistent and intuitive for controlling percussion gestures, by using end-effector trajectories, for instance mallets extremities.

### 5.2.3 Results

The physical model of the virtual character is composed of 17 joints, totalling 51 degrees of freedom. The Open Dynamic Engine [Smi09] is used for the forward dynamics simulation of the motion equations. Masses, inertia, limb lengths, as well as joint limits are estimated from real percussion performers. Concerning the hybrid control mode, many inverse kinematics methods may be involved in our cascaded control mode. These results have been obtained using an implementation of the damped-least-squares method [Wam86], a simple yet more robust adaptation of the pseudo-inverse regarding the singularity of the inverse kinematics problem. The hybrid control scheme presented previously is tested on the control of the two arms of the virtual character, tracking a sequence of percussion gestures for synthesizing whole arm movements only from the specification of the mallet extremity trajectories.

The results presented in this section compare the traditional *ID* control mode to our hybrid control scheme. It should be noted that for making such a comparison possible, the parameters that can change drastically the resulting simulations have been kept constant. Such parameters include for instance the simulation rate, the recorded motion to be simulated as well as the parameterization of the damped springs composing the dynamic layer of the virtual character.

A qualitative evaluation of the presented hybrid motion control is first con-

ducted. The results obtained by the two tracking modes used to physically control the virtual percussionist are compared: a) motion capture tracking in joint space, involving only *ID* controllers, and b) motion capture hybrid tracking in Cartesian space, involving the cascaded combination of *IK* and *ID* controllers. In a second time, we quantitatively evaluate the two tracking modes by quantifying the errors made during the simulation of various types of percussion gestures in each control case. An evaluation directed by a classification/recognition methodology analogous to chapter 4 is also conducted.

Concerning the quantitative evaluation, a critical issue to consider is the metric used for evaluating the effectiveness of our hybrid solution compared to traditional *ID* control. A few works have addressed the question of the user sensibility to various motion metrics [RP03, RAE<sup>+</sup>05]. In this work, the ballistic nature of percussion motion makes it however easy to conduct an extensive evaluation by focusing on mallet extremity trajectories. We also show that our solution leads to a coherent motion in the joint space for specific degrees of freedom that are shown to be of paramount importance in percussion playing.

### 5.2.3.1 Qualitative Evaluation

For this qualitative evaluation, we ran the simulation of a set of pre-recorded percussion gestures (*French* grip, *legato*). Figure 5.4 compares raw data from captured motion with the two modes of control (*ID* only, and the hybrid combination of *IK* and *ID*).

FIG. 5.5(a), 5.5(c) and 5.5(e) present the comparison between raw motion capture data and data generated by the two control modes for wrist, elbow and shoulder flexion angle trajectories. These results show that data generated by the *IK* involved in the hybrid control mode are consistent with real ones, especially for wrist and elbow trajectories. Data generated by our hybrid solution is more accurate compared to *ID* control, this latter shows indeed a motion exaggeration that tends to be avoided by our solution.

More specifically about shoulder flexion angle, FIG. 5.5(e), it seems that the *IK* process loses the pattern of the original motion, while leading to more accurate extrema occurring at beat impacts compared to *ID* control. This loss is however of lower importance in percussion gestures, as attested in [BWG08] where it is argued that wrist and elbow angle trajectories are one of the most important degrees of freedom in percussion arm mechanisms.

We also present the result comparison of the two control modes concerning height trajectories for the mallet extremity, wrist and elbow in FIG. 5.5(b),

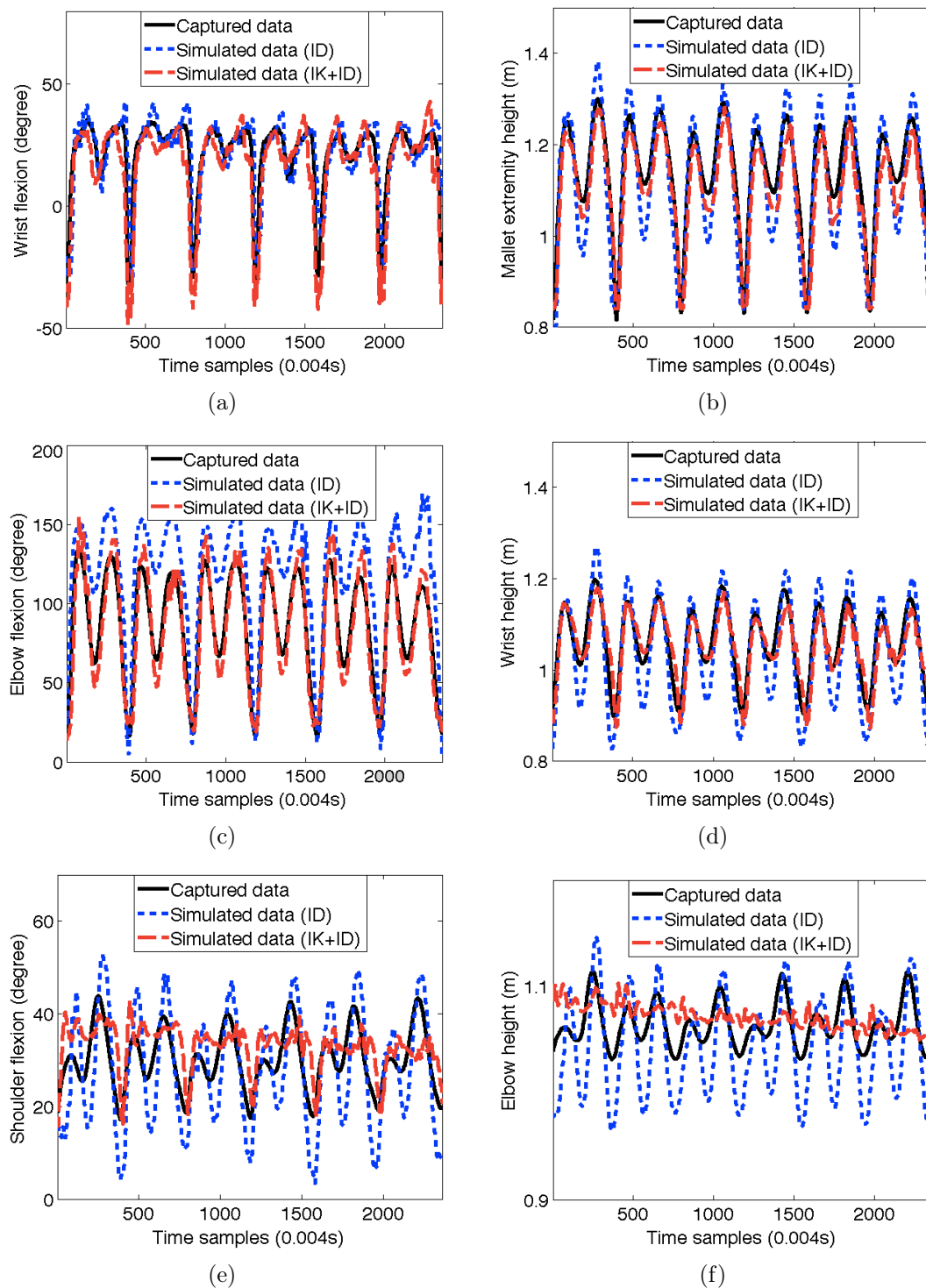


FIG. 5.4. Comparison of captured and simulated trajectories using the two motion control schemes: (a), (c) and (e) resp. for wrist, elbow and shoulder flexion angles, (b), (d) and (f) resp. for mallet, wrist and elbow height positions.

5.5(d) and 5.5(f). One interesting issue is the accuracy of the hybrid control mode compared to the simple *ID* control. The motion exaggeration observed previously in joint flexion trajectories in the case of *ID* control is propagated on joint position trajectories. Conversely, our hybrid solution leads to a more accurate tracking in joint position trajectories. This observation lies in the fact that the convergence of motion capture tracking is processed in the joint space in the case of *ID* control, adding and amplifying multiple errors on the different joints and leading to a greater error than processing the convergence in the Cartesian space for the hybrid control.

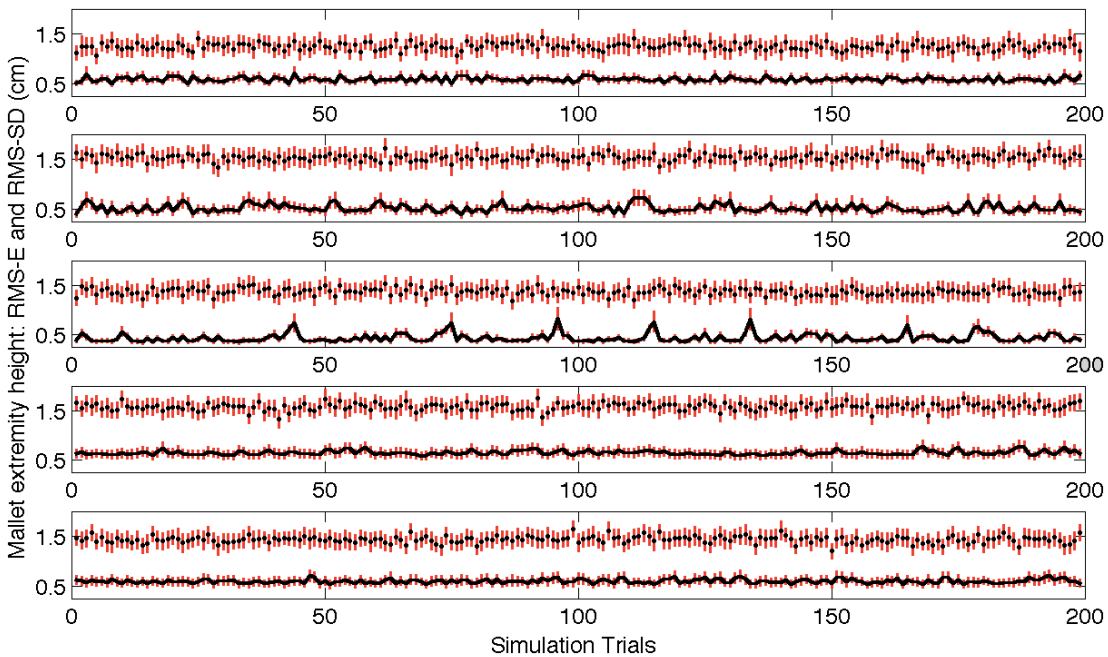
One limitation however arises in elbow position trajectories. While the *ID* control mode results again in an amplified motion, our hybrid control solution results in a loss in the motion pattern of original data. The elbow position trajectory resulting from the hybrid control mode leads to a more stiff and somewhat constant motion. This can be related to the inverse kinematics scheme which does not include any secondary goals, as already pointed out in [KH83]. This may be explained also by the fact that the *IK* algorithm favors the accuracy of the most distal joints compared to joints at the basis of the articulated chain. Another reason to this limitation could be the influence of the physics simulation on the convergence of the inverse kinematics scheme.

Although such cascaded chain of *IK* and *ID* controllers runs in real-time, the main drawback of this improvement is however the additional computational cost of the *IK* algorithm which is processed at every simulation step. It provides nevertheless a more flexible motion edition technique for controlling a physics-based virtual character solely by end-effector trajectories. It also provides a consistent control scheme as regards to the preliminary analysis conducted in chapter 4.

### 5.2.3.2 Quantitative Evaluation

An extensive evaluation of the two control modes is presented in FIG. 5.5. This takes into account the test of the two control algorithms on 200 simulation trials for various playing modes (from top to bottom: *legato*, *tenuto*, *accent*, *vertical accent* and *staccato*). FIG. 4.5 shows examples of longitudinal-vertical phases of the mallet extremity position for each playing mode. This emphasizes that these playing modes present drastically different gesture dynamics in mallet motion, especially regarding its longitudinal and vertical amplitude. Each simulation trial corresponds to a gesture unit that is simulated according to the two control modes. The root mean square error and its standard deviation are processed as regards to the mallet extremity compared to the corresponding motion initially chosen in the motion capture database.

From the results presented in FIG. 5.5, one can easily conclude that the cascaded hybrid combination of *IK* and *ID* controllers leads to a more accurate minimization of the error committed on mallet extremity trajectories of about 1 centimeter, compared to simple *ID* controllers. The standard deviation along simulation trials are also more minimized for the hybrid motion paradigm compared to the *ID* control.



**FIG. 5.5.** Comparison of RMS errors between captured and simulated mallet trajectories for various playing modes, using the two motion control schemes: *ID* (point line) and *IK + ID* (plain). For each playing mode (from top to bottom: *legato*, *tenuto*, *accent*, *vertical accent* and *staccato*), Root Mean Square Error (RMS-E, black) and Standard Deviation (RMS-SD, red) are computed on 200 simulation trials.

It should be noted that this average difference of 1 centimeter of accuracy cannot be considered as negligible, since the sound synthesis model taking mallet extremity informations around the beat impacts can change drastically the nature of the resulting sounds. For instance, in the case of modal sound synthesis models, the accuracy of mallet extremity trajectories is of paramount importance since it will change the selection of the excited mode shapes.

Another evaluation method has been considered for insuring the accuracy of

**TAB. 5.1.** SVM recognition of simulated *French* grip playing modes (in percentage of success) using the combination of mallet velocity and acceleration extrema presented in Fig. 4.8

<i>Training / Test</i>	<i>Legato</i>	<i>Tenuto</i>	<i>Accent</i>	<i>V.Accent</i>	<i>Staccato</i>
<i>Legato</i>	<b>92.6</b>	4.8	0	2.6	0
<i>Tenuto</i>	3.7	<b>94.2</b>	1.2	0.9	0
<i>Accent</i>	2.5	0	<b>96.4</b>	0	1.1
<i>V.Accent</i>	0.5	0.6	0	<b>93.9</b>	4.7
<i>Staccato</i>	0	0	0.5	4.4	<b>95.1</b>

the hybrid motion control scheme. In an analogous manner to the evaluation of percussion playing characteristics in chapter 4, a classification/recognition methodology has been adopted. A SVM classifier with RBF kernel functions has been trained with the extracted parameters described in section 4.3 from motion capture data, thus forming the training set. These gesture characteristics have also been extracted from the 200 simulations trials, which are this time forming the query set. The results of the classification/recognition of simulated playing modes (*legato*, *tenuto*, *accent*, *vertical accent* and *staccato*) are summarized in the confusion matrix in TAB. 5.1.

These results attest the consistency of the novel hybrid motion control scheme presented in this section, giving an average recognition rate of more than 94% of success for the simulated playing modes.

## 5.2.4 Conclusion

We presented in this section the derivation of the modeling and motion control of a physical model of a virtual performer which has the particularity of merging two complementary (kinematics and dynamics) representations. We also derived consequently a novel hybrid motion control scheme dedicated to percussion gestures that allows the control of the physical virtual performer solely by the specification of mallet extremity trajectories. We have particularly shown that this hybrid motion control can be more effective than simple *ID* controllers for percussion gestures. Moreover, in an edition step, managing the control of a physical model only by 3D Cartesian targets is more convenient than handling whole kinematic postures.

From a musical point of view, and according to the conclusions of the analysis study conducted in chapter 4, such a hybrid motion control scheme is also more



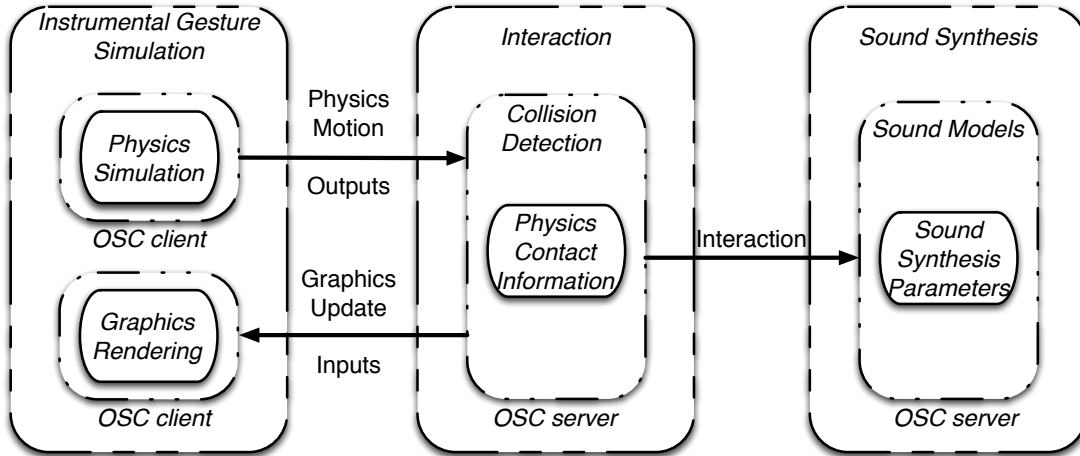


FIG. 5.6. Asynchronous server-client architecture of our system.

consistent than *ID* control as regards to percussion tasks, since our hybrid control scheme only needs the specification of mallet extremity trajectories. In the next section, we focus on the interaction between motion and sound simulations (FIG. 5.1) so that the simulated percussion actions of the virtual performer can influence the sound production process.

### 5.3 Interaction between Motion and Sound Synthesis

We propose a general architecture (FIG. 5.6) that allows to simultaneously and asynchronously run the physics simulation of percussion gestures, graphics and sound rendering processes, as well as to handle their interaction in real-time. This architecture is effective for managing different modalities and data types, and for specifying at the physics level the interaction between these former.

#### 5.3.1 Asynchronous Client-Server Architecture

Our system makes possible the multimodal integration, interaction and synchronization of visual and sounding media, and is composed of four components represented as physics, graphics, interaction and sound managers.

Multimodal processes such as graphics and sound rendering are fundamentally different, and achieving a real-time interaction between both of them can be hazardous. The first difficulty that appears when managing such media is their discrepancy in time constants: graphics rendering is usually admitted effective at

about 33Hz, whereas sound rendering needs a higher time sample rate around 44kHz. Moreover, the graphics rendering is the visible layer of a far more demanding process that is the physics simulation, requiring the handling of two other different time rates since our method relies on motion capture tracking: the original motion capture time rate and the time step of the physics simulation. An approach could consist in running synchronously every manager at the sound rate, but such an approach would fall short in real-time considerations.

We therefore propose an asynchronous client-server scheme for handling the interaction between the physics, graphics and sound managers. This enables the possible distribution of the different managers on distinct platforms, thus reducing the computational cost of each manager and its impact on the others.

For this purpose, we define a communication layer between the managers, based on the Open Sound Control (*OSC*) communication protocol. The *OSC* protocol has been primarily designed to enable the networking and interoperability between interactive computer music applications [Wri05]. Its scheme is based on a "transport-independent" server-client relationship, supporting any network technology (ethernet or wireless, UDP or TCP/IP) and serial connection. Similarly to MIDI, the *OSC* scheme is based on the interaction of communicative points and assigns values to these. *OSC* differs fundamentally from MIDI in handling namespaces, data types and time retrieval. *OSC* features intuitive naming and addressing rules (on the contrary to MIDI byte encoding), as well as an extended set of addresses and namespaces (compared to the limited and limiting six channels of MIDI). *OSC* allows also the transmission of any type of structure of any size, even MIDI data, providing a standard for many software or hardware developments. Eventually, *OSC* features message time tagging for managing priority and anticipating future events. For a complete derivation of *OSC* principles and functionalities, readers are referred to [WFM03, FS09, OSC09].

The asynchronous exchange of data is materialized here by the interaction between the output of motion synthesis and the input of sound synthesis. Events produced during the physical simulation are dealt by the interaction module which feeds the sound synthesis processes and the visual outputs.

### 5.3.2 Motion-Sound Physics Interaction

For sound synthesis, we consider in our work the modal synthesis technique. Note however that the modularity of our system allows to plug any sound synthesis model to the instrumental gesture simulation. Modal synthesis is an efficient way of representing the vibrations of resonating objects as the motion simulation of systems composed of masses connected with springs and dampers (cf. section 2.2.1.2.2). This physically-based approach enables the direct modeling of the

contact force impact and its effect on mode shapes. According to [AR01], this force depends on the state (displacement and velocity) of the colliding modal objects.

We define in this section a practical interaction between the instrumental gesture simulation and a modal model of a drum membrane. Let us suppose first that such a model is available, and consider a slightly modified version of the model derived in appendix A.6 for taking into account an interaction force on the drum membrane. The model is governed by the motion of  $N$  coupled oscillators, that can be excited by a force  $\mathbf{f}$ , EQU. (5.3).

$$\mathbf{M}.\ddot{\mathbf{z}}(t) + \mathbf{K}.\mathbf{z}(t) = \mathbf{f} \quad (5.3)$$

The factorization of the problem considers a solution of the form  $\mathbf{z}(t) = \mathbf{s}.\sin(\omega.t + \phi)$  where  $\mathbf{s}$  are again referred as the *modal shapes* of the system. This leads to the equivalent normal formulation of the problem in the orthogonal basis made of the mode shapes, EQU. (5.4) (see appendix A.6 for intermediate steps).

$$\mathbf{M}_S.\ddot{\mathbf{q}}(t) + \mathbf{K}_S.\mathbf{q}(t) = \mathbf{S}^T.\mathbf{f} \quad (5.4)$$

with  $\mathbf{S} = [s_1, \dots, s_N]$ ,  $\mathbf{z} = \mathbf{S}.\mathbf{q}$ ,  $\mathbf{M}_S = \mathbf{S}^T.\mathbf{M}.\mathbf{S}$  and  $\mathbf{K}_S = \mathbf{S}^T.\mathbf{K}.\mathbf{S}$

EQU. (5.4) shows explicitly that the *modal shapes* of the system (matrix  $\mathbf{S}$ ) defines how the external force  $\mathbf{f}$  acts on the modes. As an illustration, an external force acting only on the  $j^{th}$  mass of the model is transmitted to the  $i^{th}$  mode by a scaling factor  $s_{i,j}$  (the modal shape of the  $i^{th}$  mode at the  $j^{th}$  mass node). If the  $i^{th}$  mode has a shape that contains the  $j^{th}$  mass node of the model, then the scaling factor  $s_{i,j}$  is null, and finally no force is "applied" to this mode. According to EQU. (5.4), the vertical oscillation of the  $j^{th}$  mass of the model can be written as  $z_j(t) = \sum_{i=1}^N s_{i,j}.q_i(t)$ , such that if the  $i^{th}$  mode has a node at the  $j^{th}$  mass this mode will not be heard at that listening point.

The interaction manager includes a collision detection algorithm (FIG. 5.6) that can retrieve the physical features of any contact event produced when the drum is excited during the instrumental gesture simulation. In particular, it provides information on the impact position, velocity and force (direction and amplitude), which can be related to the previously presented inputs to a modal model of a drum membrane according to a direct one-to-one mapping.

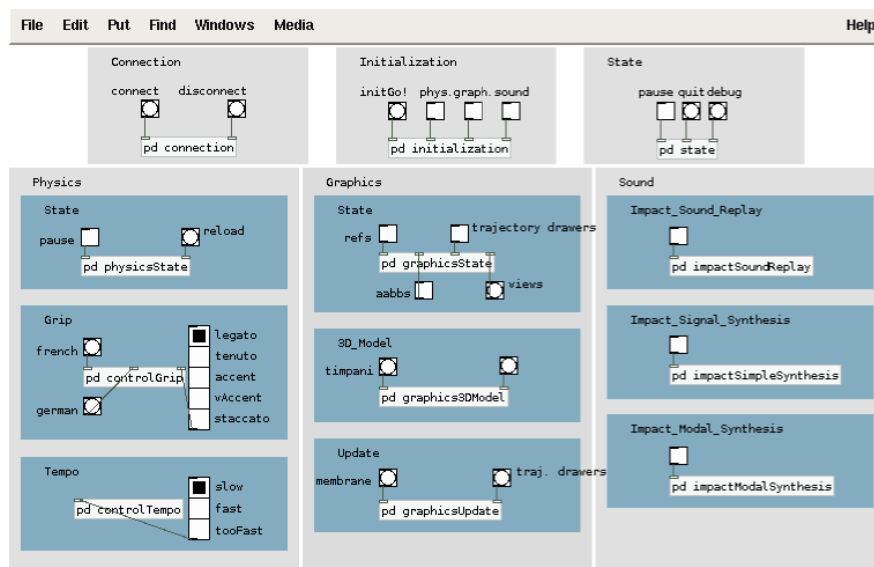
### 5.3.3 Results

Our software architecture makes use of the *OSC* protocol, without any assumption about the hosting of OSC clients and servers, making it possible to run the

graphics, physics and sound managers on distinct computers. This architecture has been successfully tested by running the graphics/physics and sound cores on two different platforms linked by an ethernet connexion, providing an effective and reliable communication between the two.

As for the sound synthesis system, a modal model of a drum membrane has been designed<sup>3</sup> using the *Modalys* framework [Adr89, Adr91, EBC05]. It allows the real-time parameterization of the membrane properties (size, mass, tension), as well as the parameters of the modal synthesis (number of modes, resonances), thus rendering varied sound feedback effects. The impact location and force on the drum membrane can also be parameterized, offering a real-time sound rendering in response to an instrumental gesture simulation.

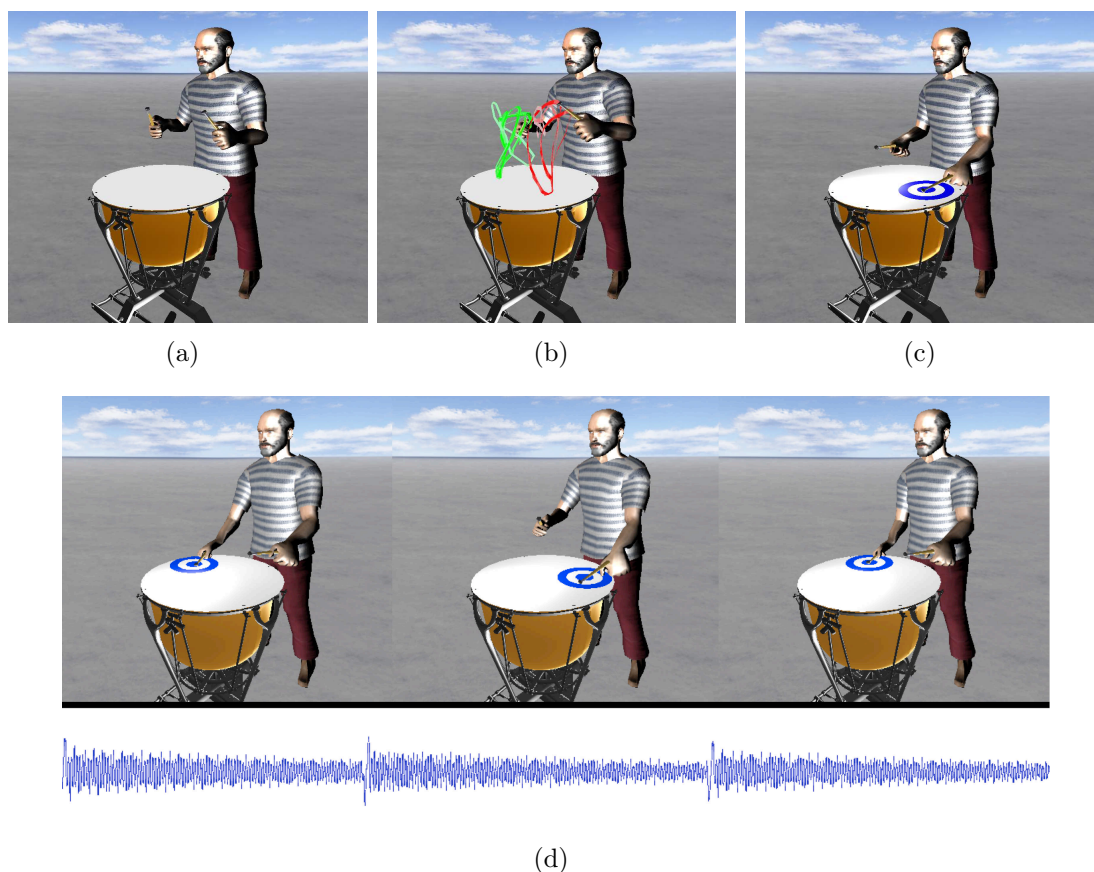
The user interface presented in FIG. 5.7 was implemented using Pure Data [Puc96], which is considered as the sound OSC server in our architecture. It shows how users can instantiate and access OSC components (top control panels) by modifying, creating and registering new interaction messages between the physics (left control panel), graphics (middle control panel) and sound (right control panel) managers.



**FIG. 5.7.** User interface: users can parameterize the percussion gesture to be simulated (drum grip, playing mode, tempo), as well as the graphics rendering and the sound feedback to be used (sound replay, signal-based and physically-based sound synthesis).

<sup>3</sup>Thanks to the help of Nicholas Ellis from *IRCAM*, who designed the overall modal model of the drum membrane.

Users can select different percussion grips (*French* or *German*), playing modes (*legato*, *tenuto*, *accent*, *vertical accent* and *staccato*), and different tempi, to be simulated by the virtual percussionist. The parameterization of the visual and sound feedback is also possible. During the simulation, the interface provides users with different sound synthesis models such as the simple playback of sound clips, signal-based sound synthesis, or physics-based sound (modal) synthesis. Every sound synthesis technique proposed by the interface can be tuned in real time. For instance for the modal sound synthesis module, one can tune the parameterization of the drum membrane physics properties (radius size, mass, tension). The available sound and visual feedback is presented in FIG. 5.8, which can be parameterized by users with the rendering of helpful visual cues such as mallet trajectories and beat impact locations of the drum membrane.



**FIG. 5.8.** Visual feedback during the simulation with (a) the final graphical model of the virtual percussionist. Users can explore and visualize the overall virtual percussion performance space, as well as (b) mallet trajectories and (c) beat impact locations. (d) Visual and sound renderings of beat impacts on the drum membrane.

## 5.4 Conclusion

We proposed in this chapter a physically-enabled environment in which a virtual percussionist can be physically controlled and interact with a physics-based sound synthesis model. The physics-based control from real percussion performances guarantees to maintain the main characteristics of human motion data while keeping the physical coherence of the interaction with the simulated instrument.

Furthermore, the hybrid control mode combining *IK* and *ID* controllers leads to a more intuitive and consistent way of editing the motion to be simulated only from mallet extremity trajectories, as regards to the analysis presented in chapter 4. This control mode is especially shown to be more accurate than traditional solutions for physically tracking motion capture data.

Moreover, the proposed asynchronous client-server architecture takes advantage of motion and sound physics formulations, generating in real-time virtual percussion performances that can be parameterized from the motion to the sound. The physical interaction of the virtual percussionist with a physics-based model of a timpani membrane is presented, leading to the control of a sound synthesis model from the actions (impacts) of the virtual percussionist.



# 6

## Musical Application and Evaluation

In this chapter, we propose a composition process for synthesizing novel (*i.e.* previously not recorded) virtual percussion performances, as well as its evaluation in a musical perspective. We namely present how new virtual percussion performances can be synthesized through the specification of scores at the gestural level, by the definition of gesture units that can be assembled and articulated. Furthermore, a musical evaluation of synthesized percussion exercises is conducted by submitting these latter to the evaluation of a professional percussionist.

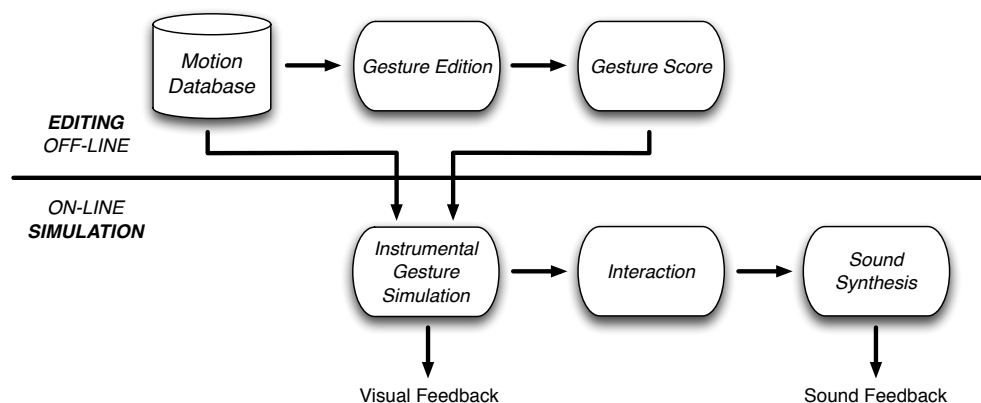
The chapter is organized as follows. Section 6.1 presents the adopted approach to the composition of novel virtual percussion performances at the gestural level, based on the assembling and articulation of gesture units. Synthesized percussion exercises as well as their musical evaluation by a professional percussionist are then presented in section 6.2. Section 6.3 then extensively discusses the advantages and limitations of our approach as regards to its musical perspectives. Finally, section 6.4 concludes this chapter with further perspectives.

### 6.1 Gesture Edition and Composition

The synthesis system presented in the previous chapter is used for creating novel percussion sequences from available motion data. We detail in this section the motivation and mechanisms involved during the edition of gesture scores that are simulated by our synthesis system. We consider this edition step as a first advance towards the composition of virtual music performances with our system.

The basis of the presented edition process is highly inspired from existing works in representing music-related materials. Since the beginning of the transcription of music performances, composers and performers have used the idea of representing sounds as independent and distinct events. The most obvious example of this idea is of course the transcription of notes on music scores. This event-based representation can be more or less suited depending on the considered





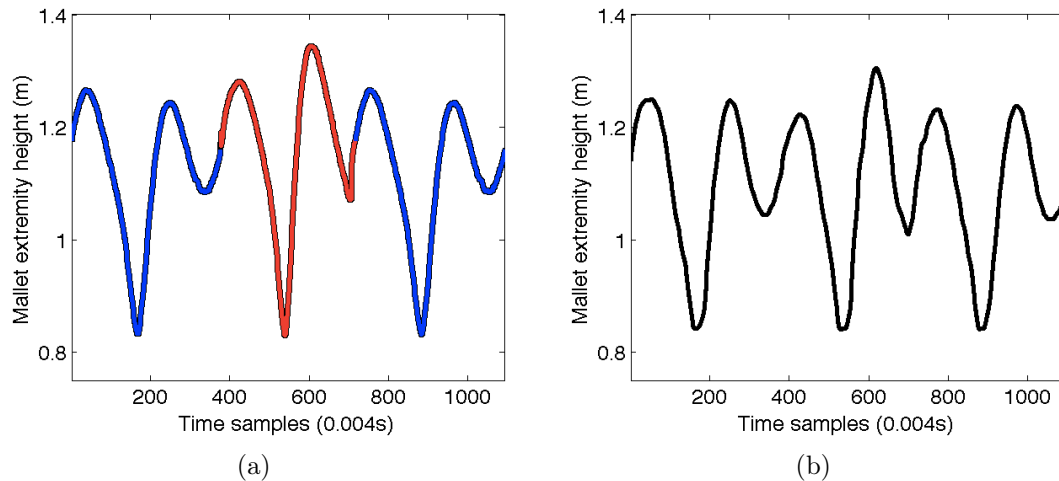
**FIG. 6.1.** Global approach to the composition process of virtual percussion performances.

music performance. For example, it is useful for representing sounds produced by instruments that involve distinct actions, such as striking or plucking. For other instruments involving continuous actions, such as a vibrating reed or bowing, such representation can also be useful by defining events with a start, an end, a state change. This is namely the approach that was adopted in the design of the widespread *MIDI* protocol [MID09].

Similarly, we define a set of events that can be used in our system for synthesizing percussion performances. However, the drastic difference from the previously detailed event representation is that our system uses the manipulation of elementary *gesture events*. In fact, these gesture events are constituted of motion signals coming directly from motion data presented in chapter 4. We will therefore refer to them as *gesture units*. FIG. 6.1 depicts the use of these gesture units in our synthesis system. The edition and assembling of these canonical gesture units leads to the specification of a *gesture score*, whose equivalent signal is then simulated by our system.

Our edition and composition process of gesture scores makes available gesture units of different types. These gesture units are directly related to the motion capture database, so that a huge amount of gesture units are available for representing the playing techniques under study (grips, playing modes, beat locations). We detail here how these gesture units are obtained, as well as how gesture scores are edited from these latter. This involves both an adequate segmentation of motion capture data, as well as an assembling (composition) of the resulting gesture units.

An interesting technical question here is the way the simulation step uses the edited gesture scores. The segmentation process discussed in section 4.3.1 leads



**FIG. 6.2.** Gesture edition and composition: (a) *beat-centered* gesture units for *legato* (blue) and *accent* (red) playing modes, (b) resulting simulation of the assembling of the two units.

to *beat-to-beat* gesture units for each of the playing techniques under study. It should be noted however that gesture scores involving such *beat-to-beat* gesture units are not usable in a simulation context. This segmentation creates two types of problem. The first one is related to the fact that linking two gesture units at the moment of the beat impact may create unwanted kinematic discontinuities, concerning the position and orientation (and their derivatives) of both mallet extremities and body segments. Another problem that can occur is the alteration of beat impact profiles. Linking two gesture units at the beat impact would lead to separate impact profiles in two phases, which may result in the modification of the action/reaction mechanical nature of the impact. Finally, placing the articulation point between two gesture units during the beat impact disagrees with the usual decomposition of instrumental gestures into preparatory, interaction and retraction gestures.

We therefore translate gesture units from *beat-to-beat* to *beat-centered* units. Handling the articulation between gesture units in such a way is more consistent since the articulation point is placed between the retraction phase of the previous gesture unit and the preparatory phase of the next one. However, even such a unit representation may lead to discontinuities in position and velocity between two gesture units. That is why we ensure a  $C^1$  continuity on drumstick extremity trajectories. Examples of *beat-centered* gesture units for *legato* and *accent* playing modes are presented in FIG. 6.2, with the corresponding simulation of the assembling of the two units.

Another issue to discuss concerning the edition of gesture scores is the way the gesture units are assembled to create a partition that our system will simulate. This includes two underlying questions, on the one hand which gesture unit will be chosen for representing a playing technique since many units are available, and on the other hand how the assembling is practically achieved?











The equivalent problem to the first question brings back, for a given playing technique, to the existence of a "best-suited" gesture unit among the available ones. This issue has also been under study in [Ram91] concerning piano fingering. Our viewpoint on this open question is to argue that performers are highly trained to replicate and adapt their gestures under a wide range of conditions. So that our solution is to pick up a particular gesture unit in a captured sequence, apart from the first and last one. This restriction ensures that no external disturbance is included in the gesture unit, as it was observed that performers tend to test the balance of their mallets at the beginning of a sequence, and exaggerate their motion when finishing it.

Once a set of gesture units is available for representing each playing technique, a strategy has to be developed for assembling these in a whole coherent gesture score. This includes again two underlying problems. The first one concerns the articulation between different position and orientation postures of the virtual character from the end of a gesture unit to the beginning of the next one. In all the percussion exercises presented in this chapter, this problem was achieved mainly by focusing only on the mallet extremity trajectories, as these latter are essential factors in percussion gestures as underlined in chapter 4. The choice of the gesture units was therefore achieved by a careful analysis of the height of mallet extremity trajectories so that a match can be found between the end of a gesture unit and the beginning of the next one. The animation engine then involves an interpolation process for transitioning from a gesture unit to another. The second problem concerns the timing issue when mixing heterogeneous gesture units. It has been easily overcome due to the high proficiency of performers in keeping a steady tempo.

## **6.2 Musical Evaluation of Virtual Percussion Performances**

One of the most interesting outcomes of such framework is the possibility of handling and assembling heterogeneous performances using a combination of a few percussion gesture units. Thanks to the physics simulation of the virtual performer, the issue of gesture articulation between movement units is partly addressed by the physics engine, thus leading to a more natural sequence of

TAB. 6.1. Timpani playing notation

<i>Legato</i>	<i>Tenuto</i>	<i>Accent</i>	<i>V.Accent</i>	<i>Staccato</i>
				
<i>One – Third</i>	<i>Center</i>	<i>Rim</i>	<i>Left</i>	<i>Right</i>
				

movements<sup>1</sup>.

To evaluate the proposed framework, we have simulated an extensive set of musical examples consisting of several timpani exercises<sup>2</sup>. The resulting simulations were evaluated by the last author (a University Percussion Professor and an active Performer)<sup>3</sup>, by focusing both on the visual and auditory feedback, first simultaneously and then separately. The exercises are divided in two main categories: validation and extrapolation exercises.

The first category, validation exercises, aims at verifying the accuracy with which the proposed framework can synthesize percussion movements (and sounds) similar to the captured movements in the database. They consist in independent exercises that explore variations on the type of grip (*French* or *German*), the type of playing mode (*legato*, *tenuto*, *accent*, *vertical accent* and *staccato*), as well as the position of the impact on the drum membrane (one-third, center or rim).

The second category, extrapolation exercises, consist in musical excerpts that go beyond those obtained with motion capture. Such exercises were not performed by the musicians and therefore no articulation data corresponding to these exercises are available. They typically mix various types of gestures and impact positions in one excerpt, as well as variations on the tempo of the performances.

<sup>1</sup>The resulting articulation is not necessarily equivalent to real performer techniques. It will nevertheless be a physically plausible solution to the problem.

<sup>2</sup>The exercises presented in this chapter are available at <http://www-valoria.univ-ubs.fr/Alexandre.Bouenard/index.php/Research/PhDThesis>

<sup>3</sup>Prof. F. Marandola (Schulich School of Music, McGill University) has chosen the percussion exercises to be simulated by the virtual character. He also qualitatively evaluated the synthesized percussion exercises.

Throughout this section, all percussion exercises exploit gesture units in the motion capture database described in section 4.2. The notation corresponding to timpani playing techniques is described in TAB. 6.1, and is used as the standard notation throughout the rest of this paper for describing percussion exercise scores.

### 6.2.1 General Comments

Although the simulations look and sound realistic, a first musical evaluation was achieved for determining the overall degree of expertise of the virtual performer. From the resulting simulations, we can see that the virtual timpanist’s performance is similar to that of a beginner/intermediate performer. Only part of the attacks were performed correctly, but with a wide range of variation in the impact locations and in the motion of the arm and forearm. Other issues in the simulations include the fact that the grip is not as realistic as expected. Moreover, the bottom part of the virtual character’s body does not move. This fact may be of little importance for a set of repeated notes at a steady tempo, but becomes more significant when variations in location, intensity, types of attack and rhythm occur.

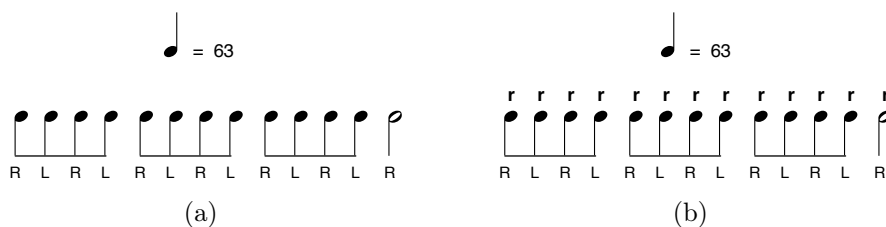
The setup described in section 4.2 actually does not include the recording of the subtleties occurring in the mallet grips, since no sensors are monitoring the grasp mechanisms involving the fingers and the mallets. Although our motion capture setup does capture the orientation of hands and mallets, this information is not sufficient for modeling the grasp subtleties that could enhance the realism of the resulting grip simulations. Furthermore, no balance strategies are involved in the simulation of the percussion exercises, although information on performers’ center of mass was captured in the motion capture sessions. This simulation choice explains the remark on the quite static bottom part of the virtual performer.

### 6.2.2 Validation Exercises

In this section we analyze how the system manages to connect independent gesture units taken from the timpani database, where each exercise consists in a sequence of similar types of strokes. The strokes are chosen from pre-selected gesture units, corresponding to a given grip, a playing mode and a location on the timpani. We focus on the qualitative evaluation of the simulation of playing modes as well as beat impact locations<sup>4</sup>.

---

<sup>4</sup>The simulation quality of *French* or *German* grips is further discussed in section 6.3.1.2.



**FIG. 6.3.** (a) Simulation exercise consisting of a sequence of *legato* strokes. (b) Simulation exercise consisting of a sequence of *legato* strokes played at the rim.

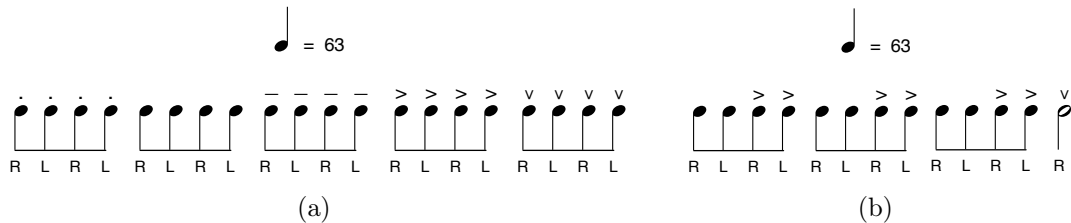
We simulated five exercises consisting of 13 strokes each (7 for the right hand and 6 for the left hand) for each playing mode. FIG. 6.4(a) shows an example of an exercise consisting only of *legato* strokes, whereas FIG. 6.4(b) shows the simulation score of beat impacts locations at the rim of the timpani membrane.

As expected, the resulting simulations show significant variations in the quality of the attacks for a given type of stroke, specifically with respect to the location of the impact point. However, the shoulders present very large movements; there is an excessive shoulder rotation for each single stroke, making the position of the elbows not accurate. In fact, given the steady position of the bottom part of the body, there should be more motion at the elbows.

The exaggeration in shoulder and elbow movements can be attributed to the physical model of the virtual percussionist. Just as any physical entity, the body parts of the virtual percussionist are characterized by inertia properties. When the physical model is put into motion, these body parts seem to acquire too much inertia, so that even if the overall motion (mallet extremities) is accurate enough compared to the recorded gesture units [BWG09b], an exaggeration of motion appears.

Considering only the sound generated, one can notice that the sound outcome is plausible and that playing modes producing them are audibly perceptible. However, the sound starts with too much attack (there is even some saturation at the beginning of the sound) and the resonance differs from a real timpani sound. Finally, the sound is not metallic enough when considering the attacks near the rim of the membrane.

These issues concern the *Modalys* model of the drum membrane. The lack of sound resonance is easily explained by the fact that the drum membrane model does not feature a resonance body. Moreover, the non-metallic nature of the resulting sounds may also come from a combination of imprecisions both from the drum membrane model itself since no metallic rim is included in the *Modalys* model, and from beat impact locations inaccuracies.



**FIG. 6.4.** (a) Simulation exercise consisting of five gesture units repeated four times: *staccato*, *legato*, *tenuto*, *accent* and *vertical accent*, respectively. (b) Simulation exercise consisting of variations between *legato* and *accent* strokes, ending with a *vertical accent*.

### 6.2.3 Extrapolation Exercises

The previous exercises consisted of simple sequences of similar strokes. Much more interesting though is to verify how the framework can deal with gestures combinations of different stroke types, tempos, and impact locations that were not initially recorded in sequence. Various combinations of gestures will be analyzed in the following subsections.

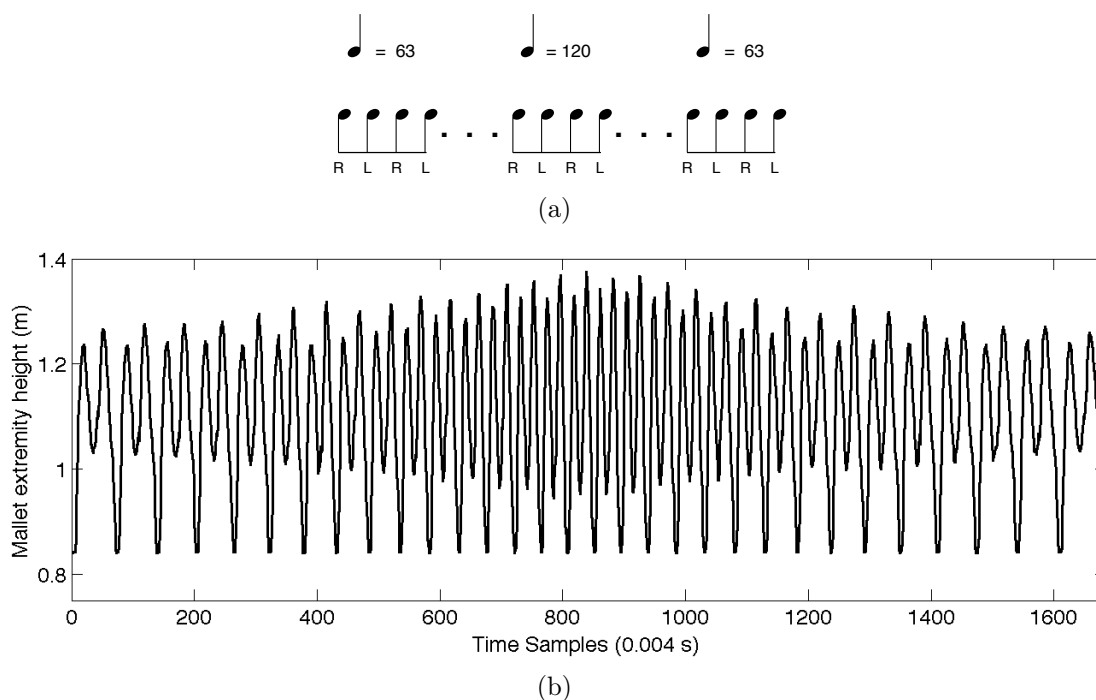
#### 6.2.3.1 Playing Modes

The following two exercises show a combination of gesture units for various playing modes. The idea is to show that the system can connect independent gesture units of each playing mode, as well as creating the articulation for a combination of these different units.

The first exercise (FIG. 6.5(a)) is a sequence of several strokes (respectively *staccato*, *legato*, *tenuto* and *vertical accent*) in groups of four strokes (two per arm). The second exercise (FIG. 6.5(b)) changes the type of stroke for each arm from *legato* to *accent* and back, finishing with a *vertical accent* performed with the right hand.

Considering the articulation between the simulated playing modes, it can be seen that articulation changes are perceptible throughout the simulation. However, problems arise in *legato-tenuto* and *tenuto-accent* pairs, so that differences between these are sometimes not very clear. These can be the result of the motion exaggeration discussed previously (cf. section 6.2.2), as well as from the fact that the effect of fingers is not available in the recorded motion data (cf. section 6.2.1).

The resulting sounds from the articulation between various playing modes were perceptible, except for the *legato-accent* sequences. This lack may result from both the *Modalys* model of the drum membrane and from the simulated



**FIG. 6.5.** (a) Simulation exercise consisting of an accelerando followed by a decelerando of *legato* strokes. Note that the timpani mocap database only has samples of *legato* strokes capture at one given tempo. (b) Simulation and articulation between *legato* beats under an accelerando-decelerando musical variation (height of the tip of the mallet).

articulation problem in this case as pointed out previously.

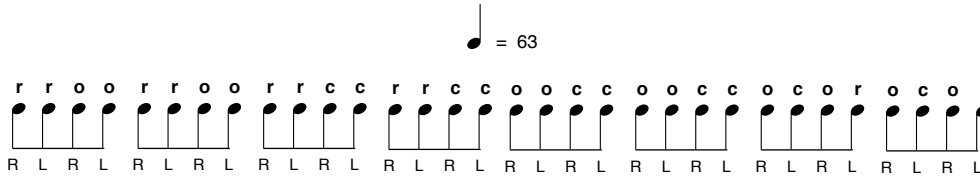
### 6.2.3.2 Tempo Variations

We have tested the simulation of articulating the same percussion gesture (*legato*) under an accelerando-decelerando musical variation<sup>5</sup>. FIG. 6.6(a) shows the percussion exercise simulated by the virtual percussionist.

Regarding the quality of the synthesized *legato* strokes under such tempo variation, it can be seen in the simulation that the motion of the shoulders increase with the speed (becoming way too wide), while the amplitude of the motion of the forearms and the space between them remain identical. In reality, timpanists would have a tendency to bring closer their forearms and to reduce the amplitude of the vertical movements of the forearms and mallets, compensating by increasing the motion of the wrists and an increased use of the fingers. Moreover, during

<sup>5</sup>This tempo variation has not been recorded initially.





**FIG. 6.6.** Simulation exercise consisting of several *legato* strokes played at various locations in the membrane (o: one-third, c: center and r: rim).

this *accelerando* phase, the flow becomes more irregular and the impact locations substantially change, causing inconsistency in the quality of the sounds produced.

FIG. 6.6(b) displays the motion amplitude increase (mallet extremity) along the *accelerando-decelerando* exercise shown in FIG. 6.6(a). This increase can be explained by a gain of inertia in the physical model of the virtual percussionist (cf. section 6.2.2).

The inconsistency reported on beat impact locations can be explained from a simulation point of view. As mentioned in section 5.2.2.2 and emphasized by equation 5.3, the control of the virtual percussionist is achieved by the control of errors between target and state mallet extremities. The *accelerando* phase seems to have a critical effect on the control of such errors, leading to the propagation of joint errors on the mallet extremity and on beat impact locations.

As for the quality of the corresponding synthesized sounds to this percussion exercise, during the *accelerando*, a *crescendo* occurs naturally by setting on resonance the membrane of timpani. It should be noticed that this feature is also used by real timpanists to build-up a *crescendo* during a roll: they increase the speed of their rolls to build up the *crescendo* while they tend to find the speed which brings out the maximum natural resonance from the timpani.

This fact attests to the natural response of the *Modalys* model of the drum membrane in response to the simulated *legato* strokes under an *accelerando* tempo variation. It shows moreover that the interaction model, and especially the collision detection of beat impacts, offers realistic features such as contact durations and forces in comparison to real timpani performances.

### 6.2.3.3 Impact Location Variations

Finally, we simulated the exercise depicted in FIG. 6.6, consisting in a sequence of several *legato* strokes played at various beat impact locations. This exercise aims to verify if the articulation of various beat impact locations results in still perceptible changes in the simulation.

In fact, the differences remain perceptible, more specifically with respect to the end of the exercise (two last groups of four strokes: o c o r). However, one

would expect that the differences would be more obvious among the three types of location, specifically for the centered strokes.

The reason for this effect is that the original recorded performances that were used to build and simulate this exercise were not performed accurately enough with respect to impact position. As shown in FIG. 6.7, the recorded beat impact locations (hollow signs) for the center location are too near to those corresponding to the one-third location.

## 6.3 Discussion: Advantages and Limitations

Simulating instrumental percussion gestures for controlling sound synthesis processes presents advantages and limitations, and these are of different orders when considering each module of our framework. In the following sections we will analyze in detail some of the sources of variability that are produced by our simulation system and may be at the origin of unwanted artifacts, as well as new interesting possibilities. We therefore consider sequentially the advantages and limitations of each module of our framework, considering first the instrumental gesture simulation and then the gesture edition step (cf. FIG. 6.1).

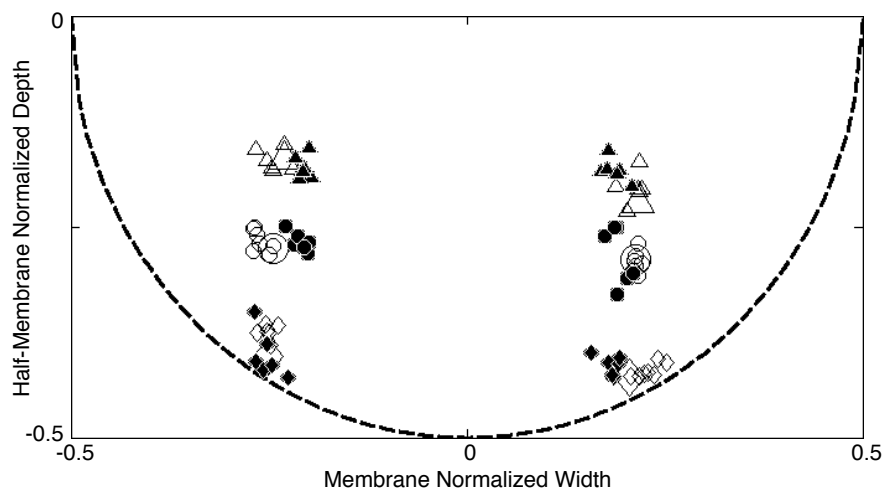
### 6.3.1 Instrumental Gesture Simulation

We here underline the importance of the parameterization of the virtual character during the simulation. A non-optimal parameterization can lead to artifacts in the resulting synthesis of percussion performances, as well as to effects on synthesized sounds. We also highlight the interest of simulating instrumental gestures with respect to the possible modulation of synthesized gestures while preserving the initial style and expressive characteristics of real percussion performances.

#### 6.3.1.1 Advantages

From a simulation point of view, once an acceptable parameterization of the simulation is achieved, associating a motion capture database and the physics-based synthesis of instrumental gestures yields an accurate simulation of movements.

A first advantage of the our physics-based framework for simulating percussion gestures is the retrieval of the dynamical aspects (beat impacts) of the recorded instrumental gestures, since no extra sensors have been used during the recording of percussion performances and no information concerning beat impact durations and forces has been recorded. Such physical approach is also of great interest for exploring the interaction of percussion gestures with sound synthesis since it makes available useful contact information that can be exploited by sound synthesis processes.



**FIG. 6.7.** Comparison of captured and simulated beat impact locations played *legato*. Small hollow signs represent strokes from motion capture data, whereas large hollow signs represent beat impact locations resulting from the chosen gesture units. Bold signs represent the simulated beat impact locations.

A second advantage of our system is its virtual nature. Indeed, the proposed framework allows the modeling of realistic phenomena occurring during real percussion performances. Furthermore, it also allows to go beyond reality, and provides the possibility of modifying the parameters of the virtual environment as wanted for creating new percussion performances. Among these parameters, it includes modifications of the physical model of the virtual percussionist, leading for instance to various stiffness properties of the synthesized gestures. Different interaction schemes between synthesized gestures and sound synthesis inputs can also be explored.

From a more musical point of view, an issue to take into account is the small variations in movements that are important to produce expressive performances. Indeed, deadpan performances where movements are kept as close to an ideal movement target as possible are considered mechanical and non expressive. By choosing a single gesture unit for representing each playing mode in the various simulations, we might actually expect to eliminate the natural variability of movements from the original performance by the (real) percussionists. This choice, although greatly simplifying the simulation, has the downside of potentially creating mechanical performances.

Nevertheless, thanks to the physics simulation layer, departing from a single gesture unit for each gesture type will nevertheless produce synthesized gestures with inherent variability due to the dynamic characteristic of the virtual charac-

ter. This micro variability can be explained by a combination of effects occurring at each level of our framework. As explained below, for the instrumental gesture simulation, numerical drifts can occur but still be controlled. Exploiting such drifts for slightly modifying the instrumental gesture may produce some form of expressiveness. For instance, single simulations of the same gesture unit yield slightly different yet consistent synthesized gesture units, as shown in FIG. 6.2. The propagation (and not the amplification) of such numerical drifts is also influenced by the edition and the articulation of gesture units, thus leading to another source of variability during the simulation of a sequence of several gesture units. These small differences in gestural trajectories therefore lead to variations in the impact locations on the membrane, thus producing variations in corresponding synthesized sound. These different effects may be explored musically, and actually lead to non-mechanical performances by the virtual percussionist.

An example is presented in FIG. 6.7, showing the simulation of several *legato* strokes for the *French* grip. Small hollow signs represent the 2D positions of motion captured strokes on the drum membrane respectively for the three impact positions. The three large hollow signs indicate the impact positions of the three captured strokes chosen as gesture units for each hand. Finally, the bold signs represent the various simulated strokes. One can note the comparable variability of the (real) percussionist's strokes from motion capture (small hollow signs), and the resulting variability of the synthesized strokes (small bold signs) from multiple simulations of each gesture unit.

### 6.3.1.2 Limitations

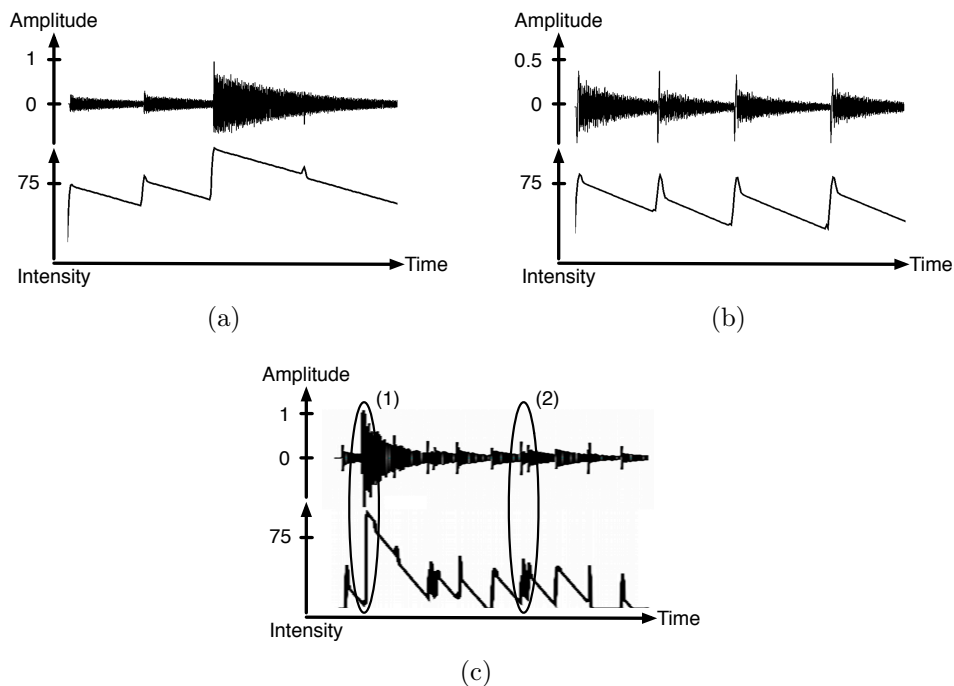
Depending on the sequence and speed of the selected gestures, as well as to the fine tuning of the mechanical joints (equation 5.3), simulation artifacts can appear and result in large variations in beat impact positions and forces. These variations are due to the adaptation of the physics simulation to the constraints in the movement data, and may yield unexpected sound phenomena.

An example is presented in FIG. 6.8, where the virtual musician performs a sequence of four beats played *legato* with different sets of physical constraints (mechanical joints). In FIG. 6.9(a), a simulation artifact can be observed on the third beat, that results in a larger sound waveform amplitude and sound intensity, actually masking the last beat. This artifact is removed by changing the physics constraints expressed by the damping and stiffness coefficients involved in equation 5.3, as shown in FIG. 6.9(b). In this latter case, one can see that the four beats are performed similarly, both in terms of amplitude and intensity<sup>6</sup>.

---

<sup>6</sup>Sounds in both simulations have been obtained using *Modalys*.

The parameterization of the mass and density of the different segments composing the physical model of the virtual percussionist should be chosen carefully, so that the inertial behavior of the system is comparable to that of humans. Otherwise accelerations or slowdowns as well as unexpected overruns may be obtained, leading to unrealistic simulations.



**FIG. 6.8.** (a) A simulation artifact causes much larger sound waveform amplitude and sound intensity on the third beat, compared to (b) a stable simulation generating fairly constant amplitude and intensity for all beats. (c) Issues (1) and (2) discussed below: (1) artifact when two close beat impacts are glued together by the sound synthesis engine to form a huge sound, and issue (2) when two close beat impacts are finally heard.

Moreover, variations in the simulations of the two percussion grips can be found in the assembly of more complex gesture sequences. *French* grip simulations work well, due to the large flexibility and smoothness of movement which lead to large amplitudes in the trajectories deployed by the human performer (cf. FIG. 6.3(b)). The simulation framework is capable of following the original gesture trajectories and can articulate the sequence of gestures. For the *German* grip though, due to the large motion-phase where the performer keeps the mallet relatively immobilized close to the membrane, the physics engine results are less satisfactory. This is mostly seen in sequences of gestures assembled from gesture units. The less successful simulation of *German* grip can be due to two reasons.

The first reason is that our motion data does not take into account the fact that expert percussionists commonly use their fingers to help performing the strokes. This is true for both *French* and *German* grips, but the effect of this technique in the *German* grip can be more important. The use of the *German* grip involves generally a technique requiring more control from the wrist and fingers and less amplitude of the arms and forearms, in order to control the speed of the mallet as well as the duration of contact between the mallet and the timpani membrane.

The second reason reveals the fact that the mallets (and therefore the percussionists' arms) stay relatively immobile during a large portion of the gesture unit (cf. FIG. 4.8). It necessitates the simulation of fixed and more rigid postures that especially involve the tuning of the physical parameters with high stiffness values, thus potentially leading to instabilities in the numerical simulation.

FIG. 6.9(c) shows the effect of this issue on the resulting sound. The virtual performer is given a task to perform a sequence of *accent* strokes using captured data of the *German* grip. From the resulting sounds, one can notice that at some moments, (1) the sound amplitude is largely superior to the average, and (2) the resulting beats are not regularly spaced, although no dynamics or tempo variation was present in the score. Issue (1) characterizes a larger stroke seen in the figure as an artifact from the sound synthesis generation, *i.e.* the simulated gesture presents strokes that are very close temporally and spatially. The collision detection events, performed in a short lapse time, are then interpreted by the *Modalys* system which "glues" them together as a single yet much stronger stroke. As for issue (2), it can be explained by the limitations of both the capture database and the simulation framework, as discussed above.

### 6.3.2 Motion Database and Gesture Edition

Our framework depends highly on the motion clips initially recorded during the acquisition process of motion data. The motion database associated with edition processes are off-line components that are of particular importance when considering the composition of a gesture score. In this section we examine the assets and drawbacks of our system with respect to these components. We initially highlight the advantages of our framework in considering the simulation of percussion performances at the gestural level. We then discuss the completeness of the available motion capture database, as well as the editing method presented in this paper.

### **6.3.2.1 Advantages**

In this work, we adopted a sketching approach to the composition of gesture scores made of elementary gesture units. This constitutes an intuitive process where the user only needs to select and assemble a limited amount of canonical gesture units, one for representing each timpani playing mode. The gesture score can then be seen as copy-and-paste and concatenation processes of these gesture units to form a whole gesture score.

Although the presented model does not take into account the subtle articulation mechanisms occurring in real percussion performances, a simple articulation (interpolation) of gesture units is implicitly achieved by the internal physics simulation when the gesture score simulation switches from a gesture unit to another. Such a "sketching" approach has the advantage to be transparent to the user.

Another advantage of our framework in considering the composition at the gestural level is the preservation of the expressive features inherent to real percussion performances. As the composition process of gesture scores relies highly on real motion data, this insures that simulations will preserve the expressive components of the gesture units used in the gesture score. More interestingly, it allows consequently to preserve the style of the real performer in the simulated percussion performances. Specifically in this work, all the exercises presented in section 6.2 have used motion data from one performer, such that these simulated exercises can be considered as supplementary exercises that we might have asked the percussionist to perform.

The presented motion capture database in section 4.2 focuses on a particular excerpt (grip, playing mode and beat location) among the multitude of percussion techniques used by timpani performers. Our system provides the interesting possibility of enriching this limited database by the editing and simulation of novel percussion performances. One can also consider our approach as a means of preventing the tedious and time-consuming task of capturing an exhaustive motion database containing all the possible combinations of playing techniques. Finally, our framework provides an interactive tool that can be used in the exploration of the relationship between instrumental gestures and sound processes.

### **6.3.2.2 Limitations**

The motion capture database presented in section 4.2 focuses on specific set of timpani playing techniques, i.e. it provides gesture examples only for these playing techniques. As seen before, this fact limits the possible simulations to be performed with the proposed system.

Another limiting issue is the choice of the gesture units used during the editing step. As already discussed in the beginning of this chapter, such issue yields inevitably to the question of the existence of a best-suited gesture unit, compared to other examples of the same playing mode. If such "best" gesture unit exists, how can one identify it among the numerous replicates available in the motion database? Otherwise, does it make sense to average multiple gesture units so that a "mean" gesture unit can be used? We believe that using an average gesture might lead to the loss of some expressive features implicitly contained in the gesture signal. We therefore chose one specific gesture example within the sequence, without trying to find an optimization criteria for selecting this gesture unit.

A final issue to discuss related to the edition process is the concatenation of gesture units, which is the basis of the composition process. The gesture composition presented in this chapter is only based on a sketching process for composing gesture scores, therefore demonstrating the feasibility of the approach. It could be improved by considering higher-level planning processes responsible for the sequencing of gestures. In real percussion playing, high-level control mechanisms are indeed in charge of the articulation of playing modes, depending on many factors such as playing style, tempo variations as well as expressive nuances. Our concatenation model typically does not take such mechanisms into account.

## 6.4 Conclusion

This chapter presented a methodology for composing percussion exercises at the gestural level. This implies a sketching approach for assembling gesture units, leading to the synthesis of percussion exercises that were not initially recorded. The simulated exercises have explored both the composition of validation and extrapolation exercises, that were qualitatively evaluated by a percussion professor.

Validation exercises consist in assembling intra-gesture units, *i.e.* the composition and simulation of gesture units of the same type. The evaluation of these exercises has underlined the accuracy and naturalness of the resulting simulations, both in terms of playing modes and impact locations. It was also pointed out some artifacts, such as gesture exaggeration and sound inaccuracies, that may be explained both in terms of the gesture simulation and sound models.

Extrapolation exercises have explored the articulation and simulation of various playing modes that were not recorded initially during the collection of motion capture data. The evaluation underlined the quality of the simulated articulation between these playing modes, as well as for the resulting sounds. Problems in these simulations were mainly caused by the lack of completeness of the motion



capture database.

# 7

## Conclusion and Future Work

### 7.1 Conclusion

This dissertation has proposed a system for synthesizing percussion performances in which a virtual percussionist controls sound synthesis processes. It includes the analysis of percussion motion that has then oriented the design of a physics-based model for controlling and animating a virtual percussionist. We have also presented a physics-based model of a drum membrane that can be controlled by the actions of the animated percussionist, as well as an architecture for easing the real-time interaction between gesture and sound simulations. We have finally conducted an informal evaluation of the musical possibilities of the proposed system.

We first propose (chapter 4) a protocol for capturing and analysing percussion (timpani) performances. The resulting motion database makes available data for various timpani playing techniques, such as mallet grips (*French* and *German*), various playing modes (*legato*, *tenuto*, *accent*, *vertical accent* and *staccato*) as well as different beat impact locations (*one-third*, *center* and *rim*).

The analysis of the motion data focuses especially on percussion grips and playing modes, as a mean of highlighting the importance of using mallet extremity trajectories to conduct the synthesis of timpani performances. The analysis methodology has consisted in the extraction of motion parameters for representing the percussion playing techniques under study. This has led to the determination of a set of parameters solely regarding mallet extremity trajectories, both for the discrimination of percussion grips and playing modes. The evaluation of such a parameterization is proved to be consistent, as attested by the high recognition rates obtained through a classification/recognition scheme.

We present in chapter 5 a physically-enabled environment in which a virtual percussionist can be physically controlled and interact with a physics-based model of a drum membrane.

The physics-based control from percussion motion data guarantees to maintain the main characteristics of human motion data while keeping the physical coherence of the interaction with the simulated instrument. More specifically, we have presented a hybrid control mode combining *IK* and *ID* controllers that leads to a more intuitive and consistent way of editing the motion to be simulated only from mallet extremity trajectories.

The proposed asynchronous client-server architecture of our system takes advantage of motion and sound physics formulations, generating in real-time virtual percussion performances that can be parameterized from the motion to the sound.

Chapter 6 finally explores the musical possibilities of the proposed system by the simulation of two types of percussion exercises: validation and extrapolation exercises. These percussion exercises have been evaluated by a percussion professor.

Validation exercises analyse the articulation and simulation of gesture units of the same type. The evaluation of these exercises has underlined the accuracy and naturalness of the resulting simulations, both in terms of playing modes and impact locations

Extrapolation exercises explore the articulation and simulation of heterogeneous gesture units, exercises that were not recorded initially during the collection of motion capture data. The evaluation underlines the quality of the simulated articulation between these playing modes, as well as for the resulting sounds.

## 7.2 Future Work

Regarding future work, there are many avenues for extending the work presented in this dissertation. These perspectives concern the analysis of additional percussion playing strategies, as well as the development of other synthesis models to improve the overall realism of the resulting percussion simulations. Such supplementary works would finally be of great interest for proposing enhancements to the musical possibilities of the presented system.

### 7.2.1 Analysis

The analysis of percussion motion data presented in this dissertation has particularly focused on grips and playing modes. This analysis might take into account a larger set of timpani playing variations.

Among these additional playing techniques, gesture dynamics such as *pp*, *mf* and *ff* as well as tempo variations could be of great interest. The study

of these playing variations could lead to identify the motion mechanisms that are involved in real percussion performances for articulating gesture features of heterogenous natures. For instance, such work has been performed in [Ras08] and have underlined the effect of articulation of bowed-string movements under various playing conditions.

Another specificity related to timpani performance is the fact that timpani performers usually play on a drum set composed of several timpani. The capture and study of percussion performances where timpanists are playing several instruments could lead to the identification of motion parameters characterizing the switching from a timpani to another, as well as balance strategies.

Other features related to percussion performances could be captured as well. Among them, the capture of finger motion data could be of interest for analysing the effect of percussion grips (*French* or *German*) on the subtle grasp mechanisms occuring in the finger-mallet system. These finger motion data could be of different natures, either purely kinematic (orientation and position) or dynamic by the use of pressure sensors for example.

Finally, dynamic features that are not taken into account in our work is the capture of beat impact forces. This could lead to the identification of gesture-sound couplings, and could be useful for comparing real mechanisms to those simulated by our system.

### 7.2.2 Synthesis

As our synthesis system is strongly related to motion capture data, the integration of additional models could benefit from the availability of a larger set of timpani playing conditions. Other models could also be involved for enhancing the realism of the resulting percussion performances.

The availability of motion data regarding both grip mechanisms and the switching from a timpani instrument to another leaves room to integrate additional models in our system.

Motion data and analysis of mallet grasp strategies could be involved in the design of a physics-based model of a hand. Related achievements regarding the simulation of grasp mechanisms [Kry05, KP06] appear promising for applying such strategies to the case of mallet grasps.

Moreover, motion data concerning the switching from timpani instruments to another could be used for design balance controllers in our synthesis framework. Such balance controllers should particularly take care of the orientation of the body as well as the beat impact location as regards to the timpani instrument to be played. The simulation of such task-oriented balance strategies would show

advances regarding motor tasks that are usually not addressed, and could be inspired by recent works [MZS09, SCCH09].

Integrating a physics-based hand model in our synthesis system might enhance the realism of the resulting percussion simulations. More generally, proposing a more accurate physics-model of the virtual character could have great benefits to that mean. Especially, one research problem that has not been addressed in this dissertation is the tuning of the mechanical joints composing the virtual character model. At the moment, there is no reliable and automatic method for tuning such mechanical joints apart from heuristic techniques [ZH02] or analytical solutions only valid for upper-body motion [ACSF07]. Apart from the interest of solving this research question for the *Computer Animation* community, providing a method for automatically tuning mechanical joints could be of great interest in a musical perspective. It could indeed propose a dynamic interpretation of the differences in posture control strategies occurring in percussion performances. For instance, such automatic determination could be an interpretation of the stiffness nature of *German*-related performances compared to *French*-related ones mentioned in chapters 4 and 6.

Other further developments that could improve the overall realism of the simulated percussion performances are related to the sound synthesis part of this work. The drum model does not feature a resonance board, so that the resulting synthesized sounds are characterized by a somewhat metallic nature. Future works include therefore the modeling of a resonance board associated to the drum membrane model. This would involve the modeling of the radiation of the pressure load inside the resonance board after a beat impact on the drum membrane, as it has been shown that this is a preponderant mechanism of energy loss in the case of timpani [Ros82].

Finally, our system simulates the control of sound synthesis processes by the actions of the animated virtual percussionist. A future avenue would consist in moving from this *control* mechanism to an extended *interaction* scheme in which the sound can influence the gesture simulation. Such interaction scheme would involve the simulation of the virtual performer actions (gestures) as regards to the tension and vibration properties of the timpani model.

### 7.2.3 Musical Applications

In the previous sections, we detailed how this dissertation can be extended, regarding both the analysis and synthesis of percussion performances. Such future works may improve the musical application possibilities of our system, especially concerning two main usecases that it provides to users. The first application is based on the proposition of pedagogical and composition tools, whereas the sec-

ond involves such synthesis system during live performances.

Our system is based on a motion capture database that takes into account many playing conditions, such as percussion grips, playing modes and dynamics, as well as beat locations. These constitute a set of playing conditions that percussion professors could show to students for demonstrating percussion techniques. Such application is motivated by the fact that percussion professors have no other choice for demonstrating a technique than performing it and then judge students' aptitude in mimicing it. Such pedagogical tool combined with an analysis of motion data as proposed in chapter 4 can be used for demonstrating percussion techniques, as well as pointing out which motion features the students have to focus on.

Moreover, the availability of heterogenous motion data from percussion performances can also be used as a composition process. It has been illustrated in chapter 6 by the definition of gesture scores that can be generated by our system. A fundamental need for this composition application would be a user interface through which users could select and articulate the gesture units of their choice, as well as modifying the tempo, gesture dynamics and beat locations. The motion parameters extracted from mallet extremity trajectories (chapter 4) could also be an intuitive motion representation for making available to users the choice to specify their own gesture units. Users might manipulate such patterns characterizing mallet trajectories that could then be used for automatically reconstructing the whole mallet trajectories.

The evaluation of the motion parameters involved in chapter 4 can be seen as first theoretical attempt to provide recognition mechanisms of real percussion playing conditions. We have shown that it is possible to recognize with a high confidence percussion grips as well as playing modes. A future avenue of the analysis of percussion playing techniques is to extend the identification of parameters and recognition mechanisms to a larger set of playing conditions (beat impact locations, gesture dynamics). It should be noted that in chapter 4 we have only focused on motion parameters for discriminating between playing variations. Sound parameters could also be used to that mean. Once the identification of adequate parameters will be achieved, such live recognition-interaction application involves two technological obstacles. The first one is to provide recognition mechanisms that can be processed in *real-time*. A second issue is the specification of interaction rules between real/virtual performers. Such rules could be inspired by research works related to the automatic music composition and interactive performance areas.



# A

## Sound Synthesis Models

### A.1 d'Alembert Equation

FIG. A.1 represents the vertical displacement  $z$  of a vibrating string of length  $L$ , depending on its longitudinal displacement  $x$  over time and the application of a tension  $T$ . The limit conditions are:  $z(0) = z(L) = 0$ .

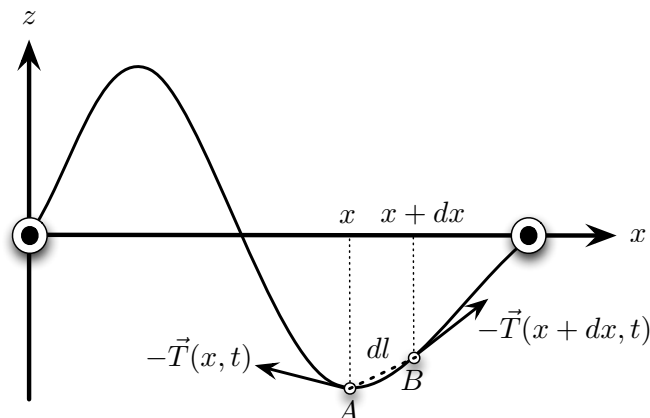


FIG. A.1. d'Alembert formulation of the vibrating string.

The d'Alembert equation make two major assumptions:

- The displacement of the string is mostly vertical and negligible compared to the longitudinal displacement, *i.e.*  $z \ll x$  and consequently  $\frac{\partial z}{\partial x} \ll 1$
- The gravity force applied on the string is negligible

With these assumptions, the infinitesimal length  $dl$  between  $A$  and  $B$  is such that:

$$dl^2 = dx^2 + [z(x + dx, t) - z(x, t)]^2 = dx^2 \cdot [1 + (\frac{\partial z}{\partial x})^2] \sim_{z \ll x} dx^2$$



Applying Newton's principle on the infinitesimal string length  $dl$  leads to the system:

$$T_x(x + dx, t) - T_x(x, t) = 0 \quad (\text{A.1})$$

$$T_z(x + dx, t) - T_z(x, t) = \mu \cdot dx \cdot \frac{\partial^2 z}{\partial t^2}(x, t) \quad (\text{A.2})$$

EQU. (A.1) expresses the independance and constance of the longitudinal component of  $T$  towards the longitudinal displacement  $x$  and time  $t$ , so that we can deduce from the boundary conditions:  $\forall x, \forall t T_x(x, t) = T$ . The tangency of the tension induces also:  $\frac{T_z}{T_x} = \frac{\partial z}{\partial x}(x, t)$ , so that  $T_z(x, t) = T \cdot \frac{\partial z}{\partial x}(x, t)$ .

The substitution of the latter expressions in EQU. (A.2) leads to:

$$T_z(x + dx, t) - T_z(x, t) = T \cdot dx \cdot \frac{\partial^2 z}{\partial x^2}(x, t) = \mu \cdot dx \cdot \frac{\partial^2 z}{\partial t^2}(x, t) \quad (\text{A.3})$$

From EQU. (A.3), we deduce the d'Alembert formulation enounced in EQU. (2.16).

## A.2 d'Alembert Equation: Fourier's Solution

Fourier proposed a general solution to d'Alembert equation of the vibrating string problem. It involves a particular form of the solution as a stationnary wave, such that  $z(x, t) = z_1(t) \cdot z_2(x)$ .

The substitution of such solution form in EQU. (2.16) leads to:

$$\frac{\ddot{z}_2}{z_2}(x) = \frac{1}{c^2} \cdot \frac{\ddot{z}_1}{z_1}(t) \text{ with } c^2 = \frac{T}{\mu} \quad (\text{A.4})$$

Due to the dependancy of the two members of EQU. (A.4) to two separable variables (space and time), these terms are inevitably equal to a real constant  $a$ . From that observation we can deduce a general form of the function  $z_1$  and  $z_2$ , with  $k_i = \frac{i\pi}{L}$ , and  $\phi_i$  depending on initial conditions:

$$\begin{aligned} \ddot{z}_1(t) = a \cdot c^2 \cdot z_1(t) &\Rightarrow z_1(t) = a_1 \cdot \sin(k_i \cdot c \cdot t + \phi_i) \\ \ddot{z}_2(x) = a \cdot z_2(x) &\Rightarrow z_2(x) = a_2 \cdot \sin(k_i \cdot x) \end{aligned}$$

The solution can then be written as a sum of *normal modes*  $z_i$  characterizing the allowed modes of deformation of the string:

$$z(x, t) = \sum_{i=1}^{\infty} z_i(x, t) \text{ with } z_i(x, t) = a'_i \cdot \sin(k_i \cdot c \cdot t + \phi_i) \cdot \sin(k_i \cdot x)$$

These normal modes can also be rewritten as:

$$z_i(x, t) = a_i. [\cos(k_i.(c.t - x) + \phi_i) - \cos(k_i.(c.t + x) + \phi_i)] \quad (\text{A.5})$$

EQU. (A.5) shows the form of the solution intuited by d'Alembert, composed of two travelling waves in opposite longitudinal directions, EQU. (2.17).

### A.3 Modal Decomposition and d'Alembert Equation

We consider here the simplest modal decomposition of a vibrating string of length  $L$  composed of  $n$  identical masses ( $m$ ) linked by identical springs ( $k$ ), FIG. A.2. This representation makes the assumption that every mass is spaced of a distance  $h$ , that is negligible, *i.e.*  $h \ll 1$  for continuity purpose and for treating the string as a continuous material.

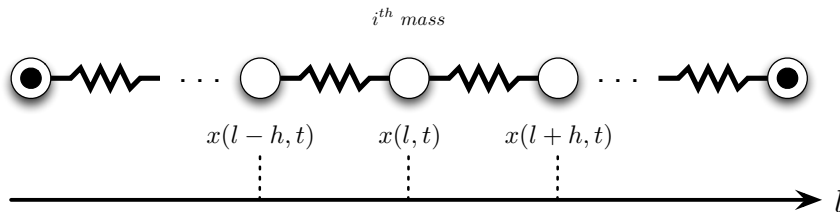


FIG. A.2. Modal decomposition of the vibrating string.

Applying Newton's principle on the  $i^{\text{th}}$  mass leads to the system, where the contributions of the previous and next masses are derived by the Hookes law:

$$\begin{aligned}
 m \cdot \frac{\partial^2 x}{\partial t^2}(l, t) &= T_{l-h}(l, t) + T_{l+h}(l, t) & (\text{A.6}) \\
 T_{l-h}(l, t) &= -k \cdot [x(l, t) - x(l-h, t)] \\
 T_{l+h}(l, t) &= k \cdot [x(l+h, t) - x(l, t)]
 \end{aligned}$$

The continuity assumption allows the derivation of the following Taylor limited developments:

$$\begin{aligned}
 x(l-h, t) &\sim_{h \ll 1} x(l, t) - h \cdot \frac{\partial x}{\partial l} + \frac{h^2}{2} \cdot \frac{\partial^2 x}{\partial l^2} \\
 x(l+h, t) &\sim_{h \ll 1} x(l, t) + h \cdot \frac{\partial x}{\partial l} + \frac{h^2}{2} \cdot \frac{\partial^2 x}{\partial l^2}
 \end{aligned}$$

The substitution of the latter expressions into EQU. (A.6) leads finally to:

$$\frac{\partial^2 x}{\partial t^2}(l, t) - \frac{1}{c^2} \cdot \frac{\partial^2 x}{\partial l^2}(l, t) = 0 \text{ with } c^2 = \frac{k \cdot h^2}{m} \quad (\text{A.7})$$

Note that the dimension of the constant  $c$  is the dimension of a velocity. Furthermore, the modal decomposition of the spring can be considered as whole spring by observing:

- the mass of the system is equally distributed, such that we can consider a lineic mass  $m_l = \frac{m}{h}$

---

### A.3. Modal Decomposition and d'Alembert Equation

---

- each spring is characterized by a constant stiffness  $k$  at its steady distance  $h$ . Knowing that the stiffness of a spring is inversely proportional to its steady distance, we can express the stiffness of the whole spring as  $k = \frac{k_l}{h}$  where  $k_l$  is a characteristic of the equivalent spring (note however that the dimension of  $k_l$  is the dimension of a force [N] and not a stiffness).

We can eventually write the equation EQU. (A.7) that governs the motion of the string as follows:

$$\frac{\partial^2 x}{\partial l^2}(l, t) - \frac{1}{c^2} \cdot \frac{\partial^2 x}{\partial t^2}(l, t) = 0 \text{ with } c^2 = \frac{k_l}{m_l} \quad (\text{A.8})$$

EQU. (A.8) is analogously to EQU. (2.16) of the form of the d'Alembert equation, justifying the use of such mechanical model for the problem of the vibration string.

## A.4 Finite-Difference and Finite-Element Formulations

### Finite-Difference

The finite-difference formulation proceed to a discretization of EQU. (2.16) both in space and time. Given a fixed space step  $\delta x$  such that  $L = X_s \delta x$ , and a fixed time step  $\delta t$  such that  $T = T_s \delta t$ , a second order and explicit scheme leads to the following Taylor limited development of the partial derivatives:

$$\begin{aligned} \frac{\delta x^2}{2} \cdot \frac{\partial^2 z}{\partial x^2} &\sim_{\delta x \ll 1} z(x - \delta x, t) - z(x, t) + \delta x \cdot \frac{\partial z}{\partial x} \\ \frac{\delta x^2}{2} \cdot \frac{\partial^2 z}{\partial x^2} &\sim_{\delta x \ll 1} z(x + \delta x, t) - z(x, t) - \delta x \cdot \frac{\partial z}{\partial x} \\ \frac{\delta t^2}{2} \cdot \frac{\partial^2 z}{\partial t^2} &\sim_{\delta t \ll 1} z(z, t - \delta t) - z(x, t) + \delta t \cdot \frac{\partial z}{\partial t} \\ \frac{\delta t^2}{2} \cdot \frac{\partial^2 z}{\partial t^2} &\sim_{\delta t \ll 1} z(z, t + \delta t) - z(x, t) - \delta t \cdot \frac{\partial z}{\partial t} \end{aligned}$$

Note that a continuity assumption is again made on the function  $z$  for writing these Taylor series; the partial derivatives involved in the d'Alembert equation can also been expressed such that:

$$\begin{aligned} \frac{\partial^2 z}{\partial x^2} &\sim_{\delta x \ll 1} \frac{z(x+\delta x, t) - 2 \cdot z(x, t) + z(x-\delta x, t)}{\delta x^2} \\ \frac{\partial^2 z}{\partial t^2} &\sim_{\delta t \ll 1} \frac{z(x, t+\delta t) - 2 \cdot z(x, t) + z(x, t-\delta t)}{\delta t^2} \end{aligned}$$

Solving the discretized problem is then equivalent to:

$$\begin{aligned} \forall t \in [1, T_s - 1], \forall x \in [1, X_s - 1] \\ \frac{z[(x+1) \cdot \delta x, t] - 2 \cdot z[x, t] + z[(x-1) \cdot \delta x, t]}{\delta x^2} \\ - \frac{1}{c^2} \cdot \frac{z[x, (t+1) \cdot \delta t] - 2 \cdot z[x, t] + z[x, (t-1) \cdot \delta t]}{\delta t^2} = 0 \end{aligned} \quad (\text{A.9})$$

### Finite-Element

The finite-element method offers a more formalized formulation of the problem. The vertical displacement function  $z$  of the string is first admitted to belong to the vectorial space  $H_{[0, L]}$ , composed of the functions whose square and square

derivative can be integrated on  $[0, L]$ . The definition of a scalar product  $\langle, \rangle$  on  $H_{[0,L]}$  is therefore possible:

$$(f, g) \in H_{[0,L]} \times H_{[0,L]} \quad \langle f, g \rangle = \int_0^L f(x).g(x).dx$$

This ensure that the integral  $\int_0^L f(x)^2.dx$  exists. Note that  $H_{[0,L]}$  is also more constrained when taking into account the limit conditions imposed by the vibrating string problem, such that every function  $f$  belonging to  $H_{[0,L]}$  respect the following conditions:  $f(0) = f(L) = 0$  and  $\frac{\partial f}{\partial x}(0,0) = \frac{\partial f}{\partial x}(L,0) = T$ , with  $T$  the tension applied on the string.

Back to the vibrating string problem, the finite-element method expresses the equation of d'Alembert EQU. (2.16) as a variational formulation which corresponds to the virtual work principle. The problem is to find the function  $z$  satisfying:

$$\forall v \in H_{[0,L]} \quad \langle \frac{\partial^2 z}{\partial x^2}, v \rangle - \frac{1}{c^2} \cdot \langle \frac{\partial^2 z}{\partial t^2}, v \rangle = 0 \quad (\text{A.10})$$

An integration by parts combined with the initial conditions leads to the following simplification:

$$\langle \frac{\partial^2 z}{\partial x^2}, v \rangle = [\frac{\partial z}{\partial x} \cdot v]_0^L - \langle \frac{\partial z}{\partial x}, \frac{\partial v}{\partial x} \rangle = - \langle \frac{\partial z}{\partial x}, \frac{\partial v}{\partial x} \rangle$$

So that EQU. (A.10) is equivalent to finding  $z$  satisfying:

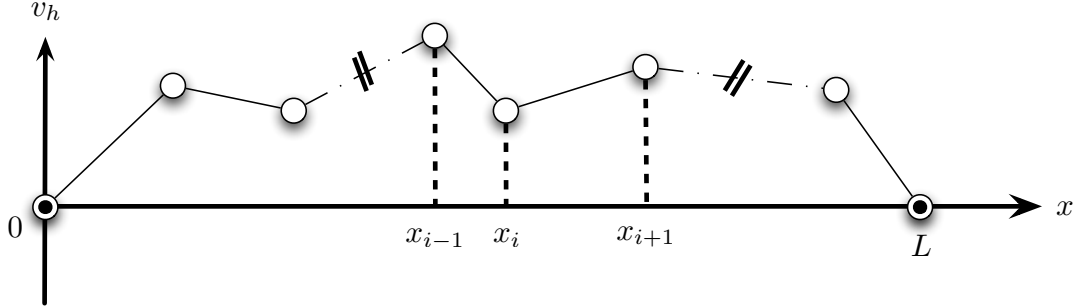
$$\forall v \in H_{[0,L]} \quad \langle \frac{\partial z}{\partial x}, \frac{\partial v}{\partial x} \rangle + \frac{1}{c^2} \cdot \langle \frac{\partial^2 z}{\partial t^2}, v \rangle = 0 \quad (\text{A.11})$$

The discretization of EQU. (A.11) is achieved by approximating the problem on a finite vectorial subspace of  $H_{[0,L]}$ . For simplicity matters, we will consider the vectorial subspace  $V_h$  of  $H_{[0,L]}$  of dimension  $N$  composed of the piecewise continuous functions with the following basis:

$$\forall i \in [1, N], \quad \forall j \in [1, N] \quad w_i(x_j) = \delta_{ij}$$

It leads to the expression of any function  $v_h$  belonging to  $V_h$ , with its geometric correspondance given in FIG. A.3 (note that the equivalent discretization of the string can be not equally distributed).

$$\forall v_h \in V_h \quad v_h(x) = \sum_{i=1}^n v_i \cdot w_i(x) \quad \text{with } v_i = v_h(x_i) \quad (\text{A.12})$$



**FIG. A.3.** Finite-element decomposition of the vibrating string.

The function  $z$  is analogously approximated in the subspace  $V_h$  by a function  $z_h$ , and leads to the discretized problem of finding  $z_h$  satisfying:

$$z_h(x, t) = \sum_{i=1}^n z_i(t) \cdot w_i(x) \quad (\text{A.13})$$

$$\forall v_h \in V_h \quad \left\langle \frac{\partial z_h}{\partial x}, \frac{\partial v_h}{\partial x} \right\rangle + \frac{1}{c^2} \cdot \left\langle \frac{\partial^2 z_h}{\partial t^2}, v_h \right\rangle = 0 \quad (\text{A.14})$$

By introducing the derivative forms of  $v_h$  and  $z_h$ , EQU. (A.12) and (A.13), into the discretized formulation, EQU. (A.14), the final formulation of the finite-element model of the vibrating string is:

$\forall t \in [0, T], \forall j \in [1, N],$  find  $z_j(t)$  so that

$$\forall i \in [1, N] \quad \sum_{j=1}^n \left( \int_0^L \frac{\partial w_j}{\partial x} \cdot \frac{\partial w_i}{\partial x} \cdot dx \right) \cdot z_j(t) + \frac{1}{c^2} \cdot \sum_{j=1}^n \left( \int_0^L w_j \cdot w_i \cdot dx \right) \cdot \frac{\partial^2 z_j}{\partial t^2}(t) = 0$$

This can also be written in a matricial form:

$\forall t \in [0, T],$  find  $\mathbf{Z}(t)$  so that

$$\mathbf{M} \cdot \ddot{\mathbf{Z}}(t) + \mathbf{K} \cdot \mathbf{Z}(t) = 0 \quad (\text{A.15})$$

$$\mathbf{Z}(t) = [z_j(t)]_{j \in [1, N]}$$

$$\mathbf{M} = [m_{ij}]_{(i,j) \in [1, N] \times [1, N]} \text{ with } m_{ij} = \frac{1}{c^2} \cdot \int_0^L w_j \cdot w_i \cdot dx$$

$$\mathbf{K} = [k_{ij}]_{(i,j) \in [1, N] \times [1, N]} \text{ with } k_{ij} = \int_0^L \frac{\partial w_j}{\partial x} \cdot \frac{\partial w_i}{\partial x} \cdot dx$$

The matrices  $\mathbf{M}$  and  $\mathbf{K}$  are called respectively the mass and stiffness matrices of the system. In our case,  $V_h$  is the vectorial subspace composed of the piecewise continuous functions, so that  $\mathbf{M}$  and  $\mathbf{K}$  can be easily expressed knowing the line equation of each basis function  $w_i$ .

EQU. (A.15) can then be discretized in time according to a similar scheme described in the previous point regarding the finite-difference formulation. By defining a time step  $\delta t$  such that  $T = T_s \delta t$ , an explicit scheme of the second order leads to solving:

$\forall t \in [1, T_s - 1]$ , find  $\mathbf{Z}[(t + 1).\delta t]$  so that

$$\mathbf{M}.\mathbf{Z}[(t + 1).\delta t] = (2\mathbf{M} + \delta t^2.\mathbf{K}).\mathbf{Z}[t.\delta t] + \mathbf{M}.\mathbf{Z}[(t - 1).\delta t] \quad (\text{A.16})$$



## A.5 Digital Waveguide Formulation

### Karplus-Strong

FIG. A.4 and EQU. (A.17) describe the formulation of Karplus and Strong for synthesizing sounds from a delay line of length  $p$  for the vibrating string problem. The delay line is initialized randomly by numbers, and the synthesis is achieved by a given modifier that acts on the delay line. The modifier in the case of the vibrating string can be considered as a low-pass filter that accounts for the decay of the tone.

$$z(n) = \frac{1}{2} \cdot [z(n - p) + z(n - p - 1)] \quad (\text{A.17})$$

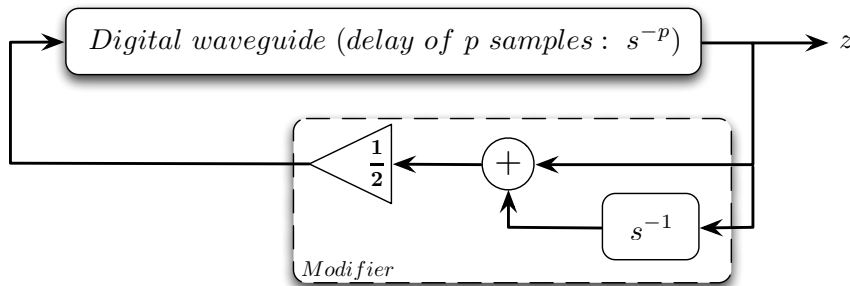
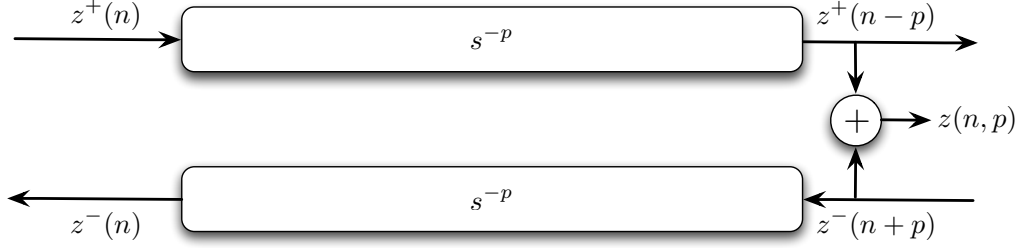


FIG. A.4. Karplus-Strong delay line of the vibrating string.

The method of Karplus and Strong is a somewhat *ad-hoc* formulation of the vibrating string problem since it introduces the design of a filter.

### Digital Waveguide and Finite-Difference Scheme Equivalence

FIG. A.5 depicts the two delay lines involved in the digital waveguide formulation of the vibrating string problem.



**FIG. A.5.** Digital waveguide delay lines of the vibrating string

The physical interpretation of the digital waveguide formulation can be obtained from the equivalence to the finite-difference method. By normalizing the simulation terms  $\delta x$ ,  $\delta t$  and  $c$ , EQU. (A.9) becomes:

$$z(n+1, p) = z(n, p+1) - z(n-1, p) + z(n, p-1) \quad (\text{A.18})$$

As stated by the digital waveguide formulation  $z(n, p) = z^+(n-p) + z^-(n+p)$ , so that:

$$\begin{aligned} z(n, p+1) - z(n-1, p) + z(n, p-1) &= z^+(n-p-1) + z^-(n+p+1) \\ &\quad - z^+(n-1-p) - z^-(n-1+p) \\ &\quad + z^+(n-p+1) + z^-(n+p-1) \\ &= z^+(n-p+1) + z^-(n+p+1) \\ &= z^+[(n+1)-p] + z^-[(n+1)+p] \\ &= z(n+1, p) \end{aligned} \quad (\text{A.19})$$

EQU. (A.18) and (A.19) are then proved to be equivalent. Moreover EQU. (A.19) can be expressed such that:

$$\begin{aligned} z(n+1, p) &= z^+[(n+1)-p] + z^-[(n+1)+p] \\ &= z^+[n-(p-1)] + z^-[n+(p+1)] \end{aligned} \quad (\text{A.20})$$

EQU. (A.20) expresses the vertical displacement of the string at the position  $p$  at the time  $n+1$  as the superposition of the left-going and right-going string displacements at the time  $n$  at the positions  $p-1$  and  $p+1$ , which is exactly the scheme depicted by FIG. A.5.

## A.6 Modal Synthesis Formulation

### Normal Shapes and Modes

Let us recall EQU. (2.19), governing the motion of  $N$  coupled oscillators:

$$\mathbf{M}.\ddot{\mathbf{z}}(t) + \mathbf{K}.\mathbf{z}(t) = 0 \quad (\text{A.21})$$

We look for a factorized solution of the oscillation form  $\mathbf{z}(t) = \mathbf{s} \cdot \sin(\omega.t + \phi)$ ,  $\mathbf{s}$  are called the *modal shapes* of the system, so that the previous equation becomes:

$$\omega^2.\mathbf{M}.\mathbf{s} = \mathbf{K}.\mathbf{s} \Leftrightarrow \mathbf{M}^{-1}.\mathbf{K}.\mathbf{s} = \omega^2.\mathbf{s} \quad (\text{A.22})$$

The system described by EQU. (A.22) is then characterized by  $N$  eigenvalues  $\omega^2$  and eigenvectors  $\mathbf{s}$  (the *modal shapes*) of the matrix  $\mathbf{M}^{-1}.\mathbf{K}$ . The *modal shapes* form then an orthogonal basis of the system, such that:

$$\begin{aligned} \forall (i, j) \in [1, N] \times [1, N] \\ s_j^T.\mathbf{M}.s_i = \delta_{ij}.m_i \\ s_j^T.\mathbf{K}.s_i = \delta_{ij}.k_i \text{ with } k_i = \omega_i^2.m_i \end{aligned}$$

The *modal shapes* define therefore a coordinate *modal transformation* from a system of  $N$  coupled oscillators to a system of  $N$  *uncoupled* normal modes noted  $\mathbf{q}$ , such that:

$$\mathbf{z} = \mathbf{S}.\mathbf{q} \Leftrightarrow \mathbf{q} = \mathbf{S}^T.\mathbf{z} \quad (\text{A.23})$$

Solving EQU. (A.21) is then equivalent to solve the system described by EQU. (A.24):

$$\begin{aligned} \mathbf{M}_S.\ddot{\mathbf{q}}(t) + \mathbf{K}_S.\mathbf{q}(t) = 0 \\ \text{with } \mathbf{M}_S = \mathbf{S}^T.\mathbf{M}.\mathbf{S} \text{ and } \mathbf{K}_S = \mathbf{S}^T.\mathbf{K}.\mathbf{S} \end{aligned} \quad (\text{A.24})$$

$\mathbf{M}_S$  and  $\mathbf{K}_S$  are diagonal mass and stiffness matrices by the transformation  $\mathbf{S}$ . Once the system described by EQU. (A.24) is solved,  $\mathbf{z}$  is then obtained through a linear combination of the *normal modes*  $\mathbf{q}$  weighted by the *modal shapes*, accordingly to EQU. (A.23):

$$\forall j \in [1, N] \quad z_j(t) = \sum_{i=0}^N s_{i,j} \cdot q_i(t)$$

### Continuous Equivalence

Back to the vibrating string problem, let us recall the d'Alembert equation EQU. (2.16):

$$T \cdot \frac{\partial^2 z}{\partial x^2}(x, t) - \mu \cdot \frac{\partial^2 z}{\partial t^2}(x, t) = 0$$

From section A.2, we already know that  $z$  is of the form  $z(x, t) = \sum_{i=1}^N z_i(x, t)$  with  $z_i(x, t) = s_i(x) \cdot q_i(t)$ . By introducing this latter expression in EQU. (2.16), multiplying by  $s_i$  and integrating over the string length leads to:

$$\forall i \in [1, N] \left[ \mu \cdot \int_0^L s_i^2(x) \cdot dx \right] \cdot \frac{d^2 q_i}{dt^2}(t) - \left[ T \cdot \int_0^L \frac{d^2 s_i}{dx^2}(x) \cdot s_i(x) \cdot dx \right] \cdot q_i(t) = 0 \quad (\text{A.25})$$

An integration by parts combined with the initial conditions leads to the following simplification:

$$\forall i \in [1, N] \int_0^L \frac{d^2 s_i}{dx^2}(x) \cdot s_i(x) \cdot dx = \left[ \frac{ds_i}{dx}(x) \cdot s_i(x) \right]_0^L - \int_0^L \left[ \frac{ds_i}{dx}(x) \right]^2 \cdot dx = - \int_0^L \left[ \frac{ds_i}{dx}(x) \right]^2 \cdot dx$$

EQU. (A.25) becomes therefore:

$$\forall i \in [1, N] \left[ \mu \cdot \int_0^L s_i^2(x) \cdot dx \right] \cdot \frac{d^2 q_i}{dt^2}(t) + \left[ T \cdot \int_0^L \left[ \frac{ds_i}{dx}(x) \right]^2 \cdot dx \right] \cdot q_i(t) = 0$$

This latter expression can also be expressed as a matricial equation, EQU. (A.26) is then exactly of the form of EQU. (A.21)

$$\mathbf{M} \cdot \ddot{\mathbf{q}}(t) + \mathbf{K} \cdot \mathbf{q}(t) = 0 \quad (\text{A.26})$$

$$\mathbf{M} = [m_i]_{i \in [1, N]} \text{ with } m_i = \mu \cdot \int_0^L s_i^2(x) \cdot dx$$

$$\mathbf{K} = [k_i]_{i \in [1, N]} \text{ with } k_i = T \cdot \int_0^L \left[ \frac{ds_i}{dx}(x) \right]^2 \cdot dx$$



# B

## Motion Capture Protocol

## Timpani MoCap Session Score

The image displays a musical score for a timpani MoCap session. It consists of six horizontal staves, each representing a different playing technique. Each staff begins with a double bar line and ends with a double bar line. The notes are quarter notes, and there are ten notes per staff. The techniques are: Legato (smooth transitions), Accent (>), Staccato (staccato dots), Tenuto (horizontal dashes), Vertical accent (^ and v), and Piston (diagonal slashes).

For each item, perform the score at the three following locations :

- => normal (one-third of the membrane)
- => center of the membrane
- => circle

**FIG. B.1.** Timpani MoCap protocol



Schulich School of Music  
of McGill University  
555 Sherbrooke Street West  
Montreal, Quebec H3A 1E3



Schulich School of Music  
École de musique Schulich

École de Musique Schulich  
de l'Université McGill  
555, rue Sherbrooke Ouest  
Montréal (Québec) H3A 1E3

☎ (514) 398-4535  
☎ (514) 398-8061  
www.mcgill.ca/music

### Information for Prospective Participants

Dear Participant:

We are interested in studying the influence of expressive body movements in musical performance. Not much is known about the origin and importance of these movements, nor whether they contribute or negatively influence the performance.

In the present study, you will be asked to perform standard pieces in the repertoire adapted to your mastery level. Your performance will be filmed and/or measured using a movement tracker. If using the tracker, small markers will be attached to your clothes and your instrument using tape. Your participation will require less than one hour.

No judgment will be made on your technical abilities or your personal performance style. Your name will not be disclosed at any time. Your performance will be assigned a number and will be grouped with those of other participants and later analyzed according to specific variables. You may discontinue participation in this study at any point during the process. If the task is tiring or you need a break during the recording sessions, kindly inform the test administrator.

If you would like to learn about the results of the study, please contact Prof. Wanderley at 398-4535 (ext. 00917). We thank you for your interest in this study.

Sincerely,

Marcelo M. Wanderley, PhD.  
Assistant Professor.

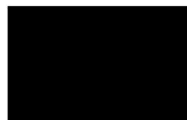
### Participant Consent Form

My participation in the Study "Analysis of Expressive Performer Movements" is voluntary. I understand that I may discontinue participation at any point during the experiment and that my name will not be disclosed at any time during the analysis or the dissemination of findings.

Participant's name



Signature



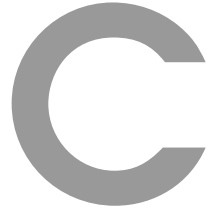
Date

2-10-3/2007

FIG. B.2. Timpani MoCap ethics form







# Bibliography

## C.1 Computer Animation

- [ABC96] K. Amaya, A. Bruderlin, and T. Calvert. Emotion from Motion. In *Proc. of the International Conference on Graphics Interface (GI)*, pages 222–229, 1996.
- [ACD<sup>+</sup>09] C. Awad, N. Courty, K. Duarte, T. Le Naour, and S. Gibet. A Combined Semantic and Motion Capture Database for Real-Time Sign Language Synthesis. In *Proc. of the International Conference on Intelligent Virtual Agents (IVA)*, pages 342–348, 2009.
- [ACSF07] B. Allen, D. Chu, A. Shapiro, and P. Faloutsos. On the Beat! Timing and Tension for Dynamic Characters. In *Proc. of the SIGGRAPH/Eurographics Symposium on Computer Animation (SCA)*, pages 239–247, 2007.
- [AdSP07] Y. Abe, M. da Silva, and J. Popović. Multiobjective Control with Frictional Contacts. In *Proc. of the SIGGRAPH/Eurographics Symposium on Computer Animation (SCA)*, pages 249–258, 2007.
- [AFO03] O. Arikan, D. Forsyth, and J. O’Brien. Motion Synthesis from Annotations. *ACM Transactions on Graphics*, 22(3):402–408, 2003.
- [AFS93] R. Agrawal, C. Faloutsos, and A. Swami. Efficient Similarity Search in Sequence Databases. In *Proc. of the International Conference on Foundations of Data Organization and Algorithms (FODO)*, pages 69–84, 1993.
- [AG85] W. Armstrong and M. Green. The Dynamics of Articulated Rigid Bodies for Purposes of Animation. *The Visual Computer*, 1:231–240, 1985.

- [AJG09] M. Aubry, F. Julliard, and S. Gibet, 2009. Modeling Joint Synergies to Synthesize Realistic Movements. To appear in *Gesture-Based Communication in Human-Computer Interaction*, Springer Verlag.
- [AM00] M. Alexa and W. Müller. Representing Animations by Principal Components. *Computer Graphics Forum*, 19(3):411–426, 2000.
- [Arn88] B. Arnaldi. *Conception du Noyau d’un Système d’Animation de Scènes Tridimensionnelles intégrant les Lois de la Mécanique*. PhD thesis, université de Rennes I, France, 1988.
- [AW90] G. Andersson and J. Winters, 1990. Role of Muscle in Postural Tasks: Spinal Loading and Postural Stability. In *Multiple Muscle Systems: Biomechanics and Movement Organization*, pp. 377–408, Springer-Verlag.
- [Bar96] D. Baraff. Linear-Time Dynamics using Lagrange Multipliers. In *Proc. of the Annual conference on Computer Graphics and Interactive Techniques (SIGGRAPH)*, pages 137–146, 1996.
- [BB88] R. Barzel and A. Barr. A Modeling System based on Dynamic Constraints. *SIGGRAPH Computer Graphics*, 22(4):179–188, 1988.
- [BB98] P. Baerlocher and R. Boulic. Task-priority Formulations for the Kinematic Control of Highly Redundant Articulated Structures. In *Proc. of the International Conference on Intelligent Robots and Systems (IROS)*, pages 323–239, 1998.
- [BB04] P. Baerlocher and R. Boulic. An Inverse Kinematic Architecture enforcing an Arbitrary Number of Strict Priority Levels. *The Visual Computer*, 20(6):402–417, 2004.
- [BDT<sup>+</sup>08] N. Bonneel, G. Drettakis, N. Tsingos, I. Viaud-Delmon, and D. James. Fast Modal Sounds with Scalable Frequency-Domain Synthesis. *ACM Transactions on Graphics*, 27(3):1–9, 2008.
- [BGW09a] A. Bouënard, S. Gibet, and M. M. Wanderley, 2009. Real-time Simulation and Interaction of Percussion Gestures with Sound Synthesis. In *HAL Open Archives*, Technical Report.
- [BGW09b] A. Bouënard, S. Gibet, and M. M. Wanderley. Hybrid Motion Control combining Inverse Kinematics and Inverse Dynamics Controllers for Simulating Percussion Gestures. In *Proc. of the International Conference on Computer Animation and Social Agents (CASA)*, pages 17–20, 2009.

- [BGW09c] A. Bouënard, S. Gibet, and M. M. Wanderley. Real-time Simulation and Interaction of Percussion Gestures with Sound Synthesis. *Graphical Models*, 2009. Submitted.
- [BH00] M. Brand and A. Hertzmann. Style Machines. In *Proc. of the Annual conference on Computer Graphics and Interactive Techniques (SIGGRAPH)*, pages 183–192, 2000.
- [BMT95] R. Boulic, R. Mas, and D. Thalmann. Position Control of the Center of Mass for Articulated Figures in Multiple Support. In *Proc. of the Eurographics Workshop on Animation and Simulation (EG)*, pages 130–143, 1995.
- [BMT96] R. Boulic, R. Mas, and D. Thalmann. A Robust Approach for the Center of Mass Position Control with Inverse Kinetics. *Computers and Graphics*, 20(5):693–701, 1996.
- [BP08] J. Barbič and J. Popović. Real-time Control of Physically based Simulations using Gentle Forces. *ACM Transactions on Graphics*, 27(5):1–10, 2008.
- [BSP+04] J. Barbič, A. Safonova, J. Y. Pan, C. Faloutsos, J. Hodgins, and N. Pollard. Segmenting Motion Capture Data into Distinct Behaviors. In *Proc. of the International Conference on Graphics Interface (GI)*, pages 185–194, 2004.
- [BW71] N. Burtnyk and M. Wein. Computer Generated Keyframe Animation. *Journal of the Society of Motion Picture and Television Engineers*, 80(3):149–153, 1971.
- [BW95] A. Bruderlin and L. Williams. Motion Signal Processing. In *Proc. of the Annual conference on Computer Graphics and Interactive Techniques (SIGGRAPH)*, pages 97–104, 1995.
- [CA08] N. Courty and E. Arnaud. Inverse Kinematics using Sequential Monte Carlo Methods. In *Proc. of the International Conference on Articulated Motion and Deformable Objects (AMDO)*, pages 1–10, 2008.
- [Cat78] E. Catmull. The Problems of Computer-Assisted Animation. *SIGGRAPH Computer Graphics*, 12(3):348–353, 1978.
- [CB06] B. Le Callenec and R. Boulic. Interactive Motion Deformation with Prioritized Constraints. *Graphical Models*, 68(2):175–193, 2006.

- [CBB<sup>+</sup>02] M. Cardle, S. Brooks, L. Barthe, M. Hassan, and P. Robinson. Music-Driven Motion Editing. In *Proc. of the Annual conference on Computer Graphics and Interactive Techniques (SIGGRAPH)*, page 222, 2002.
- [CBBJR03] M. Cardle, S. Brooks, Z. Bar-Joseph, and P. Robinson. Sound-by-Numbers: Motion-Driven Sound Synthesis. In *Proc. of the SIGGRAPH/Eurographics Symposium on Computer Animation (SCA)*, pages 349–356, 2003.
- [CBT07] S. Carvalho, R. Boulic, and D. Thalmann. Interactive Low-dimensional Human Motion Synthesis by Combining Motion Models and PIK. *Computer Animation and Virtual Worlds*, 18(4-5):493–503, 2007.
- [CBvdP09] S. Coros, P. Beaudoin, and M. van de Panne. Robust task-based control policies for physics-based characters. *ACM Transactions on Graphics*, 2009. To appear.
- [CH07] J. Chai and J. Hodgins. Constraint-based Motion Optimization using a Statistical Dynamic Model. *ACM Transactions on Graphics*, 26(3):8:1–8:9, 2007.
- [CHt76] G. Cavagna, N. Heglund, and C. Taylor. The Sources of External Work in Level Walking and Running. *Journal of Physiology*, 262:639–657, 1976.
- [Coh92] M. Cohen. Interactive Spacetime Control for Animation. In *Proc. of the Annual conference on Computer Graphics and Interactive Techniques (SIGGRAPH)*, pages 293–302, 1992.
- [DAH90] G. Dumont, B. Arnaldi, and G. Héron. Automatic Animation Control, 1990. Publication Interne no. 511 Irisa- 1170 Inria, IRISA-INRIA.
- [DG67] W. Dempster and G. Gaughran. Properties of Body Segments based on Size and Weight. *American Journal of Anatomy*, 120(1):33–54, 1967.
- [DTS<sup>+</sup>08] H. Ding, G. Trajcevski, P. Scheuermann, X. Wang, and E. Keogh. Querying and Mining of Time Series Data: Experimental Comparison of Representations and Distance Measures. In *Proc. of the International Conference on Very Large Human-Motion Data Bases*, pages 1542–1552, 2008.

- [DZS08] P. DiLorenzo, V. Zordan, and B. Sanders. Laughing out Loud: Control for Modeling Anatomically inspired Laughter using Audio. *ACM Transactions on Graphics*, 27(5):1–8, 2008.
- [ES03] G. ElKoura and K. Singh. Handrix: Animating the Human Hand. In *Proc. of the SIGGRAPH/Eurographics Symposium on Computer Animation (SCA)*, pages 110–119, 2003.
- [Fel66] A. Feldman. Functional Tuning of the Nervous System with Control of Movement or Maintenance of a Steady Posture – ii. Controllable Parameters of the Muscles. *Biophysics*, 3(11):565–578, 1966.
- [FF05] K. Forbes and E. Fiume. An Efficient Search Algorithm for Motion Data using Weighted PCA. In *Proc. of the SIGGRAPH/Eurographics Symposium on Computer Animation (SCA)*, pages 67–76, 2005.
- [FP03] A. Fang and N. Pollard. Efficient Synthesis of Physically Valid Human Motion. *ACM Transactions on Graphics*, 22(3):417–426, 2003.
- [FvdPT01] P. Faloutsos, M. van de Panne, and D. Terzopoulos. Composable Controllers for Physics-Based Character Animation. In *Proc. of the Annual conference on Computer Graphics and Interactive Techniques (SIGGRAPH)*, pages 251–260, 2001.
- [Gir91] M. Girard, 1991. Constrained Optimization of Articulated Animal Movement in Computer Animation. In *Making them Move: Mechanics, Control and Animation of Articulated Figures*, pp. 209–229, Morgan Kaufmann Publishers Inc.
- [Gle98] M. Gleicher. Retargetting Motion to New Characters. In *Proc. of the Annual conference on Computer Graphics and Interactive Techniques (SIGGRAPH)*, pages 33–42, 1998.
- [GM85] M. Girard and A. Maciejewski. Computational Modeling for the Computer Animation of Legged Figures. *Computer Graphics*, 19(3):263–270, 1985.
- [GM94] S. Gibet and P.-F. Marteau. A Self-organized Model for the Control, Planning and Learning of Nonlinear Multi-Dimensional Systems using a Sensory Feedback. *Journal of Applied Intelligence*, 4(4):337–349, 1994.

- [GM03] S. Gibet and P. F. Marteau. Expressive Gesture Animation Based on Non Parametric Learning of Sensory-Motor Models. In *Proc. of the International Conference on Computer Animation and Social Agents (CASA)*, pages 79–85, 2003.
- [GM07] S. Gibet and P. F. Marteau. Learning for the Control of Dynamical Motion Systems. In *International Conference on Intelligent Systems Design and Applications (ISDA)*, pages 454–459, 2007.
- [GMHP04] K. Grochow, S. Martin, A. Hertzmann, and Z. Popović. Style-Based Inverse Kinematics. *ACM Transactions on Graphics*, 23(3):522–531, 2004.
- [Gre59] T. Greville. The Pseudoinverse of a Rectangular or Singular Matrix and its Application to the Solution of Systems of Linear Equations. *Society for Industrial and Applied Mathematics Review*, 1(1):38–43, 1959.
- [HA09] H-Anim, 2009. *Humanoid Animation Working Group*, <http://h-anim.org>.
- [HCGM06] A. Héloir, N. Courty, S. Gibet, and F. Multon. Temporal Alignment of Communicative Gesture Sequences. *Computer Animation and Virtual Worlds*, 17(3-4):347–357, 2006.
- [HGLG95] J. Hahn, J. Geigel, J. Won Lee, and L. Gritz. An Integrated Approach to Motion and Sound. *The Journal of Visualization and Computer Animation*, 6:109–123, 1995.
- [HPP05] E. Hsu, K. Pulli, and J. Popović. Style Transition for Human Motion. *ACM Transactions on Graphics*, 24(3):1082–1089, 2005.
- [HR91a] J. Hodgins and M. Raibert. Adjusting Step Length for Rough Terrain Locomotion. *Transactions on Robotics and Automation*, 7(3):289–298, 1991.
- [HR91b] J. Hodgins and M. Raibert. Biped Gait Transitions. In *Proc. of the International Conference on Robotics and Automation (ICRA)*, pages 2092–2097, 1991.
- [HWBO95] J. Hodgins, W. Wooten, D. Brogan, and J. O’Brien. Animating Human Athletics. In *Proc. of the Annual conference on Computer Graphics and Interactive Techniques (SIGGRAPH)*, pages 71–78, 1995.

- [IC87] P. Isaacs and M. Cohen. Controlling Dynamic Simulation with Kinematic Constraints, Behavior Functions and Inverse Dynamics. *SIGGRAPH Computer Graphics*, 21(4):215–224, 1987.
- [Joh03] M. Johnson. *Exploiting Quaternions to Support Expressive Interactive Character Motion*. PhD thesis, Massachusetts Institute of Technology, USA, 2003.
- [JYL09] S. Jain, Y. Ye, and K. Liu. Optimization-Based Interactive Motion Synthesis. *ACM Transactions on Graphics*, 28(1):1–10, 2009.
- [KB84] D. Kochanek and R. Bartels. Interpolating Splines with Local Tension, Continuity, and Bias Control. *SIGGRAPH Computer Graphics*, 18(3):33–41, 1984.
- [KB96] H. Ko and N. Badler. Animating Human Locomotion with Inverse Dynamics. *IEEE Computer Graphics and Applications*, 16(2):50–59, 1996.
- [KG03] L. Kovar and . Gleicher. Flexible Automatic Motion Blending with Registration Curves. In *Proc. of the SIGGRAPH/Eurographics Symposium on Computer Animation (SCA)*, pages 214–224, 2003.
- [KGP02] L. Kovar, M. Gleicher, and F. Pighin. Motion Graphs. *ACM Transactions on Graphics*, 21(3):473–482, 2002.
- [KH83] C. Klein and C.-H. Huang. Review of Pseudoinverse Control for Use with Kinematically Redundant Manipulators. *Transactions on Systems, Man and Cybernetics*, 13(3):245–250, 1983.
- [KHL05] T. Komura, E. Ho, and W. H. Lau. Animating Reactive Motion Using Momentum-Based Inverse Kinematics. *Computer Animation and Virtual Worlds*, 16(3-4):213–223, 2005.
- [KKK<sup>+</sup>03] T. Komura, A. Kuroda, S. Kudoh, T. Lan, and Y. Shinagawa. An Inverse Kinematics Method for 3D Figures with Motion Data. In *Proc. of Computer Graphics International (CGI)*, pages 266–271, 2003.
- [KM05] R. Kulpa and F. Multon. Fast Inverse Kinematics and Kinetics Solver for Human-like Figures. In *Proc. of the IEEE-RAS International Conference on Humanoids Robots (HUMANOIDS)*, pages 38–43, 2005.
- [Kor85] J. Korein. *A Geometric Investigation of Reach*. MIT press, 1985.



- [KP06] P. Kry and D. Pai. Interaction Capture and Synthesis. *ACM Transactions on Graphics*, 25(3):872–880, 2006.
- [KPZ<sup>+</sup>04] E. Keogh, T. Palpanas, V. Zordan, D. Gunopulos, and M. Cardle. Indexing Large Human-Motion Databases. In *Proc. of the International Conference on Very Large Human-Motion Data Bases*, pages 780–791, 2004.
- [Kry05] P. Kry. *Interaction Capture and Synthesis of Human Hands*. PhD thesis, University of British Columbia, Canada, 2005.
- [Las87] J. Lasseter. Principles of Traditional Animation Applied to 3D Computer Animation. *SIGGRAPH Computer Graphics*, 21(4):35–44, 1987.
- [LHP05] C. Liu, A. Hertzmann, and Z. Popović. Learning Physics-Based Motion Style with Nonlinear Inverse Optimization. *ACM Transactions on Graphics*, 24(3):1071–1081, 2005.
- [Lié77] A. Liégeois. Automatic Supervisory Control of the Configuration and Behavior of Multi-body Mechanisms. *IEEE Transactions on Systems Man and Cybernetics*, 7(12):868–871, 1977.
- [Lin06] Y. Lin. Efficient Human Motion Retrieval in Large Databases. In *Proc. of the International Conference on Computer Graphics and Interactive Techniques in Australasia and South East Asia (GRAPHITE)*, pages 31–37, 2006.
- [LP02] K. Liu and Z. Popović. Synthesis of Complex Dynamic Character Motion from Simple Animations. In *Proc. of the Annual conference on Computer Graphics and Interactive Techniques (SIGGRAPH)*, pages 408–416, 2002.
- [LT06] S. H. Lee and D. Terzopoulos. Heads Up! Biomechanical Modeling and Neuromuscular Control of the Neck. *ACM Transactions on Graphics*, 25(3):1188–1198, 2006.
- [LvdPF00] J. Lazlo, M. van de Panne, and E. Fiume. Interactive Control of Physically-Based Animation. In *Proc. of the Annual conference on Computer Graphics and Interactive Techniques (SIGGRAPH)*, pages 201–208, 2000.

- [LWS02] Y. Li, T. Wang, and H.-Y. Shum. Motion Texture: a Two-level Statistical Model for Character Motion Synthesis. In *Proc. of the Annual conference on Computer Graphics and Interactive Techniques (SIGGRAPH)*, pages 465–472, 2002.
- [Lyt90] W. Lytle, 1990. More Bells and Whistles. In SIGGRAPH Computer Animation Festival.
- [Lyt01] W. Lytle, 2001. Pipe Dream. In SIGGRAPH Computer Animation Festival.
- [Lyt09] W. Lytle, 2009. *Animusic*, <http://www.animusic.com>.
- [LZWP03] F. Liu, Y. Zhuang, F. Wu, and Y. Pan. 3D Motion Retrieval with Motion Index Tree. *Computer Vision and Image Understanding*, 92(2):265–284, 2003.
- [Man04] M. Mandel. Versatile and Interactive Virtual Humans: Hybrid use of Data-driven and Dynamics-based Motion Synthesis. Master’s thesis, Carnegie Mellon University, USA, 2004.
- [McM84] T. McMahon. Mechanics of Locomotion. *International Journal of Robotics Research*, 3(2):4–28, 1984.
- [MF07] A. Majkowska and P. Faloutsos. Flipping with Physics: Motion Editing for Acrobatics. In *Proc. of the SIGGRAPH/Eurographics Symposium on Computer Animation (SCA)*, pages 35–44, 2007.
- [MK88] A. Maciejewski and C. Klein. Numerical Filtering for the Operation of Robotic Manipulators through Kinematically Singular Configurations. *Journal of Robotic Systems*, 5(6):527–552, 1988.
- [MK08] M. Mahmudi and M. Kallmann. Fast Path Planning using Motion Graphs. In *Proc. of the International Symposium on Interactive 3D Graphics and Games (I3D)*, pages 1–1, 2008.
- [MKY+06] K. Mitobe, T. Kaiga, T. Yukawa, T. Miura, H. Tamamoto, A. Rodgers, and N. Yoshimura. Development of a Motion Capture System for a Hand using a Magnetic Three Dimensional Position Sensor. In *Proc. of the Annual conference on Computer Graphics and Interactive Techniques (SIGGRAPH)*, page 102, 2006.
- [Mot09] Meta Motion, 2009. <http://www.metamotion.com>.

- [MPS06] J. McCann, N. Pollard, and S. Srinivasa. Physics-Based Motion Retiming. In *Proc. of the SIGGRAPH/Eurographics Symposium on Computer Animation (SCA)*, pages 205–214, 2006.
- [MRC05] M. Müller, T. Röder, and M. Clausen. Efficient Content-Based Retrieval of Motion Capture Data. *ACM Transactions on Graphics*, 24(3):677–685, 2005.
- [MZS09] A. Macchietto, V. Zordan, and C. Shelton. Momentum Control for Balance. *ACM Transactions on Graphics*, 28(3):1–8(80), 2009.
- [Nef05] M. Neff. *Aesthetic Exploration and Refinement: a Computational Framework for Expressive Character Animation*. PhD thesis, University of Toronto, Canada, 2005.
- [NF02] M. Neff and E. Fiume. Modeling Tension and Relaxation for Computer Animation. In *Proc. of the SIGGRAPH/Eurographics Symposium on Computer Animation (SCA)*, pages 81–88, 2002.
- [NF04] M. Neff and E. Fiume. Methods for exploring expressive stance. In *Proc. of the SIGGRAPH/Eurographics Symposium on Computer Animation (SCA)*, pages 49–58, 2004.
- [Noc01] O. Nocent. *Animation Dynamique de Corps Déformables Continus : Application à la Simulation de Textiles Tricotés*. PhD thesis, Université de Reims Champagne-Ardenne, France, 2001.
- [OCE01] J. O’Brien, P. Cook, and G. Essl. Synthesizing Sounds from Physically Based Motion. In *Proc. of the Annual conference on Computer Graphics and Interactive Techniques (SIGGRAPH)*, pages 529–536, 2001.
- [OSG02] J. O’Brien, C. Shen, and C. Gatchalian. Synthesizing Sounds from Rigid Body Simulations. In *Proc. of the SIGGRAPH/Eurographics Symposium on Computer Animation (SCA)*, pages 175–181, 2002.
- [PB91a] C. Philips and N. Badler, 1991. Interactive Behaviors for Bipedal Articulated Figures. *Technical Report N° MS-CIS-91-03*, University of Pennsylvania, USA.
- [PB91b] C. Philips and N. Badler. Interactive Behaviors for Bipedal Articulated Figures. In *Proc. of the Annual conference on Computer Graphics and Interactive Techniques (SIGGRAPH)*, pages 359–362, 1991.

- [Pol09] Polhemus, 2009. <http://www.polhemus.com>.
- [PSE<sup>+</sup>00] J. Popović, S. Seitz, M. Erdmann, Z. Popović, and A. Witkin. Interactive Manipulation of Rigid Body Simulations. In *Proc. of the Annual conference on Computer Graphics and Interactive Techniques (SIGGRAPH)*, pages 209–217, 2000.
- [PW99] Z. Popović and A. Witkin. Physically Based Motion Transformation. In *Proc. of the Annual conference on Computer Graphics and Interactive Techniques (SIGGRAPH)*, pages 11–20, 1999.
- [Qua09] Qualisys, 2009. <http://www.qualisys.com>.
- [RAE<sup>+</sup>05] L. Ren, P. Alton, A. Efros, J. Hodgins, and J. Rehg. A Data-driven Approach to Quantifying Natural Human Motion. *ACM Transactions on Graphics*, 24(3):1090–1097, 2005.
- [Rai86] M. Raibert. *Legged Robots that Balance*. MIT Press, 1986.
- [RB09] D. Raunhardt and R. Boulic. Motion Constraint. *The Visual Computer*, 25:509–518, 2009.
- [RCB98] C. Rose, M. Cohen, and B. Bodenheimer. Verbs and Adverbs: Multidimensional Motion Interpolation using Radial Basis Functions. *IEE Computer Graphics and Applications*, 18(5):32–41, 1998.
- [RGBC96] C. Rose, B. Guenter, B. Bodenheimer, and M. Cohen. Efficient Generation of Motion Transitions using Spacetime Constraints. In *Proc. of the Annual conference on Computer Graphics and Interactive Techniques (SIGGRAPH)*, pages 147–154, 1996.
- [RH91] M. Raibert and J. Hodgins. Animation of Dynamic Legged Locomotion. *SIGGRAPH Computer Graphics*, 25(4):349–358, 1991.
- [RL06] N. Raghuvanshi, , and M. C. Lin. Interactive Sound Synthesis for Large Scale Environments. In *Proc. of the International Symposium on Interactive 3D Graphics and Games (I3D)*, pages 101–108, 2006.
- [RP03] P. Reitsma and N. Pollard. Perceptual Metrics for Character Animation: Sensitivity to Errors in Ballistic Motion. *ACM Transactions on Graphics*, 22(3):537–542, 2003.
- [RSC01] C. Rose, P. P. Sloan, and M. Cohen. Artist-directed Inverse Kinematics Using Radial Basis Function Interpolation. *Computer Graphics Forum*, 20(3):239–250, 2001.

- [SCCH09] T. Shiratori, B. Coley, R. Cham, and J. Hodgins. Simulating Balance Recovery Responses to Trips Based on Biomechanical Principles. In *Proc. of the SIGGRAPH/Eurographics Symposium on Computer Animation (SCA)*, pages 37–46, 2009.
- [Sho85] K. Shoemake. Animating Rotation with Quaternion Curves. *SIGGRAPH Computer Graphics*, 19(3):245–254, 1985.
- [SHP04] A. Safonova, J. Hodgins, and N. Pollard. Synthesizing Physically Realistic Human Motion in Low-Dimensionnal, Behavior-Specific Spaces. *ACM Transactions on Graphics*, 23(3):514–521, 2004.
- [SKP08] S. Sueda, A. Kaufman, and D. K. Pai. Musculotendon Simulation for Hand Animation. *ACM Transactions on Graphics*, 27(3):83:1–8, 2008.
- [Smi09] R. Smith, 2009. *Open Dynamics Engine*, <http://www.ode.org>.
- [SPF03] A. Shapiro, F. Pighin, and P. Faloutsos. Hybrid Control For Interactive Character Animation. In *Proc. of the Pacific Conference on Computer Graphics and Applications (PG)*, pages 455–461, 2003.
- [SS91] B. Siciliano and J. J. Slotine. A General Framework for Managing Multiple Tasks in Highly Redundant Robotic Systems. In *Proc. of International Conference on Advanced Robotics*, pages 1211–1216, 1991.
- [ST09] Advanced Industrial Science and Technology, 2009. *Human Body Properties Database*, <http://riodb.ibase.aist.go.jp/dhbodydb/properties/index-e.html>.
- [TGB00] D. Tolani, A. Goswami, and N. Badler. Real-Time Inverse Kinematics Techniques for Anthropomorphic Limbs. *Graphical Models*, 62(5):353–388, 2000.
- [TH92] T. Takala and J. Hahn. Sound Rendering. *SIGGRAPH Computer Graphics*, 26(2):211–220, 1992.
- [Uni09] Carnegie Mellon University, 2009. *Motion Capture Database*, <http://mocap.cs.cmu.edu>.
- [UPBS08] L. Unzueta, M. Peinado, R. Boulic, and A. Suescun. Full-Body Performance Animation with Sequential Inverse Kinematics. *Graphical Models*, 70(5):87–104, 2008.

- [VAV<sup>+</sup>07] D. Vlastic, R. Adelsberger, G. Vannucci, J. Barnwell, M. Gross, W. Matusik, and J. Popović. Practical Motion Capture in Everyday Surroundings. *ACM Transactions on Graphics*, 26(3):35:1–35:9, 2007.
- [vdPFV90] M. van de Panne, E. Fiume, and Z. Vranesic. Reusable Motion Synthesis using State-Space Controllers. In *Proc. of the Annual conference on Computer Graphics and Interactive Techniques (SIGGRAPH)*, number 4, pages 225–234, 1990.
- [Vic09] Vicon, 2009. <http://www.vicon.com>.
- [vWvBE<sup>+</sup>09] H. van Welbergen, B. van Basten, A. Egges, Z. Ruttkay, and M. Overmars. Real-time Animation of Virtual Humans: A Trade-off between Naturalness and Control. In *Proc. of the Eurographics Workshop on Animation and Simulation (EG)*, pages 45–72, 2009.
- [Wam86] C. Wampler. Manipulator Inverse Kinematic Solutions based on Vector Formulations and Damped Least Squares. *IEEE Transactions on Systems, Man and Cybernetics*, 16(1):93–101, 1986.
- [WB85] J. Wilhelms and B. Barsky. Using Dynamic Analysis to Animate Articulated Bodies as Humans and Robots. In *Proc. of the International Conference on Graphics Interface (GI)*, pages 97–104, 1985.
- [WC91] L. C. Wang and C. C. Chen. A Combined Optimization Method for Solving the Inverse Kinematics Problems of Mechanical Manipulators. *IEEE Transactions on Robotics and Automation*, 7(4):489–499, 1991.
- [WE84] W. Wolovitch and H. Elliot. A Computational Technique for Inverse Kinematics. In *Proc. of the Conference on Decision and Control*, pages 1359–1362, 1984.
- [WFH09] J. Wang, D. Fleet, and A. Hertzmann. Optimizing Walking Controllers. *ACM Transactions on Graphics*, 2009. To appear.
- [WG97] A. Wood-Gaines. Modelling Expressive Movement of Musicians. Master’s thesis, Simon Fraser University, Canada, 1997.
- [WH97a] D. Wiley and J. Hahn. Interpolation Synthesis of Articulated Figure Motion. *IEEE Transactions on Man-Machine Systems*, 17:39–45, 1997.

- [WH97b] W. Wooten and J. Hodgins. Transitions Between Dynamically Simulated Motions: Leaping, Tumbling, Landing and Balancing. In *ACM SIGGRAPH 97 Visual Proceedings: The Art and Interdisciplinary Programs of SIGGRAPH'97*, page 217, 1997.
- [Whi69] D. Whitney. Resolved Motion Rate Control of Manipulators and Human Prostheses. *IEEE Transactions on Man-Machine Systems*, 10(2):47–53, 1969.
- [Wil86] J. Wilhelms. Virya - A Motion Control Editor for Kinematic and Dynamic Animation. In *Proc. of the International Conference on Graphics Interface (GI)*, pages 141–146, 1986.
- [Wil91] J. Wilhelms, 1991. Dynamic experiences. In *Making them Move: Mechanics, Control and Animation of Articulated Figures*, pp 265–280, Morgan Kaufmann Publishers.
- [WJM06] P. Wrotek, O. Chadwicke Jenkins, and M. McGuire. Dynamo: Dynamic, Data-Driven Character Control with Adjustable Balance. In *Proc. of the SIGGRAPH Symposium on Videogames (Sandbox)*, pages 61–70, 2006.
- [WK88] A. Witkin and M. Kass. Spacetime Constraints. *SIGGRAPH Computer Graphics*, 22(4):159–168, 1988.
- [WLZ06] Y. Wang, Z. Liu, and L. Zhou. Key-styling: Learning Motion Style for Real-time Synthesis of 3D Animation. *Computer Animation and Virtual Worlds*, 17(3-4):229–237, 2006.
- [Woo98] W. Wooten. Simulation of Leaping, Tumbling, Landing and Balancing. Master's thesis, Georgia Institute of technology, USA, 1998.
- [Wri05] M. Wright. Open Sound Control: an Enabling Technology for Musical Networking. *Organized Sound*, 10(3):193–200, 2005.
- [XZ07] J. Xiang and H. Zhu. Ensemble HMM Learning for Motion Retrieval with Non-linear PCA Dimensionality Reduction. In *Proc. of the International Conference on International Information Hiding and Multimedia Signal Processing (IIH-MSP)*, pages 604–607, 2007.
- [YCP03] K. Yin, M. Cline, and D. Pai. Motion Perturbation based on Simple Neuromotor Control Models. In *Proc. of the Pacific Conference on Computer Graphics and Applications (PG)*, pages 445–449, 2003.

- [YDYY05] F. Yang, L. Ding, C. Yang, and X. Yuan. An Algorithm for Simulating Human Arm Movement considering the Comfort Level. *Simulation Modelling Practice and Theory*, 13(5):437–449, 2005.
- [YLS04] P. F. Yang, J. Laszlo, and K. Singh. Layered Dynamic Control for Interactive Character Swimming. In *Proc. of the SIGGRAPH/Eurographics Symposium on Computer Animation (SCA)*, pages 39–47, 2004.
- [YLvdP07] K. Yin, K. Loken, and M. van de Panne. SIMBICON: Simple Biped Locomotion Control. *ACM Transactions on Graphics*, 26(3):105:1–105:10, 2007.
- [ZCCD04] V. Zordan, B. Celly, B. Chiu, and P. DiLorenzo. Breathe easy: Model and control of simulated respiration for animation. In *Proc. of the SIGGRAPH/Eurographics Symposium on Computer Animation (SCA)*, pages 29–37, 2004.
- [ZH99] V. Zordan and J. Hodgins. Tracking and modifying upper-body human motion data with dynamic simulation. In *CAS'99: Proc. of Computer Animation and Simulation, Eurographics*, pages 13–22, 1999.
- [ZH02] V. Zordan and J. Hodgins. Motion Capture-Driven Simulations that Hit and React. In *Proc. of the SIGGRAPH/Eurographics Symposium on Computer Animation (SCA)*, pages 89–96, 2002.
- [ZMCF05] V. Zordan, A. Majkowska, B. Chiu, and M. Fast. Dynamic Response for Motion Capture Animation. *ACM Transactions on Graphics*, 24(3):697–701, 2005.
- [ZMM<sup>+</sup>07] V. Zordan, A. Macchietto, J. Medina, M. Soriano, and C. C. Wu. Interactive Dynamic Response for Games. In *Proc. of the SIGGRAPH Symposium on Videogames (Sandbox)*, pages 9–14, 2007.
- [ZW90] F. Zajac and J. Winters, 1990. Modeling Musculoskeletal Movement Systems: Joint and Body Segmental Dynamics, Musculoskeletal Actuation and Neuromuscular Control. In *Multiple Muscle Systems: Biomechanics and Movement Organization*, pp 121–148, Springer-Verlag.



## C.2 Computer Music

- [AC06] F. Avanzini and P. Crosato. Integrating Physically-based Sound Models in a Multimodal Rendering Architecture. *Computer Animation and Virtual Worlds*, 17(3-4):411–419, 2006.
- [Adr89] J. M. Adrien. *Etude de Structures Complexes Vibrantes, Applications à la Synthèse par Modèles Physiques*. PhD thesis, Université de Paris VI, France, 1989.
- [Adr91] J. M. Adrien, 1991. The Missing Link: Modal Synthesis. In *Representations of Musical Signals*, pp. 269–298, MIT Press.
- [Aim02] R. Aimi. New Expressive Percussion Instruments. Master’s thesis, Massachusetts Institute of Technology, USA, 2002.
- [Aim06] R. Aimi. *Hybrid Percussion: Extending Physical Instruments using Sampled Acoustics*. PhD thesis, Massachusetts Institute of Technology, USA, 2006.
- [ALF00] M. Aird, J. Laird, and J. Fitch. Modelling a Drum by Interfacing 2-D and 3-D Waveguide Meshes. In *Proc. of the International Computer Music Conference (ICMC)*, pages 82–85, 2000.
- [AR01] F. Avanzini and D. Rochesso. Controlling Material Properties in Physical Models and Sounding Objects. In *Proc. of the International Computer Music Conference (ICMC)*, pages 91–94, 2001.
- [AR04] F. Avanzini and D. Rochesso. Physical Modeling of Impacts: Theory and Experiments on Contact Time and Spectral Centroid. In *Proc. of the International Conference on Sound and Music Computing (SMC)*, pages 287–293, 2004.
- [Ban06] B. Bank. *Physics-based Sound Synthesis of String Instruments including Geometric Nonlinearities*. PhD thesis, Budapest University of Technology and Economics, Hungary, 2006.
- [BCT<sup>+</sup>99] I. Bork, A. Chaigne, L.-C. Trebuchet, M. Kosfelder, and D. Pillot. Comparison between Modal Analysis and Finite Element Modeling of a Marimba Bar. *Acta Acustica united with Acustica*, 85(2):258–266, 1999.
- [BDT<sup>+</sup>08] N. Bonneel, G. Drettakis, N. Tsingos, I. Viaud-Delmon, and D. James. Fast Modal Sounds with Scalable Frequency-Domain Synthesis. *ACM Transactions on Graphics*, 27(3):1–9, 2008.

- 
- [Ben03] J. Bensoam. *Représentation Intégrale appliquée à la Synthèse Sonore par Modélisation Physique : Méthode des Eléments Finis*. PhD thesis, Université du Maine, France, 2003.
- [BGW08] A. Bouënard, S. Gibet, and M. M. Wanderley. Enhancing the Visualization of Percussion Gestures by Virtual Character Animation. In *Proc. of the International Conference on New Interfaces for Musical Expression (NIME)*, pages 38–43, 2008.
- [Bil05a] S. Bilbao. A Finite Difference Scheme for Plate Synthesis. In *Proc. of the International Computer Music Conference (ICMC)*, 2005.
- [Bil05b] S. Bilbao. Sound Synthesis for Nonlinear Plates. In *Proc. of the International Conference on Digital Audio Effects (DAFx)*, 2005.
- [BMS89] R. Boie, M. Mathews, and A. Schloss. The Radio Drum as a Synthesizer Controller. In *Proc. of the International Computer Music Conference (ICMC)*, pages 42–45, 1989.
- [Bor95] I. Bork. Practical Tuning of Xylophones Bars and Resonators. *Applied Acoustics*, 46:103–127, 1995.
- [Bou06] A. Bouënard. Réalisation d’un Système Expérimental pour la Simulation de Gestes de Percussion. Master’s thesis, Université de Rennes I, France, 2006.
- [BRC02] F. Bevilacqua, J. Ridenour, and D. Cuccia. Mapping Music to Gesture: a Study using 3D motion capture data. In *Proc. of the Symposium on Sensing and Input Mediacentric Systems (SIMS)*, 2002.
- [BRS97] G. Borin, D. Rochesso, and F. Scalcon. A Physical Piano Model for Music Performance. In *Proc. of the International Computer Music Conference (ICMC)*, pages 350–353, 1997.
- [BSM99] J. Bretos, C. Santamaria, and J. Moral. Finite Element Analysis and Experimental Measurements of Natural Eigenmodes and Random Responses of Wooden Bars in Musical Instruments. *Applied Acoustics*, 56:141–156, 1999.
- [BT01] C. Bahn and D. Trueman. Interface: Electronic Chamber Ensemble. In *ACM CHI 2001 New Interfaces for Musical Expression (NIME) Workshop*, 2001.
-

- [Buc09a] D. Buchla, 2009. *Thunder MIDI Controller*, <http://www.buchla.com/historical/thunder>.
- [Buc09b] D. Buchla, 2009. *Lightning II MIDI Controller*, <http://www.buchla.com/lightning>.
- [BWG09] A. Bouënard, M. M. Wanderley, and S. Gibet. Advantages and Limitations of Simulating Percussion Gestures for Sound Synthesis. In *Proc. of the International Computer Music Conference (ICMC)*, pages 255–261, 2009.
- [BWGM09] A. Bouënard, M. M. Wanderley, S. Gibet, and F. Marandola. Synthesis of Music Performances: Advantages and Limitations of Simulating Percussion Gestures for Sound Synthesis. *Computer Music Journal*, 2009. Submitted.
- [CA09] CORDIS-ANIMA, 2009. *ACROE*, <http://www-acroe.imag.fr>.
- [Cad79] C. Cadoz. *Synthèse de Sons par Simulation de Mécanismes Vibratoires : Application aux Sons Musicaux*. PhD thesis, Institut National Polytechnique de Grenoble, France, 1979.
- [Cad94] C. Cadoz. Le Geste Canal de Communication Homme/Machine. La Communication "Instrumentale". *Technique et Science Informatique*, 13(1):31–61, 1994.
- [CC09] N. Castagné and C. Cadoz. Genesis<sup>3</sup> – Plate-forme pour la Création Musicale à l'aide des Modèles Physiques CORDIS-ANIMA. In *Proc. of the Journées d'Informatique Musicale (JIM)*, pages 161–170, 2009.
- [CCH04] A. Cont, T. Coduys, and C. Henry. Real-time Gesture Mapping in PD Environment using Neural Networks. In *Proc. of the International Conference on New Interfaces for Musical Expression (NIME)*, pages 39–42, 2004.
- [CCW09] M. Collicutt, C. Casciato, and M. M. Wanderley. From Real to Virtual: a Comparison of Input Devices for Percussion Tasks. In *Proc. of the International Conference on New Interfaces for Musical Expression (NIME)*, pages 1–6, 2009.
- [CD97] A. Chaigne and V. Doutaud. Numerical Simulation of Xylophones. I. Time-domain modeling of the vibrating bars. *Journal of the Acoustical Society of America*, 101(1):539–557, 1997.

- [Cha97] J. Chadabe. *Electric Sound: the Past and Promise of Electronic Music*. Prentice-Hall, 1997.
- [CLF90] C. Cadoz, L. Lisowski, and J. L. Florens. A Modular Feedback Keyboard Design. *Computer Music Journal*, 14(2):47–51, 1990.
- [CLF93] C. Cadoz, A. Luciani, and J. L. Florens. CORDIS-ANIMA: a Modeling and Simulation System for Sound and Image Synthesis - the General Formalism. *Computer Music Journal*, 17(1):19–29, 1993.
- [Coo01] P. Cook. Principles for Designing Computer Music Controllers. In *ACM CHI 2001 New Interfaces for Musical Expression (NIME) Workshop*, pages 1–4, 2001.
- [COW07] K. Chuchacz, S. O’Modhrain, and R. Woods. Physical Models and Musical Controllers: Designing a Novel Electronic Percussion Instrument. In *Proc. of the International Conference on New Interfaces for Musical Expression (NIME)*, pages 37–40, 2007.
- [Cra09] S. Crab, 2009. *120 Years of Electronic Music*, <http://120years.net>.
- [CS99] P. Cook and G. Scavone. The Synthesis Toolkit (STK). In *Proc. of the International Computer Music Conference (ICMC)*, pages 164–166, 1999.
- [CS09] P. Cook and G. Scavone, 2009. *The Synthesis ToolKit*, <http://ccrma.stanford.edu/software/stk>.
- [CvdDLH07] R. Corbett, K. van den Doel, J. Lloyd, and W. Heidrich. Timbre-fields: 3d Interactive Sound Models for Real-Time Audio. *Presence: Teleoperators and Virtual Environments*, 16(6):643–654, 2007.
- [DD07] N. D’Alessandro and T. Dutoit. Handsketch bi-manual controller: Investigation on expressive control issues of an augmented tablet. In *Proc. of the International Conference on New Interfaces for Musical Expression (NIME)*, pages 78–81, 2007.
- [DMC98] V. Doutaud, D. Matignon, and A. Chaigne. Numerical Simulation of Xylophones. II. Time-domain Modeling of the Resonator and of the Radiated Sound Pressure. *Journal of the Acoustical Society of America*, 104(3):1633–1647, 1998.
- [Dol86] M. Dolson. The Phase Vocoder: a Tutorial. *Computer Music Journal*, 10(4):14–27, 1986.

- [Doo02] P. Doornbusch. Composers' Views on Mapping in Algorithmic Composition. *Organized Sound*, 7(2):152–156, 2002.
- [dSWS05] A. da Silva, M. M. Wanderley, and G. Scavone. On the Use of Flute Air Jet as a Musical Control Variable. In *Proc. of the International Conference on New Interfaces for Musical Expression (NIME)*, pages 105–108, 2005.
- [EBC05] N. Ellis, J. Bensoam, and R. Caussé. Modalys Demonstration. In *Proc. of the International Computer Music Conference (ICMC)*, pages 101–102, 2005.
- [Ego95] E. Egozy. Deriving Musical Control Features from a Real-time Timbre Analysis of the Clarinet. Master's thesis, Massachusetts Institute of Technology, USA, 1995.
- [FC91] J. L. Florens and C. Cadoz, 1991. The Physical Model: Modeling and Simulating the Instrumental Universe. In *Representations of Musical Signals*, pp. 227–268, MIT Press.
- [FG66] J. Flanagan and R. Golden. The Phase Vocoder. *The Bell System Technical Journal*, 45:1493–1509, 1966.
- [FR95] F. Fontana and D. Rochesso. A New Formulation of the 2D-waveguide mesh for percussion instruments. In *Proc of the XI Colloquium on Musical Informatics*, pages 27–30, 1995.
- [FS09] A. Freed and A. Schmeder. Features and Future of Open Sound Control version 1.1 for NIME. In *Proc. of the International Conference on New Interfaces for Musical Expression (NIME)*, pages 116–120, 2009.
- [Gar87] G. Garnett. Modeling Piano Sound using Digital Waveguide Filtering Techniques. In *Proc. of the International Computer Music Conference (ICMC)*, pages 89–95, 1987.
- [Gib87] S. Gibet. *Codage, Représentation et Traitement du Geste Instrumental*. PhD thesis, Institut National Polytechnique de Grenoble, France, 1987.
- [GM90] S. Gibet and P. F. Marteau. Gestural Control of Sound Synthesis. In *Proc. of the International Computer Music Conference (ICMC)*, pages 387–391, 1990.

- [HK00] A. Hunt and R. Kirk, 2000. Mapping Strategies for Musical Performance. In *Trends in Gestural Control of Music*, pp. 231–258.
- [HR71] L. Hiller and P. Ruiz. Synthetizing Musical Sounds by Solving the Wave Equation for Vibrating Objects: Parts I and II. *Journal of Audio Engineering Society*, 19(6 and 7):462–470 and 542–551, 1971.
- [HST96] R. Hänninen, L. Savioja, and T. Takala. Virtual concert performance - synthetic animated musicians playing in an acoustically simulated room. In *Proc. of the International Computer Music Conference (ICMC)*, pages 402–404, 1996.
- [HWK00] A. Hunt, M. M. Wanderley, and R. Kirk. Towards a Model for Instrumental Mapping in Expert Musical Interactions. In *Proc. of the International Computer Music Conference (ICMC)*, pages 209–212, 2000.
- [JBP06] D. James, J. Barbič, and D. Pai. Precomputed Acoustic Transfer: Output-sensitive Accurate Sound Generation for Geometrically Complex Vibration Sources. *ACM Transactions on Graphics*, 25(3):987–995, 2006.
- [Jen07] A. Jensenius. *Action – Sound: Developing Methods and Tools to Study Music-related Body Movement*. PhD thesis, University of Oslo, Norway, 2007.
- [JKGB05] S. Jordá, M. Kaltenbrunner, G. Geiger, and R. Bencina. The reacTable\*. In *Proc. of the International Computer Music Conference (ICMC)*, pages 379–382, 2005.
- [JS83] D. Jaffe and J. O. Smith. Extensions of the Karplus-Strong Plucked-String Algorithm. *Computer Music Journal*, 7(2):56–69, 1983.
- [Kap05] A. Kapur. A History of Robotic Musical Instruments. In *Proc. of the International Computer Music Conference (ICMC)*, pages 21–28, 2005.
- [Kap07] A. Kapur. *Digitizing North Indian Music: Preservation and Extension using Multimodal Sensor Systems, Machine Learning and Robotics*. PhD thesis, University of Victoria, Canada, 2007.
- [KEA<sup>+</sup>02] M. Karjalainen, P. Esquef, P. Ansalo, A. Mäkitvirta, and V. Välimäki. Frequency-zooming ARMA Modeling of Resonant and Reverberant Systems. *Journal of Audio Engineering Society*, 50(12):1012–1029, 2002.

- [KEDC03] A. Kapur, G. Essl, P. Davidson, and P. Cook. The Electronic Tabla Controller. *Journal of New Music Research*, 32(4):351–359, 2003.
- [KL62] J. Kelly and C. Lochbaum. Speech Synthesis. In *Proc. of the International Congress on Acoustics (ICA)*, pages 1–4, 1962.
- [KLD<sup>+</sup>04] A. Kapur, A. Lazier, P. Davidson, R. Wilson, and P. Cook. The Electronic Sitar Controller. In *Proc. of the International Conference on New Interfaces for Musical Expression (NIME)*, pages 7–12, 2004.
- [KSB<sup>+</sup>07] A. Kapur, E. Singer, M. S. Benning, G. Tzanetakis, and Trimpin. Integrating Hyperinstruments, Musical Robots, & Machine Musicianship for North Indian Classical Music. In *Proc. of the International Conference on New Interfaces for Musical Expression (NIME)*, pages 238–241, 2007.
- [KTS<sup>+</sup>07] A. Kapur, Trimpin, E. Singer, A. Suleman, and G. Tzanetakis. A Comparison of Solenoid-Based Strategies for Robotic Drumming. In *Proc. of the International Computer Music Conference (ICMC)*, pages 393–396, 2007.
- [KV93] M. Karjalainen and V. Välimäki. Model-based analysis/synthesis of the acoustic guitar. In *Proc. of the Stockholm Music Acoustics Conference (SMAC)*, pages 443–447, 1993.
- [kvdDP96] k. van den Doel and D. Pai. Synthesis of Shape Dependent Sounds with Physical Modeling. In *Proc. of the International Conference on Auditory Display (ICAD)*, 1996.
- [kvdDP98] k. van den Doel and D. Pai. The Sounds of Physical Shapes. *Presence: Teleoperators and Virtual Environments*, 7(4):382–395, 1998.
- [Läh09] O. Lähdeoja. Augmenting Chordophones with Hybrid Percussive Sound Possibilities. In *Proc. of the International Conference on New Interfaces for Musical Expression (NIME)*, pages 102–105, 2009.
- [Lau89] R. Laughlin. Approximating Harmonic Amplitude Envelopes of Musical Instrument Sounds with Principal Component Analysis. Master’s thesis, Simon Fraser University, Canada, 1989.
- [LFCC07] A. Luciani, J. L. Florens, D. Couroussé, and C. Cadoz. Ergotic Sounds: a New Way to improve Playability, Believability and Presence of Digital Musical Instruments. In *Proc. of the International*

- 
- Conference on Enactive Interfaces (ENACTIVE)*, pages 373–376, 2007.
- [Luc85] A. Luciani. *Un Outil Informatique de Création d'Images Animées : Modèles d'Objets, Langage, Contrôle Gestuel en Temps-réel – Le Syst'eme ANIMA*. PhD thesis, Institut National Polytechnique de Grenoble, France, 1985.
- [LVKL96] T. Laakso, V. Välimäki, M. Karjalainen, and U. Laine. Splitting the Unit Delay - Tools for Fractional Delay Filters Design. *IEEE Signal Processing Magazine*, 13(1):30–60, 1996.
- [Mal07] J. Malloch. A Consort of Gestural Musical Controllers: Design, Construction, and Performance. Master's thesis, McGill University, Canada, 2007.
- [MBW05] E. Motuk, S. Bilbao, and R. Woods. Implementation of Finite Difference Schemes for the Wave Equation on FPGA. In *Proc. of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 237–240, 2005.
- [MID09] MIDI, 2009. *MIDI Manufacturers Association*, <http://www.midi.org>.
- [Mod09] Modalys, 2009. *IRCAM*, <http://support.ircam.fr/doc-modalys/>.
- [Moo03] B. Moore. *An Introduction to the Psychology of Hearing*. Academic Press, 2003. Fifth Edition.
- [MPG62] M. Mathews, J. Pierce, and N. Guttman. Musikalische Klänge von Digitalrechnern (Musical Sounds from Digital Computers). *Gravesaner Blätter*, (23-24):109–118:119–128, 1962.
- [MRM02] M. Marshall, M. Rath, and B. Moynihan. The Virtual Bodhran - The Vodhran. In *Proc. of the International Conference on New Interfaces for Musical Expression (NIME)*, pages 179–180, 2002.
- [MSW83] M. McIntyre, R. Schumacher, and J. Woodhouse. On the Oscillations of Musical Instruments. *Journal of the Acoustical Society of America*, 74(5):1325–1345, 1983.
- [MW06] E. Miranda and M. M. Wanderley. *New Digital Musical Instruments: Control and Interaction Beyond the Keyboard*. A-R Editions., 2006.
-



- [MW07] J. Malloch and M. M. Wanderley. The T-Stick: From Musical Interface to Musical Instrument. In *Proc. of the International Conference on New Interfaces for Musical Expression (NIME)*, pages 66–69, 2007.
- [Nak00] T. Marrin Nakra. *Inside the Conductor’s Jacket: Analysis, Interpretation and Musical Synthesis of Expressive Gesture*. PhD thesis, Massachusetts Institute of Technology, USA, 2000.
- [OB91] F. Ordūna-Bustamante. Nonuniform Beams with Harmonically-related Overtones for Use in Percussion Instruments. *Journal of the Acoustical Society of America*, 90(6):2935–2941, 1991.
- [OCE01] J. O’Brien, P. Cook, and G. Essl. Synthesizing Sounds from Physically Based Motion. In *Proc. of the Annual conference on Computer Graphics and Interactive Techniques (SIGGRAPH)*, pages 529–536, 2001.
- [OSC09] OSC, 2009. *Open Sound Control*, <http://opensoundcontrol.org>.
- [OSG02] J. O’Brien, C. Shen, and C. Gatchalian. Synthetizing Sounds from Rigid Body Simulations. In *Proc. of the SIGGRAPH/Eurographics Symposium on Computer Animation (SCA)*, pages 175–181, 2002.
- [PDH<sup>+</sup>09] X. Pestova, E. Donald, H. Hindman, J. Malloch, M. Marshall, F. Rocha, S. Sinclair, A. Stewart, M. M. Wanderley, and S. Ferguson. The CIRMMT/McGill Digital Orchestra Project. In *Proc. of the International Computer Music Conference (ICMC)*, pages 295–298, 2009.
- [Pie65] J. Pierce. Portrait of the Machine as a Young Artist. *Playboy*, 12(6):184–185, 1965.
- [Pla05] C. Plack. *The Sense of Hearing*. Psychology Press, 2005.
- [Pol91] G. De Poli, 1991. A Tutorial on Digital Sound Synthesis Techniques. In *The Music Machine*, pp. 429–447, MIT Press.
- [PQ08] C. Palacio-Quintin. Eight Years of Practice on the Hyper-Flute: Technological and Musical Perspectives. In *Proc. of the International Conference on New Interfaces for Musical Expression (NIME)*, pages 293–298, 2008.
- [Puc96] M. Puckette. Pure Data. In *Proc. of the International Computer Music Conference (ICMC)*, pages 269–272, 1996.

- 
- [PvdDJ<sup>+</sup>01] D. Pai, K. van den Doel, D. James, J. Lang, J. Lloyd, J. Richmond, and S. Yau. Scanning Physical Interaction Behavior of 3D Objects. In *Proc. of the Annual conference on Computer Graphics and Interactive Techniques (SIGGRAPH)*, pages 87–96, 2001.
- [Ram91] C. Ramstein. *Analyse, Représentation et Traitement du Geste Instrumental*. PhD thesis, Institut National Polytechnique de Grenoble, France, 1991.
- [Raz86] A. Razafindrakoto. *Le Système ANIMA: Editeur d’Objets Producteurs d’Images, Implantation d’Algorithmes de Simulation Temps-réel*. PhD thesis, Université Scientifique, Technologique et Médicale de Grenoble, France, 1986.
- [RDFL94] X. Rodet, P. Depalle, G. Fleury, and F. Lazarus, 1994. Modèles de Signaux et Modèles Physiques d’Instruments : Etudes et Comparaisons. In *Modèles Physiques, Création Musicale et Ordinateurs*, Maison des Sciences de l’Homme, pp. 351–370.
- [RH00] J. Rován and V. Hayward, 2000. Typology of Tactile Sounds and their Synthesis in Gesture-driven Computer Music Performances. In *Trends in Gestural Control of Music*, pp. 355–368.
- [RL06] N. Raghuvanshi, , and M. C. Lin. Interactive Sound Synthesis for Large Scale Environments. In *Proc. of the International Symposium on Interactive 3D Graphics and Games (I3D)*, pages 101–108, 2006.
- [RM69] J. C. Risset and M. Mathews. Analysis of Musical Instrument Tones. *Physics Today*, 22:23–30, 1969.
- [Roa96] C. Roads. *The Computer Music Tutorial*. MIT Press, 1996.
- [RS96] D. Rochesso and F. Scalcon. Accurate Dispersion Simulation for Piano Strings. In *Proc. of the Nordic Acoustical Meeting*, pages 407–414, 1996.
- [RS99] D. Rochesso and F. Scalcon. Bandwidth of Perceived Inharmonicity for Physical Modeling of Dispersive Strings. *Transactions on Speech and Audio Processing*, 7(5):597–601, 1999.
- [Rui69] P. Ruiz. *A Technique for Simulating the Vibrations of Strings with a Digital Computer*. PhD thesis, University of Illinois, USA, 1969.
- [Rul95] G. Rule. Keyboard Reports: Korg Wavedrum. *Keyboard*, 21(3):72–78, 1995.
-

- [Sai06] J. H. Saiac, 2006. Analyse Numérique des Equations aux Dérivées Partielles, CNAM Course Material.
- [SDD90] M.-H. Serra, D. Dubine, and R. Dannenberg. Analysis and Synthesis of Tones by Spectral Interpolation. *Journal of Audio Engineering Society*, 38(3):111–128, 1990.
- [SHV97] L. Savioja, J. Huopaniemi, and R. Väänänen. Virtual Environment Simulation - Advances in the DIVA Project. In *Proc. of the International Conference on Auditory Display (ICAD)*, pages 43–47, 1997.
- [SK83] A. Strong and K. Karplus. Digital Synthesis of Plucked Strings and Drum Timbres. *Computer Music Journal*, 7(2):43–55, 1983.
- [Smi83] J. O. Smith. *Techniques for Digital Filter Design and System Identification with Application to the Violin*. PhD thesis, Stanford University, USA, 1983.
- [Smi87] J. O. Smith, 1987. Music Applications of Digital Waveguides. *Technical Report STAN-M-39*, CCRMA, Dept. of Music, Stanford University, USA.
- [Smi92] J. O. Smith. Physical Modeling using Digital Waveguides. *Computer Music Journal*, 16(4):74–91, 1992.
- [Smi98] J. O. Smith, 1998. Principles of Digital Waveguide Models of Musical Instruments. In *Applications of Digital Signal Processing to Audio and Acoustics*, pp. 417–466, Kluwer Academic Publishers.
- [Smi05] J. O. Smith. Virtual Acoustic Musical Instruments: Review and Update. *Journal of New Music Research*, 33(3):283–304, 2005.
- [SS90] X. Serra and J. Smith. Spectral Modeling Synthesis: a Sound Analysis/Synthesis System based on a Deterministic plus Stochastic Decomposition. *Computer Music Journal*, 14(4):12–24, 1990.
- [Str80] J. Strawn. Approximation and Syntactic Analysis of Amplitude and Frequency Functions for Digital Sound Synthesis. *Computer Music Journal*, 4(3):3–24, 1980.
- [SW08] S. Sinclair and M. M. Wanderley. A Run-time Programmable Simulator to Enable Multimodal Interaction with Rigid Body Systems. *Interacting with Computers*, 21(1-2):54–63, 2008.

- 
- [TDW03] C. Traube, P. Depalle, and M. M. Wanderley. Indirect Acquisition of Instrumental Gesture Based on Signal, Physical and Perceptual Information. In *Proc. of the International Conference on New Interfaces for Musical Expression (NIME)*, pages 42–47, 2003.
- [TH92] T. Takala and J. Hahn. Sound Rendering. *SIGGRAPH Computer Graphics*, 26(2):211–220, 1992.
- [the29] Radio-Victor Corporation of America, RCA Theremin Service Notes, 1929.
- [TVK98] T. Tolonen, V. Välimäki, and M. Karjalainen, 1998. Evaluation of Modern Sound Synthesis Methods. *Technical Report n° 48*, Helsinki University of Technology, Finland.
- [vdDKP01] K. van den Doel, P. Kry, and D. Pai. FoleyAutomatic: Physically-Based Sound Effects for Interactive Simulation and Animation. In *Proc. of the Annual conference on Computer Graphics and Interactive Techniques (SIGGRAPH)*, pages 537–544, 2001.
- [vdDKP04] K. van den Doel, D. Knott, and D. Pai. Interactive Simulation of Complex Audiovisual Scenes. *Presence: Teleoperators and Virtual Environments*, 13(1):99–111, 2004.
- [vDS93] S. van Duyne and J. O. Smith. Physical Modeling with the 2-D Digital Waveguide Mesh. In *Proc. of the International Computer Music Conference (ICMC)*, pages 40–47, 1993.
- [vDS96] S. van Duyne and J. O. Smith. The 3D Tetrahedral Digital Waveguide Mesh with Musical Applications. In *Proc. of the International Computer Music Conference (ICMC)*, pages 9–16, 1996.
- [VHKJ96] V. Välimäki, J. Huopaniemi, M. Karjalainen, and Z. Jánosy. Physical Modeling of Plucked String Instruments with Application to Real-time Sound Synthesis. *Journal of Audio Engineering Society*, 44(5):331–353, 1996.
- [vNWD04] D. van Nort, M. M. Wanderley, and P. Depalle. On the Choice of Mappings Based on Geometric Properties. In *Proc. of the International Conference on New Interfaces for Musical Expression (NIME)*, pages 87–91, 2004.
- [VT98] V. Välimäki and T. Tolonen. Development and Calibration of a Guitar Synthesizer. *Journal of Audio Engineering Society*, 46(9):766–778, 1998.
-

- [Wan01] M. M. Wanderley. *Interaction Musicien-Instrument: Application au Contrôle Gestuel de la Synthèse Sonore*. PhD thesis, Université Paris VI, France, 2001.
- [WD04] M. M. Wanderley and P. Depalle. Gestural Control of Sound Synthesis. *IEEE Special Issue on Engineering and Music - Supervisory Control and Auditory Communication*, 92(4):632–644, 2004.
- [Wei95] R. Weidenaar. *Magic Music from the Telharmonium*. The Scarecrow Press, 1995.
- [WFM03] M. Wright, A. Freed, and A. Momeni. OpenSound Control: State of the Art 2003. In *Proc. of the International Conference on New Interfaces for Musical Expression (NIME)*, pages 153–159, 2003.
- [Woo91] P. Wood. Recollections with John Robinson Pierce. *Computer Music Journal*, 15(4):17–28, 1991.
- [ZS06] M. Zadel and G. Scavone. Different Strokes: a Prototype Software System for Laptop Performance and Improvisation. In *Proc. of the International Conference on New Interfaces for Musical Expression (NIME)*, pages 168–171, 2006.
- [ZS08] M. Zadel and G. Scavone. Recent Developments in the Different Strokes Environment. In *Proc. of the International Computer Music Conference (ICMC)*, pages 1–4, 2008.

### C.3 Instrumental Performance Analysis

- [AR04] F. Avanzini and D. Rochesso. Physical Modeling of Impacts: Theory and Experiments on Contact Time and Spectral Centroid. In *Proc. of the International Conference on Sound and Music Computing (SMC)*, pages 287–293, 2004.
- [Ava01] F. Avanzini. *Computational Issues in Physically-based Sound Models*. PhD thesis, University of Padova, Italy, 2001.
- [BGS<sup>+</sup>07] F. Bevilacqua, F. Guédy, N. Schnell, E. Fléty, and N. Leroy. Wireless Sensor Interface and Gesture-follower for Music Pedagogy. In *Proc. of the International Conference on New Interfaces for Musical Expression (NIME)*, pages 124–129, 2007.
- [Bla95] J. Blades. *Percussion Instruments and their History*. Bold Strummer Ltd, 1995. Second Edition.
- [BWG08] A. Bouënard, M. M. Wanderley, and S. Gibet. Analysis of Percussion Grip for Physically Based Character Animation. In *Proc. of the International Conference on Enactive Interfaces (ENACTIVE)*, pages 22–27, 2008.
- [BWG09a] A. Bouënard, M. M. Wanderley, and S. Gibet. Analysis of Timpani Preparatory Gesture Parameterization. In *Proc. of the International Gesture Workshop (GW)*, pages 61–62, 2009.
- [BWG09b] A. Bouënard, M. M. Wanderley, and S. Gibet. Virtual Gesture Control of Sound Synthesis: Analysis and Synthesis of Percussion Gestures. *Acta Acustica united with Acustica*, 2009. Submitted.
- [BZS<sup>+</sup>09] F. Bevilacqua, B. Zamborlin, A. Sypniewski, N. Schnell, and F. Guédy, 2009. Continuous Realtime Gesture Following and Recognition. To appear in *Gesture in Embodied Communication and Human-Computer Interaction*, Vol. 5934, Springer Verlag.
- [Car68] E. Carter, 1968. Eight Pieces for Four Timpani (one player). In *Associated Music Publishers* (eds).
- [Coo97] G. Cook. *Teaching Percussion*. Schirmer Books, 1997. Second edition.
- [Dah97] S. Dahl, 1997. Spectral Changes in the Tom-Tom Related to the Striking Force. *Speech, Music and Hearing Quarterly Progress and Status Report*, KTH Royal Institute of Technology, Sweden.

- [Dah00] S. Dahl. The Playing of an Accent - Preliminary Observations from Temporal and Kinematic Analysis of Percussionists. *Journal of New Music Research*, 29(3):225–233, 2000.
- [Dah04] S. Dahl. Playing the Accent: Comparing Striking Velocity and Timing in Ostinato Rhythm Performed by Four Drummers. *Acta Acustica united with Acustica*, 90(4):762–776, 2004.
- [Dah05] S. Dahl. *On the beat: Human Movement and Timing in the Production and Perception of Music*. PhD thesis, KTH Royal Institute of Technology, Sweden, 2005.
- [Dah06] S. Dahl, 2006. Movements and Analysis of Drumming. In *Music, Motor Control and the Brain*, pp. 125–138, Oxford University Press.
- [Dem08] M. Demoucron. *On the Control of Virtual Violins: Physical Modelling and Control of Bowed String Instruments*. PhD thesis, Université Paris VI, France and KTH Royal Institute of Technology, Sweden, 2008.
- [DF07] S. Dahl and A. Friberg. Visual Perception of Expressiveness in Musicians’ Body Movements. *Music Perception*, 24(5):433–453, 2007.
- [GF88] S. Gibet and J. L. Florens. Instrumental Gesture Modeling by Identification with Time-Varying Mechanical Models. pages 28–40, 1988.
- [Gib87] S. Gibet. *Codage, Représentation et Traitement du Geste Instrumental*. PhD thesis, Institut National Polytechnique de Grenoble, France, 1987.
- [LP07] J. Loehr and C. Palmer. Cognitive and Biomechanical Influences in Pianists’ Finger Tapping. *Experimental Brain Research*, 178(4):518–528, 2007.
- [Noa84] F. W. Noak. *Timpani Sticks*. Percussion Anthology. The Instrumentalist, 1984. Third edition.
- [Pat08] D. Patterson. *John Cage: Music, Philosophy and Intention, 1933–1950*. Routledge, 2008.
- [Pet84] G. B. Peters. *Un-contestable Advice for Timpani and Marimba Players*. Percussion Anthology. The Instrumentalist, 1984. Third edition.
- [Ras08] N. Rasamimanana. *Geste Instrumental du Violoniste en Situation de Jeu : Analyse et Modélisation*. PhD thesis, Université Paris VI, France, 2008.

- [RB08] N. Rasamimanana and F. Bevilacqua. Effort-based Analysis of Bowing Movements: Evidence of Anticipation Effects. *Journal of New Music Research*, 37(4):339–351, 2008.
- [RBWB08] N. Rasamimanana, D. Bernardin, M. M. Wanderley, and F. Bevilacqua, 2008. String Bowing Gestures at Varying Bow Stroke Frequencies: A Case Study. In *Gesture-Based Human-Computer Interaction and Simulation*, LNCS 5085, pp. 216–226, Springer Verlag.
- [RF91] T. Rossing and N. Fletcher. *The Physics of Musical Instruments*. Springer Verlag, 1991.
- [RFB06] N. Rasamimanana, E. Fléty, and F. Bevilacqua, 2006. Gesture Analysis of Violin Bow Strokes. In *Gesture-Based Communication in Human-Computer Interaction*, LNAI 3881, pages 145–155, Springer Verlag.
- [Ros82] T. Rossing. The Physics of Kettledrums. *Scientific American*, 247(5):172–178, 1982.
- [Ros00] T. Rossing. *Science of Percussion Instruments*. Series in Popular Science, World Scientific, 2000.
- [Wag06] A. Wagner. Analysis of Drumbeats - Interaction between Drummers, Drumstick and Instrument. Master’s thesis, KTH Royal Institute of Technology, Sweden, 2006.
- [WVM<sup>+</sup>05] M. Wanderley, B. Vines, N. Middleton, C. McKay, and W. Hatch. The Musical Significance of Clarinetists’ Ancillary Gestures: an Exploration of the Field. *Journal of New Music Research*, 34(1):97–113, 2005.



*Bibliography*

---



## List of Figures

2.1	Kinematics and physics representations of a virtual character . . .	13
2.2	Motion capture systems . . . . .	18
3.1	Overview of the approach . . . . .	52
4.1	Timpani equipment . . . . .	56
4.2	Vibrating modes of a timpani membrane . . . . .	57
4.3	Radiation poles of a timpani membrane . . . . .	57
4.4	Mallet grips and beat impact locations . . . . .	59
4.5	Playing modes . . . . .	60
4.6	Timpani performer and the motion capture setup . . . . .	61
4.7	Semi-automatic detection of beat impacts . . . . .	63
4.8	Characterization of timpani playing techniques . . . . .	64
4.9	Position, velocity and acceleration profiles . . . . .	65
5.1	System architecture and multimodal outputs . . . . .	74
5.2	Physics-based modeling and control . . . . .	75
5.3	Hybrid physics-based motion control and synthesis . . . . .	78
5.4	Comparison of captured and simulated trajectories . . . . .	82
5.5	Comparison of RMS errors between captured and simulated mallet trajectories for various playing modes . . . . .	84
5.6	Asynchronous server-client architecture of our system . . . . .	86
5.7	User interface . . . . .	89
5.8	Visual feedback during the simulation . . . . .	90
6.1	Global approach to the composition process of virtual percussion performances. . . . .	94
6.2	Gesture edition and composition . . . . .	95
6.3	Simulated exercises: validation of playing modes and impact locations . . . . .	99
6.4	Simulated exercises: extrapolation of playing modes . . . . .	100

List of Figures

---

6.5	Simulated exercises: extrapolation of tempo variations . . . . .	101
6.6	Simulated exercises: extrapolation of impact location variations .	102
6.7	Comparison of captured and simulated beat impact locations played <i>legato</i> . . . . .	104
6.8	Instrumental gesture simulation: artifacts and their influence on sound synthesis . . . . .	106
A.1	d'Alembert formulation of the vibrating string . . . . .	117
A.2	Modal decomposition of the vibrating string . . . . .	120
A.3	Finite-element decomposition of the vibrating string . . . . .	124
A.4	Karplus-Strong delay line of the vibrating string . . . . .	126
A.5	Digital waveguide delay lines of the vibrating string . . . . .	127
B.1	Timpani MoCap protocol . . . . .	132
B.2	Timpani MoCap ethics form . . . . .	133

# E

## List of Tables

4.1	Bowl diameters and corresponding pitch intervals . . . . .	56
4.2	MoCap subjects playing characteristics . . . . .	62
4.3	Mallet trajectories and extracted extrema: statistical features . .	66
4.4	SVM recognition of timpani grips . . . . .	67
4.5	SVM recognition of <i>French</i> grip playing modes . . . . .	67
4.6	SVM recognition of <i>German</i> grip playing modes . . . . .	67
4.7	KNN recognition of timpani grips . . . . .	70
4.8	KNN recognition of <i>French</i> grip playing modes . . . . .	70
4.9	KNN recognition of <i>German</i> grip playing modes . . . . .	70
5.1	SVM recognition of simulated <i>French</i> grip playing modes . . . . .	85
6.1	Timpani playing notation . . . . .	97

*List of Tables*

---

# F

## List of Equations

2.1	Virtual character kinematic representation . . . . .	12
2.2	Virtual character kinematic configuration . . . . .	12
2.3	Virtual character dynamic representation . . . . .	14
2.4	Virtual character dynamic configuration . . . . .	14
2.5	Forward kinematics formulation . . . . .	15
2.6	Inverse kinematics formulation . . . . .	16
2.7	Jacobian-based inverse kinematics formulation . . . . .	16
2.8	Jacobian-based inverse kinematics formulation, secondary tasks .	17
2.9	Forward dynamics formulation . . . . .	21
2.10	Newton motion laws . . . . .	21
2.11	Inverse dynamics formulation . . . . .	22
2.12	<i>PD</i> control formulation . . . . .	25
2.13	<i>AA</i> control formulation . . . . .	26
2.14	<i>AA</i> control formulation and Equilibrium-Point control . . . . .	26
2.15	Spectral additive synthesis . . . . .	33
2.16	1-D d'Alembert equation . . . . .	34
2.17	1-D d'Alembert travelling waves solution . . . . .	35
2.18	Digital waveguide formulation . . . . .	36
2.19	Modal synthesis formulation . . . . .	38
2.20	Modal synthesis formulation, linear combination of normal modes	38
5.1	<i>ID</i> control involved in the hybrid motion control . . . . .	78
5.2	<i>IK</i> control involved in the hybrid motion control . . . . .	78
5.3	Modal synthesis formulation under the influence of an external force	88
5.4	Modal synthesis formulation under the influence of an external force: normal formulation . . . . .	88





## List of Algorithms

5.1	Formulating joint limits in the quaternion space . . . . .	77
5.2	Hybrid motion control combining <i>IK</i> and <i>ID</i> controllers . . . . .	79







## Résumé

Ces dernières années ont vu l'émergence de nombreuses interfaces musicales ayant pour objectif principal d'offrir de nouvelles expériences instrumentales. La spécification de telles interfaces met généralement en avant l'expertise des musiciens à appréhender des données sensorielles multiples et hétérogènes (visuelles, sonores et tactiles). Ces interfaces mettent ainsi en jeu le traitement de ces différentes données pour la conception de nouveaux modes d'interaction.

Cette thèse s'intéresse plus spécifiquement à l'*analyse*, la *modélisation* ainsi que la *synthèse* de situations instrumentales de percussion. Nous proposons ainsi un système permettant de synthétiser les retours visuel et sonore de performances de percussion, dans lesquelles un percussionniste virtuel contrôle des processus de synthèse sonore.

L'étape d'analyse montre l'importance du contrôle de l'extrémité de la mailloche par des percussionnistes experts jouant de la timbale. Cette analyse nécessite la capture préalable des gestes instrumentaux de différents percussionnistes. Elle conduit à l'extraction de paramètres à partir des trajectoires extrémité capturées pour diverses variations de jeu. Ces paramètres sont quantitativement évalués par leur capacité à représenter ces variations.

Le système de synthèse proposé dans ce travail met en oeuvre l'animation physique d'un percussionniste virtuel capable de contrôler des processus de synthèse sonore. L'animation physique met en jeu un nouveau mode de contrôle du modèle physique par la seule spécification de la trajectoire extrémité de la mailloche. Ce mode de contrôle est particulièrement pertinent au regard de l'importance du contrôle de la mailloche mis en évidence dans l'analyse précédente. L'approche physique adoptée est de plus utilisée pour permettre l'interaction du percussionniste virtuel avec un modèle physique de timbale.

En dernier lieu, le système proposé est utilisé dans une perspective de composition musicale. La construction de nouvelles situations instrumentales de percussion est réalisée grâce à la mise en oeuvre de partitions gestuelles. Celles-ci sont obtenues par l'assemblage et l'articulation d'unités gestuelles canoniques disponibles dans les données capturées. Cette approche est appliquée à la composition et la synthèse d'exercices de percussion, et évaluée qualitativement par un professeur de percussion.

**Mots clé :** Animation par Ordinateur, Informatique Musicale, Interaction Geste Instrumental-Son.

## Abstract

The last few decades have witnessed the emergence of a plethora of musical interfaces aiming at expanding musical performance experiences. The design of these interfaces generally emphasizes the expertise of musicians in managing heterogeneous sensory informations (visual, sound and tactile). Such musical interfaces involve therefore the processing of these various sensory data for designing novel interaction modes.

This thesis addresses more specifically the *analysis*, *modeling* and *synthesis* of percussion performances. We propose a system that realizes the synthesis of the visual and sound feedback of percussion performances in which a virtual percussionist controls sound synthesis processes.

The analysis step of our work shows the importance of the fine control of mallet extremity trajectories by expert percussion performers playing timpani. It includes the collection of instrumental gesture data from several percussionists. We extract movement parameters from the recorded mallet extremity trajectories for different percussion playing variations. Such parameters are quantitatively evaluated with respect to their ability to represent the various playing variations under study.

We then propose a system for synthesizing timpani performances involving the physical modeling of a virtual percussionist that interacts with sound synthesis processes. The physical framework includes a novel scheme for controlling the motion of the virtual percussionist solely by the specification of trajectories of mallet tips. This control mode is shown to be consistent with the predominant control of mallet extremity presented in the previous analysis step. The physical approach is also used for allowing the virtual percussionist to interact with a physical model of a timpani.

Finally, the proposed system is used in a musical performance perspective. A composition process based on gesture scores is proposed to achieve the synthesis of novel percussion performances. Such gesture scores are obtained by the assembly and articulation of gesture units available in the recorded data. This compositional approach is applied to the synthesis of several percussion exercises, and is informally evaluated by a percussion professor.

**Key words:** Computer Animation, Computer Music, Instrumental Gesture-Sound Interaction.

