Playing (with) Sound

Of the Animation of Digitized Sounds and their Reenactment by Playful Scenarios in the Design of Interactive Audio Applications

Dissertation by

Norbert Schnell

Submitted for the degree of

Doktor der Philosophie

Supervised by

Prof. Gerhard Eckel
Prof. Rolf Inge Godøy

Institute of Electronic Music and Acoustics
University of Music and Performing Arts Graz, Austria
October 2013

Abstract

Investigating sound and interaction, this dissertation has its foundations in over a decade of practice in the design of interactive audio applications and the development of software tools supporting this design practice. The concerned applications are sound installations, digital instruments, games, and simulations. However, the principal contribution of this dissertation lies in the conceptualization of fundamental aspects in sound and interactions design with recorded sound and music.

The first part of the dissertation introduces two key concepts, *animation* and *reenactment*, that inform the design of interactive audio applications. While the concept of *animation* allows for laying out a comprehensive cultural background that draws on influences from philosophy, science, and technology, *reenactment* is investigated as a concept in interaction design based on recorded sound materials.

Even if rarely applied in design or engineering – or in the creative work with sound – the notion of animation connects sound and interaction design to a larger context of artistic practices, audio and music technologies, engineering, and philosophy. Starting from Aristotle's idea of the *soul*, the investigation of animation follows the parallel development of philosophical concepts (i.e. soul, mind, spirit, agency) and technical concepts (i.e. mechanics, automation, cybernetics) over many centuries. Particular attention is hereby paid to sound reproduction and music technologies.

The concept of reenactment is rooted in the ideas of *enaction* and *embodied cognition*. From this perspective, *listening* is examined in relationship with further interactions that support or accompany listening. An investigation of sound reproduction technologies insists in particular on the concept of *congruency*, as a central factor in the understanding of interaction. Sound and interaction design here appears as the composition of – technical and semantic – congruencies that define the users' engagement into interaction with an application. After an exploration of epistemological and hermeneutical aspects of interactive technologies, the reenactment concept is further enriched by the investigation of related concepts such as *metaphor*, *affordance*, and *play*. Finally, *playing technique* and *scenarios* are proposed as fundamental concepts in sound and interaction design that, beyond affordance, allows for taking into account cultural factors as well as the notion of practice.

The second part of the dissertation presents nine interactive audio applications. All of the applications have been realized in the past few years in collaborations on artistic and pedagogical projects or as demo applications showcasing design ideas and technologies. The descriptions of the applications show how the concept of *reenactment* applies to the scenarios of interactive audio applications and how the concept allows for efficiently highlighting important aspects of interaction design.

Kurzfassung

(Mit) Klang spielen – Über die Animation digitalisierter Klänge und ihr Reenactment durch spielerische Szenarien in der Gestaltung interaktiver Klangapplikationen

In dieser Dissertation werden Konzepte der Gestaltung interaktiver Applikationen mit digitalisierten Klängen untersucht. Die Aufmerksamkeit gilt insbesondere den beiden Schlüsselbegriffen *Animation* und *Reenactment* (wörtlich aus dem Englischen: *Wieder-in-Handlung-setzen*), die es erlauben, fundamentale Aspekte der Gestaltung interaktiver Klangprozesse auf der Basis von Musik- und Klangaufnahmen zu erfassen. Die Untersuchung stützt sich auf eine langjährige Praxis in der Gestaltung von Klanginstallationen, digitalen Musikinstrumenten und Spielen sowie Klangsimulationen. Sie präsentiert eine Argumentation, die sich im Wesentlichen aus der Auseinandersetzung mit Literatur aus Philosophie und Kulturwissenschaften, Wissenschaft und Technik entwickelt.

Neben einer Einleitung (Kapitel 1) und der abschließenden Zusammenfassung (Kapitel 4) umfasst die Arbeit zwei Hauptkapitel. Im ersten dieser Kapitel (2), das den wesentlichen Teil der Dissertation ausmacht, werden umfassend die Schlüsselbegriffe, *Animation* und *Reenactment* bearbeitet, die im darauffolgenden Kapitel (3) anhand der Gestaltung von neun Klanginstallationen und Applikationen angewendet und verfeinert werden. Während der Begriff der *Animation* hauptsächlich dazu dient das historische und aktuelle kulturelle Umfeld der Gestaltung interaktiver Prozesse mit Klang sowie der verwendeten Technologien zu beleuchten, wird der Begriff *Reenactment* zunächst als ein präzises Konzept im Bereich der Gestaltung mit digitalisierten Klängen vorgestellt, das anschließend durch die Untersuchung weiterer nahestehender Begriffe bereichert wird.

In der Argumentation zu Beginn des zweiten Kapitels wird *Animation* zunächst mit wesentlichen Eigenschaften von Klang in Verbindung gebracht, welcher stets in direkter Beziehung zu lebhafter Bewegung und Interaktion steht. Jedoch ist Animation, im allgemeinen Sinn sowie im Bereich der künstlerischen Gestaltung, auch stets dem Unbewegten und Nichtlebendigen zugewandt – eben um es zu *animieren*. Zur Erkundung dieses Aspektes wird zunächst kulturwissenschaftliche Literatur herangezogen, die sich insbesondere mit Technologien der Klangreproduktion und ihres geschichtlichen und kulturellen Umfelds beschäftigt. Diese Technologien werden im Weiteren unter dem Blickwinkel des Begriffs der *Automation* betrachtet. *Animation* und *Automation* erscheinen hierbei als teils gleichbedeutende, teils gegensätzliche Begriffe. Sie erlauben es, wesentliche Aspekte von Technikentwicklung von einem kulturellen Standpunkt aus zu untersuchen, auch über klangbezogene Technologien hinaus. Für diese Untersuchung ist es notwendig, weitergehend auf den kulturellen und philosophischen Hintergrund des Begriffs der Animation im Sinne von *Beseelung*

einzugehen. Ausgehend von Ideen, die Aristoteles in seinem Werk *Über die Seele* entwickelt hat, werden hierzu parallele Entwicklungen philosophischer Konzepte (z.B. Seele, Geist, freier Wille) und technischer Konzepte (z.B. Mechanik, Automation, Kybernetik) über mehrere Jahrhunderte verfolgt. Dieser historische Überblick endet mit dem gleichzeitigen Aufkommen der Phänomenologie mit – zunächst analogen – Informationstechnologien und der Aufgabe der konzeptuellen Trennung von *Körper* und *Geist* (oder *Seele*). Die Ausführungen münden hier in eine Untersuchung der Begriffe der Aktion und Interaktion, die auf phänomenologischen Standpunkten sowie weiteren philosophischen und sozialwissenschaftlichen Strömungen beruht und eine wichtige Grundlage für die weitere Argumentation der Dissertation darstellt. Hier werden insbesondere Deweys Begriff der *Transaktion* und Latours *Akteur-Netzwerke* eingeführt, die im zweiten Teil des Kapitels weiter ausgeführt und angewendet werden. Im Lichte dieser Ausführungen erscheinen *Animation* und *Interaktionsgestaltung* (engl.: *interaction design*) als synonyme Begriffe. Die Zusammenfassung philosophischer und kulturgeschichtlicher Einflüsse dieses ersten Teils des Kapitels schließt mit einem Exkurs in kulturwissenschaftliche Studien, die den Begriff der Animation schließlich in den Zusammenhang mit dem Zeichentrickfilm stellen. In diesem Epilog, der insbesondere das narrative Potenzial von Animation betrifft, wird der Begriff mit dem Erforschen der Grenzen zwischen dem *Lebendigen* oder auch *Menschlichen* und dem *Leblosen* oder *Mechanischen* in Verbindung gebracht.

Der zweite Teil des Kapitels betrifft den Begriff des *Reenactment* als ein Schlüsselkonzept in der Gestaltung interaktiver Prozesse mit aufgenommenen Klängen. Durch das Konzept werden die *Aktionen* des Benutzers einer interaktiven Applikation (z.B. *Zuhörer* oder *Spieler*) im Zusammenhang mit den Aktionen betrachtet, die von den verwendeten Klangmaterialien evoziert werden (Gesten und Bewegungen, konkrete Ereignisse und Handlungen, Praktiken sowie weitere Konnotationen). *Reenactment* wird zunächst aus dem Blickwinkel von *Enaction* und *Embodied Cognition* untersucht, die auf der Grundlage phänomenologischer und psychologischer Konzepte seit der zweiten Hälfte des 20. Jahrhunderts wesentliche philosophische, wissenschaftliche und technische Entwicklungen beeinflusst haben. Zunächst wird die Verflechtung von Perzeption, Aktion und Kognition erörtert, bevor insbesondere die Aktivität des Hörens und ihre Wechselwirkung mit weiteren Handlungen, die entweder vom Hören ausgehen oder das Hören begleiten, untersucht wird. Ein weiterer Gegenstand sind die Interaktionen, die im Bereich der Klangreproduktion auftreten, wobei alle möglichen materiellen und kulturellen Interaktionen in Betracht gezogen werden, unabhängig davon, ob sie Aufnahme oder Wiedergabe, die Interaktion zwischen Menschen oder die von Menschen mit Technologien betreffen. Besondere Aufmerksamkeit gilt hier dem Konzept der *Reproduktion* selbst sowie dem der Kongruenz, der eine Reflexion über die Wahrnehmung von Kausalitäten und Wechselwirkungen in sehr verschiedenen Zeitbereichen ermöglicht. Diese umfassen gleichermaßen Klangsignale, Bewegung und Gesten sowie Handlungen und Phänomene, die sich über größere Zeiträume entwickeln. Von diesem Standpunkt aus erscheint die Gestaltung von Interaktionen als eine Frage der Komposition von Kongruenzen, in der die Komplementarität, Auslöschung und Verstärkung vielfältiger technischer und kultureller Kongruenzen schließlich die Interaktion mit einer gegebenen Applikation und den von ihr präsentierten Akteuren (engl.: *actors*) definieren. Am Ende dieser Argumentation wird *Reenactment* also als eine aktionsbezogene Erweiterung des Begriffes der Reproduktion definiert – Reproduktion *in* und *durch* Aktion.

Diesen Ausführungen, die hauptsächlich Technik und Wahrnehmung betreffen, folgen weitergehende Untersuchungen epistemologischer und semiotischer Aspekte der Gestaltung interaktiver digitaler Technologien. Hier stützt sich die Argumentation auf eine kritische Auseinandersetzung mit der Unterscheidung von *Inskriptions-* und *Inkorporationspraktiken* (engl.: *inscription and incorporation practices*), wie sie in kultur- und sozialwissenschaftlichen Studien vorgenommen wird. Diese Unterscheidung scheint im Falle interaktiver digitaler Applikationen nicht anwendbar zu sein, da bei ihrer Gestaltung sowie in der Interaktion mit ihnen beide Sichtweisen, *Inskription* und *Inkorporation*, stets weitgehend verwoben sind. Jedoch erlaubt der Vergleich interaktiver – aktionsbezogener – Techniken mit repräsentationsbezogenen Techniken die Untersuchung hermeneutischer Aspekte interaktiver Applikationen. Interaktive digitale Technologien erscheinen aus diesem Blickwinkel als ein Medium, das eine *Reflexion* durch *nichtsprachliche* und *multimodale* Interaktion ermöglicht. Diese Reflexion umfasst auch Elemente wie *Assoziation*, *Abstraktion*, *Hypothese*, *Demonstration* und *Übermittlung*, die üblicherweise repräsentationsbezogenen Praktiken zugeordnet werden.

Ähnlich der Erweiterung des Begriffs der Reproduktion zuvor wird *Reenactment* hierdurch als eine aktionsbezogene Erweiterung von *Reproduktion* verstanden – Repräsentation *in* und *durch* Aktion. Um die Prinzipien der *Reflexion durch Interaktion* weiter zu hinterfragen, werden im Folgenden die Konzepte *Metapher* und *Affordanz* (engl.: *affordance*) eingeführt, die wiederum die Auseinandersetzung mit dem *Geschriebenen* und dem *Agierenden* widerspiegeln. Die Gestaltung interaktiver Applikationen wird hiermit zu einem Wechselspiel beider Sichtweisen, wobei insbesondere die Anwendung und das Verständnis von Metaphern als ein wichtiges Element in der Gestaltung sowie der Beobachtung von Interaktionen erscheint.

Dieser zweite Teil schließt mit einer Untersuchung des Begriffs der *Spielweisen* – zusammen mit *Strategien* und *Szenarien* – als ein Faktor und Studienobjekt in der Gestaltung von Interaktionen. Während sich die Gestaltung von Applikationen anstatt mit Repräsentationen vorwiegend mit tatsächlicher – vorhandener oder erwarteter – Aktion beschäftigt, geben Spielweisen und Szenarien die Gelegenheit, *Möglichkeiten* solcher Aktion zu betrachten. Spielweisen erlauben hierbei, über den Begriff der Affordanz hinaus, die Einbeziehung sozialer und kultureller Faktoren sowie die Berücksichtigung bestehender Praktiken und geben den Anlass, den Begriff des *Spielens* und der spielerischen Auseinandersetzung genauer zu beleuchten.

Im dritten Kapitel werden neun interaktive Klangapplikationen vorgestellt. Bei diesen Projekten, die alle in den letzten Jahren realisiert wurden, handelt es sich um Klanginstallationen sowie um Anwendungen, die entwickelt wurden um technische Neuentwicklungen und Gestaltungsideen zu demonstrieren. Alle Applikationen wurden bereits einem größeren Publikum präsentiert und einige waren bereits der Gegenstand eigenständiger wissenschaftlicher Publikationen. Während der erste Teil des Kapitels sieben Applikationen und ihre Szenarien unter dem Blickwinkel der Interaktion mit musikalischen Inhalten durch metaphorische Handlungen präsentiert, werden die im zweiten Teil beschriebenen Projekte als eine Auseinandersetzung mit dem Begriff des Spielens behandelt, indem musikalisches Spiel anderen Formen von Spiel – im Sinne von *Sport und Spiel* – gegenübergestellt und mit ihnen verwoben wird.

Die Beschreibungen in diesem Kapitel tragen im doppelten Sinne dem Verständnis der Abhandlungen des vorherigen Kapitels bei. Einerseits geben sie die Möglichkeit, die entwickelten Konzepte auf die Praxis der Gestaltung anzuwenden und dabei weiter zu verfeinern. Andererseits werden die Begriffe in den Beschreibungen auf die Probe gestellt und müssen ihre Effizienz in der Betrachtung von und Reflexion über Applikationen und den enthaltenden Interaktionen beweisen.

Über eine rein theoretische Untersuchung der Gestaltung interaktiver Klangapplikationen hinaus ist diese Dissertation tief in der Praxis dieser Gestaltung verwurzelt. Das Anliegen dieser Dissertation ist es, durch die Untersuchung grundlegender Konzepte auch zur praktischen Arbeit beizutragen. Die Perspektiven, die sich durch die Abhandlungen dieser Dissertation ergeben, betreffen somit nicht nur Studien der Gestaltung von Interaktion sowie die Betrachtung von Interaktion und Interaktivität selbst, sondern auch die Praxis der Gestaltung – Bereiche, die in interdisziplinären Forschungsarbeiten, schwer zu trennen sind. Insbesondere die Idee der Reflexion durch nichtsprachliche multimodale Interaktion und die Rolle von Metaphern sollen hierbei wegweisende Impulse für zukünftige Arbeiten zu geben.

Danksagung

Meinem Doktorvater Prof. Gerhard Eckel, der sich diesen Titel geduldig verdienen musste und mir in vielem wegweisend war, sowie – dem -*Onkel* – Prof. Rolf Inge Godøy für seinen kostbaren Rat und seine lehrreichen Beiträge zum Fach.

Den "Grazer" Kollegen, Freunden und Wegbegleitern Martin Rumori, Thomas Grill, David Pirrò, IOhannes Zmölnig, Johanna Gampe, Hanns Holger Rutz und Peter Plessas, die entweder schneller waren oder als nächstes dran sind...

Meinen lieben Kollegen der IMTR Forschungsgruppe am IRCAM, Frédéric Bevilacqua, Riccardo Borghesi, Diemo Schwarz, Emmanuel Flety, Jean-Philippe Lambert und Andrea Cera, ohne die es vieles was hier beschrieben ist nicht gäbe, sowie den Ehemaligen, Baptiste Caramiaux, Nicolas Rasamimanana, Julien Bloit, Tommaso Bianco und Bruno Zamborlin, die auch tatkräftig dabei waren. Fred ganz besonders für alles was wir zusammen geschafft haben, seine Freundschaft und seine Unterstützung.

Meinen Kollegen und Projektpartnern, Christian Graupner, Fabrice Guédy, Roland Cahen, Xavier Boissarie, Jean-Louis Frechin, Uros Petrevski und Atau Tanaka, die vieles inspiriert und beigetragen haben.

Meinen *Engeln*, Seth Thorn und Antony Hequet, die mir geholfen haben (über das Englische hinaus) so einiges zu richten.

Meinen ersten wohlwollend Zuhörern, Butch Rovan (auch mein Retter in der Not) und Katherine Bergeron, denen nach all den Jahren noch vieles bekannt vorkommen könnte.

Meinen teuren Freunden Peter Böhm und Corinne Schweizer, die schon bei den ersten Projekten dabei waren.

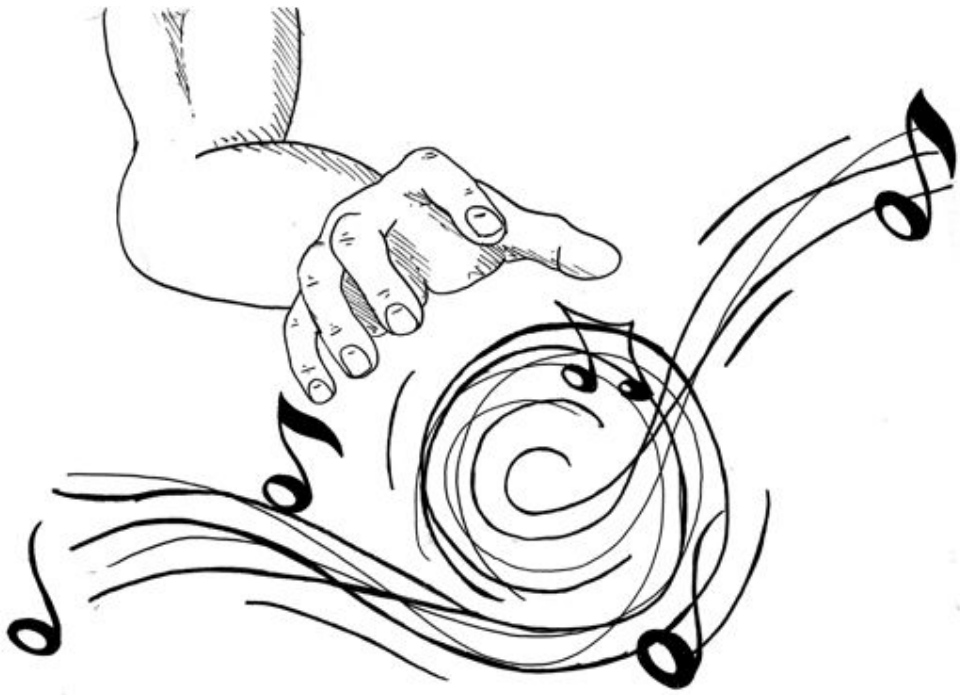
Meinem treusten Freund Wolf K., der vielleicht gar nicht so genau wissen wollte, was Reenactment ist.

Meinem Sohn Arthur, der schon vor einigen Jahren alles in einer Zeichnung zusammen gefasst hat, und Antonia, für Ihre Hilfe bei der Korrektur.

Meiner Mutter, die vorgemacht hat wie man sich noch später in Richtung neuer intellektueller Horizonte und akademischer Grade aufmacht, und meinem Vater für seine bedingungslose Unterstützung.

Meinen Schwiegereltern, Lino und Vittoria, die mich herzlich aufgenommen haben, auch wenn sie mich bisher vorwiegend als jemanden kennen, der hinter seinem Notebook auf ihrem Sofa sitzt und an seiner Dissertation arbeitet.

Meiner Frau Pina für ihre Liebe und Geduld, ohne die hier überhaupt nichts stehen würde.



Sonic Action (ArtKub, 2008)

Contents

CHAPTER 1

Introduction

1.1 Context and Motivation

This dissertation is about sound and interaction. It examines key concepts that inform the design of interactive audio applications such as audio players and displays, sound installations, digital musical instruments, games, and simulations. The practice of sound and interaction design that has been the center of my work for almost two decades essentially consists in creating applications that put the *bodily movements and gestures*[1] of their users into relationship with *digital sound processes* based on *recorded sound materials*.

In many of these applications, digital technologies may appear as a form of mediation that enables users to explore recorded sounds and the sound scenes that they represent. Interactive audio players and audio displays, for example, employ digital processing to navigate through audio content and adapt the recorded scenes to the listening situation. Sound transformations here allow for creating different perceptual perspectives on the reproduced recordings and for focusing on particular details. This category of media also includes virtual or augmented reality environments where recorded materials appear as the sound produced by fictional characters, objects, and landscapes.

A very different relationship between the players' actions and recorded sound, generally constitutes the design of digital musical instruments. Through digital processing the articulation of sound here becomes an integral part of the players' actions of performing an instrument as it is known from acoustic instruments. While the first category of applications primarily supports *listening* as the exploration of sound and music mediated through digital technologies, the players of digital musical instruments primarily *perform*.

However, these categories of sound-based interactions are challenged by the design of many applications, where listening, exploring, and performing converge into complex interactions that often involve multiple players. In the legacy of hip-hop, sophisticated audio players support functionalities that, beyond listening, allow for remixing and restaging recorded sound and music. Moreover, the design of digital musical instruments does not always follow the example of acoustic instruments. Many digital instruments and performance oriented interactive applications develop recorded sound materials into generative sound processes that are controlled

[1] In this dissertation, the combined term *movements and gestures* is used consistently without further defining and distinguishing *movement*, *gesture*, and *posture*.

by their players. The sound and interaction design of such instruments is very similar to that of sound installations, music and audio games that, rather than supporting performance on stage, directly invite their audience into playful interactions. All of these applications finally have a double affinity to musical instruments and sound reproduction technologies. They inherit essential functions and techniques from the *violin* as well as from the *gramophone*. Notions like performing, listening, exploring, controlling, and navigating here become complementary perspectives on the players engagement into interaction in which playing a record, playing an instrument, and playing a game converge into *playing (with) sound*.

Interactive audio and multimedia applications have evolved over the past decades along with digital technologies like personal and mobile computing platforms, motion capture and network technologies, as well as advanced signal processing techniques. Through miniaturized motion sensing and wireless technologies, the interaction with these applications involves more and more the entire human body and can transparently integrate into our interactions with everyday objects and environments.[2] In the design of interactive audio applications, these technologies provide malleable materials for creating applications that transcend the established paradigms of *sound reproduction*, *musical instruments*, *virtual reality*, and *simulation*. In this context, designing with digital technologies becomes a play with interactions that may evoke metaphorically the interactions within these paradigms as well as interactions known from everyday life, existing sociocultural practices, and fiction. Different horizons like music, games, sports, and communication converge into playful scenarios composed of *real* and *virtual* elements. This trend is shared throughout many domains including art, digital games, pedagogy, and industrial design.

Beyond the design of interactive audio applications, the body of work that has contributed to the elaboration of this dissertation includes the development of techniques for the analysis and synthesis of sound, movement, and music. Most of this work has been achieved in the framework of interdisciplinary research and development, as well as artistic projects. The majority of these projects have been the object of independent publications in the domain of music technology and human computer interaction. These publications primarily concern the development of techniques used in designing interactive sound and music applications. However, this dissertation, rather than focusing on particular techniques used in design, reflects on design itself.

Although it includes the confrontation with a considerable corpus of literature, in its essence this dissertation relies on the examination of my own practice of design that has evolved in constant exchange with colleague technologists, designers, and artists. An important challenge in writing this dissertation was to find a coherent perspective on design, research, and develop-

[2] Concepts that focus on different aspects emerging in this context include *embodied interaction*, *pervasive computing*, *tangible interfaces*, *natural interfaces*, *communicating objects*, and the *Internet of things*.

ment that also permits me to communicate my fascination for this work, in general as well as each of the accomplished projects and many of their details. An idea that has particularly fascinated me in the work with recorded materials is that they always very concretely refer to events that actually have happened at a specific moment and place when and where somebody has recorded them. The recorded events have involved particular actors that – whether they are recognizable or not – have been present at the scene of recording. At the same time, playing back recorded sound in absence of these actors has a strong power to inspire the listeners' *imagination* and sense of *abstraction*. Not only when listening to music as *organized sound*, as Varèse put it, but also when discovering unfamiliar scenes, rather than as the actions of recognizable actors, sound may be experienced in terms of abstract sonic *gestalts*.

Sound – played back from a recording or *live* – resonates in all aspects of our existence. It is physical vibration, arousal and affect, entrains our action and evokes our memories. Beyond these rather personal *reactions* that are part of our listening experience, recorded sound is also a cultural artifact. This applies in particular to recordings that not only capture arbitrary sound scenes, but specific *events* and *practices* such as music recordings, recordings of speech, or ethnographical recordings. Music recordings, not only represent a particular – sometimes even *historical* – moment when musicians gathered at a particular place to perform, and they do not just make us feel and dance, or remind us of something that has happened the last time we listened to the same recording. They also represent a particular interpretation of a particular piece and refer to a practice of composition and performance, an epoch, style, technique, and formal aspects of music. They imply cultural, social, and political connotations, for example when considering a particular hymn or dance. In addition, the act of recording, being itself a specific event and practice, comes with its own equally rich network of references that are associated with the recording. Any approach to sound design based on recorded sound materials has to position itself towards these manifold aspects associated with recorded sounds that may be *preserved* or *transformed*, *attenuated* or *amplified*, *neutralized* or *combined* in a particular design.

While from one point of view, recorded sound materials appear as overloaded with *significations*, from another point of view they can be simply regarded as *materials*. Sound materials have particular qualities that precede the design as a system of opportunities and constraints, similar to how physical materials afford and constrain the design of musical instruments, tools, or furniture. Sound materials may be, percussive or sustained, pitched or noisy, harmonic or dissonant, brilliant or dull, ascending or descending, fast or slow, periodic, chaotic, or static. Both, significations and material qualities of sound are part of the experience of listening.[3]

[3] The manifolds aspects of listening have been investigated by authors like Pierre Schaeffer, Murray Schafer, and Roland Barthes (Schaeffer 1966; Schafer 1977; Barthes 1982). Related to his idea of an *acoustic ecology*, Schafer (1994) explicitly underlines the crucial role of listening in sound design, regarding it as a '*matter of the retrieval of a significant aural culture*'.

This dissertation investigates sound and interaction design as it emerges from listening. Design here becomes an act of exploring and sharing perspectives on sound and music that are, inevitably, also perspectives on listening. The engagement into exploration that designers share with the users of their designed products has been compellingly described by Gaver (2009) in his manifesto for *ludic design*.

> *This is an engagement that has no fixed path or end, but instead involves a wide-ranging conversation with the circumstances and situations that give it rise. Rules may emerge and goals may be sought, but these will be provisional inventions, makeshift tools to help the advance of curiosity and exploration.*[4]

1.2 Related Domains and Practices

The sound and interaction design of audio applications concerned by this dissertation shares concepts and techniques with many other practices in which technologies mediate the interaction *with* or *through* digitized sound. When considering each of its aspects separately, further similarities and relationships to other practices appear.

From a rather technical perspective, the design of an interactive audio application, essentially consists of the following three elements:

- sound recording, analysis and annotation
- movement capture and analysis
- real-time interactive audio rendering

The first two elements of this list are concerned with representing the recorded sound materials and the users' gestures and movement within the digital system of the application. The recording, analysis, and annotation of recorded sounds can be seen as an extension of listening that is instrumented through audio recording and automatic analysis techniques – also *machine listening* techniques – as well as through tools that allow for annotating recorded sound materials manually. Whether the recordings used by an application have explicitly been recorded for the application or not, the sound and interaction design of interactive audio applications based on recorded sounds relies on *audio recording* practices and technologies that generally include *audio editing* and *post-production*.

The automatic analysis of sound is the object of research and development in the field of digital *signal processing* specifically concerned with audio signals (see for ex. Peeters 2004; Jehan 2005). Digital audio processing is strongly informed by research in *acoustics* and *psychoacoustics* (see for ex. Möser 2009; Meyer 2004; Fastl and Zwicker 2007). Two specific fields dedicated to the analysis of particular audio signals are *music information retrieval* (Fingerhut

[4] In (Gaver 2009).

2004) and *speech recognition* (see for ex. Pieraccini 2012).[5] The development of automatic analysis techniques in these fields may be informed by studies in *musicology* and *linguistics*.[6] Besides musicology, considerable contributions to the formalization of recorded sound have been made by composers (see for ex. Schaeffer 1966; Schafer 1977).

The capture and analysis of movement and gesture is a technical element that interactive audio applications share with many other interactive applications. Motion capture technologies and techniques for the estimation of perceptual parameters, as well as the recognition of gestures and postures, are the object of studies in the domain of human-computer interaction (see for ex. Camurri and Volpe 2004; Gibet et al. 2006). However, the investigation of specific movements and gestures necessary for creating meaningful digital representations, is distributed over multiple fields such as *physiology*, *neuroscience*, and *psychology*.[7] Moreover, interdisciplinary studies generally include the domain in which particular movements and gestures occur such as *music*, *linguistics*, *dance*, and *sports*. The gestures and movements related to music performance, listening, and dance are increasingly studied in the field of *musicology* (see for ex. Godøy and Leman 2010), but also in *music technology* (see for ex. Müller 2007), and especially in relationship to the design of digital musical instruments (see for ex. Miranda and Wanderley 2006).

The third of the three technical elements of the design of interactive audio applications as they are listed above is concerned with the mediation of the first two elements. This mediation essentially consists in rendering[8] the recorded sound materials in response to the users' movements and gestures. In addition to the actual interactive *playback*, *analysis/resynthesis*, and *transformation* of the recorded sound materials, audio rendering may integrate generative processes like *simulations* and *automations*. Further processing such as sound *effects* and *spatialization* allows for controlling how the generated sound is projected into the users' auditory space. While many of the involved techniques belong to the domain of *digital audio processing*, the design of generative processes involved in audio rendering relies on models and formaliza-

[5] Growing out of the field of speech processing, music information retrieval has been established as an independent field of research and development in the 1990s (Fingerhut 2004).

[6] While speech processing always has been closely related to studies in linguistics, studies in music information retrieval, have been often oriented towards music distribution rather than musicology. Nevertheless, especially lately, also musicology and music information retrieval have started to converge (see for ex. Neubarth et al. 2011; von Loesch and Weinzierl 2011).

[7] The *International Society for Gesture Studies* has started about a decade ago to establish *Gesture Studies* as an independent interdisciplinary field of research.

[8] The term *rendering* is consistently used to refer to the digital real-time processing sub-system of an application that deals with the generation of sound. Generally, this sub-system can be relatively easily distinguished from the sub-system analyzing the users' movement and gestures. However, in certain cases it also can be difficult to make this distinction since both sub-systems together constitute the interactive system of an application.

tions that are also used in *graphic animation* and *algorithmic techniques* of composition (see for ex. Wishart 1996; Supper 2001; Nierhaus 2009).[9]

Both domains converge in *computer music* (Roads 1996; Collins 2010). Some of the generative models used in audio rendering – especially statistical and physical models – may be as well applied to recognition and anticipation of the users' movements and gestures. This underlines how the different parts of the interactive system overlap and may even converge into a single algorithm (see for ex. Bevilacqua et al. 2009; Françoise et al. 2013a). This convergence may also further concern the modeling of musical structures applied to digitized sound (see for ex. Assayag et al. 2006b).

Figure *1.1* associates the three technical elements of design as they have been detailed above, with central design aspects (i.e. recorded action, user action, mediation) and, moreover, insists on the relationships between these elements.

- animating recorded sound
- creating relationships between actual and recorded action
- designing playing techniques and scenarios

The *animation* of recorded sound, as investigated in this dissertation, refers to the idea of creating interactive sound processes out of recorded sound materials. Although from this perspective sound and interaction design is very similar to particular techniques of graphic animation, sound- and music-based applications instead refer to terms like *procedural audio*,[10] *generative music*, and *algorithmic composition*. The relationship and obvious similarities of computer-based composition and sound design with visual animation allows for connecting sound and interaction design to the investigation of animation in *culture studies*, especially *animation studies*, and *philosophy* (see for ex. Cholodenko 1993, 2007).

A specific object of study in musicology and music technology is the relationship between movements and sound within music-related practices. Particular studies have investigated the gestural control of sound in the performance of acoustic instruments (see for ex. Rasamimanana 2008) and the possibilities in design of electronic instruments (see for ex. Miranda and Wanderley 2006). But also the relationship between sound and the movements of listeners and dancers has been the object of individual studies (see for ex. Haga 2008; Naveda 2011; Godøy 2009b). Some experimental studies, have examined situations in which the usual relationships of controlling sound by movement and moving according to sound are inverted. This inversion occurs, for example, in the sonification of the movements of dancers and listeners (Naveda and Leman

[9] Complementary to signal and acoustic models, generative models include mechanical models (e.g. *mass-spring-damper*, *ballistics*, *magnetism*), statistical models (e.g. *probability distributions*, *Markov models*, *bayesian networks*), and further models such as *cellular automata* or *genetic algorithms* as well as *grammar* or *rule systems*.

[10] Beyond using physical models of vibratory phenomena for the generation of sound (see for ex. Farnell 2007b), *procedural audio* also includes the modeling of sound textures and articulations of sound on larger temporal scales (see for ex. Farnell 2007a).

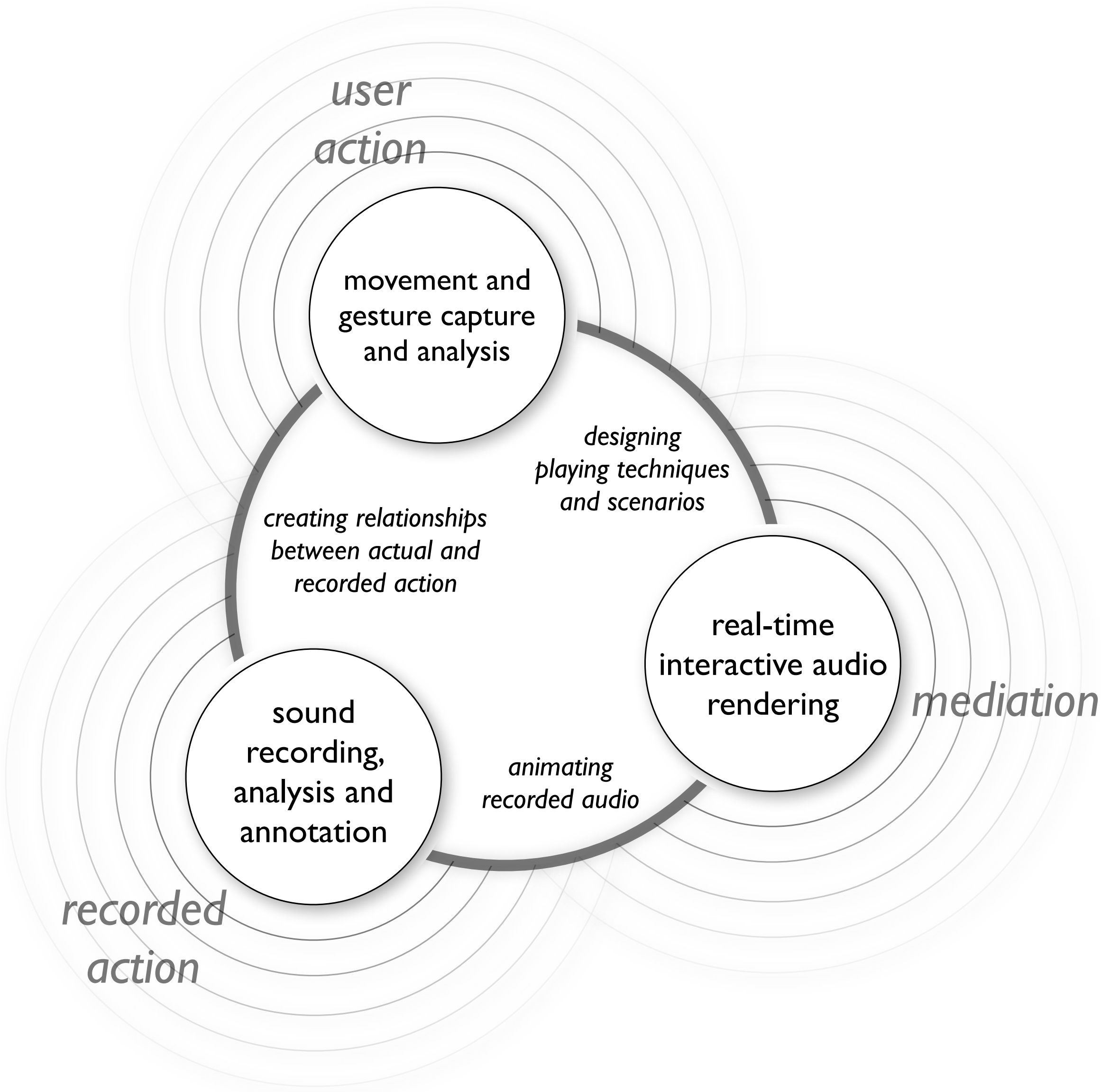


Figure 1.1: Schematic overview of the principal elements occurring in the design of interactive audio applications based on recorded sounds.

2008; Jensenius 2012) and in the performance of *air-instruments* (Godøy et al. 2005). These studies are closely related to studies in psychoacoustics on the perception of action and actors through sound (see for ex. Ballas 1993; Bregman 1994). The notion of *playing techniques* (see for ex. Deegan 1941; Kvifte and Jensenius 2006; Schellberg 1994) appears in the context of sound and interaction design, when considering the relationship between movement and sound in term of a musician's interaction with a musical instrument. Similar aspects concerning the immediate relationship to artifacts have been theorized in terms of *affordance* (Gaver 1991; Norman 1999) and *metaphor* (see for ex. Ludden et al. 2012) that connect interaction design to studies in *psychology* (Gibson 1977) as well as *philosophy*, *linguistics*, and *semiotics* (Johnson 1981; Lakoff and Johnson 1980; Cienki and Müller 2008; Eco 1986; Peirce and Moore 1998).

Further concepts theorized in *design studies* appear when widening the focus on interaction. Different authors have investigated *scenario-based* (Carroll 1995) and *experience-centered* (Wright and McCarthy 2010) approaches to design, as well as *interaction aesthetics* (Löwgren 2009) and *ludic design* (Gaver 2009). Many recent investigations of design rely on concepts of *perception*, *cognition*, *experience*, and *knowledge* based on in *phenomenology* (see for ex. Husserl 1913; Merleau-Ponty 1962; Heidegger and Stambaugh 1996). Central arguments of this thesis refer to the concepts of *enaction* and *embodied cognition* (Varela et al. 1992) that, based on phenomenology, have been developed in interaction with various domains, such as psychology, biology, and neuroscience. Recently, the ideas of enaction and embodied cognition also have considerably influenced fields that directly concern the design of interactive audio applications such as *interaction design* (Dourish 2004), *musicology* (see for ex. Leman 2008), *digital musical instrument design* (see for ex. Armstrong 2006; Essl and O'Modhrain 2006; Wessel 2006), and *music pedagogy* (see for ex. Juntunen 2004). Magnusson (2009) gives a comprehensive overview of philosophical currents – essentially focusing on phenomenology –, as well as of epistemological, semiotic, and hermeneutic aspects of the design of digital musical instruments and other interactive audio applications.

The applications particularly concerned by this dissertation generally invite their users to engage very actively into interaction. Even if cinema limits the users' – or *spectators'* – interactions essentially to watching and listening, the sound design of interactive audio applications has strong similarities with sound design for cinema. Foley art (Ament 2012) is particularly interesting in this context since it shares with the design of interactive applications the primary concern of creating relationships between preexisting actions and sound.[11] In addition, movie sound design is certainly an important reference for sound design in many domains for what concerns aesthetic, conceptual, and technical aspects. In the past years, the practices of sound design for cinema and the sound and interaction design of computer games have developed in a close relationship.

As mentioned above, from a certain point of view, interactive audio applications can be seen as extensions of audio players in the legacy of gramophones, record players, and tape machines (see for ex. Gould 1966; Kittler 1986). Their design shares fundamental concepts and techniques with the production and diffusion of audio recordings (see for ex. Dickreiter et al. 2008; Weinzierl 2008) and their interactions often explicitly or implicitly refer to the interactions with their ancestors. *Interactivity*, rather than a quality distinguishing so-called *interactive* audio applications from others, here becomes a perspective that equally applies to

[11] The users' movements and gestures anticipated in the process of designing interactive applications, in a certain sense, *precede* the articulations of sound that they will induce in the final application.

the interaction with classical audio reproduction technologies. However, the practices of using record players in music performance that emerged in hip-hop culture (Hansen 2010) merit particularly attention as they have reversed the relationship between music performance and recording.

Another well established practice of sound and interaction design that should be mentioned in this context, is the design of sample-based instruments and sample libraries used for the simulation of acoustic instrument and sound effects in music performance and composition (see also Warner 2003). When regarding sample-based instruments (i.e. *samplers*) in terms of the mediation between the actions a musician applies to a keyboard[12] and recorded sound materials, the three elements of design shown in figure *1.1* clearly appear. Relying on preexisting digital music controllers (i.e. *MIDI controllers*), the art of creating sample-based instruments primarily consists in recording (i.e. *sampling*) musicians playing single notes and of relating the captured actions of the instrument performer (i.e. *MIDI events* to the transformed rendering of the recorded notes (i.e. *samples*).[13]

The sound and interaction design of interactive audio applications is an interdisciplinary craftsmanship that is not only informed by a wide range of different domains and practices, but also requires a wide range of competences (see also Magnusson 2009). The design of interactive audio applications often creates the opportunity for collaborating in teams that unite the core competences of sound and interaction design with that of music composition and performance, music pedagogy, musicology, visual arts, game design, and industrial design.

1.3 Methodology and Structure

As already mentioned above, this dissertation investigates key concepts that inform the design of interactive audio applications based on recorded sounds. The perspective privileged in these investigations, considers recorded sound and *listening* as being in the center of the process of design. This does not mean that in practice the recorded sound materials have to precede the other elements of the design. However, in most arguments of this thesis, other elements of design are regarded in reference to recorded sound and whatever they represent and evoke in the listeners' attention.

Figure *1.2* shows the essential aspects of design as they are considered in this dissertation. Sound and interaction design, here is depicted as a process in which *recorded sound materials* are used to create *interactive sound processes* that support particular *user interactions* with the

[12]Even if also other musical controllers may be used to interact with samplers and sample libraries, they are generally designed for keyboard-based interactions.

[13]Recent sample-based software instruments provide scripting languages for adapting the details of the rendering as a function of the control to the expected behavior of a particular simulated instrument – especially related to different playing techniques.

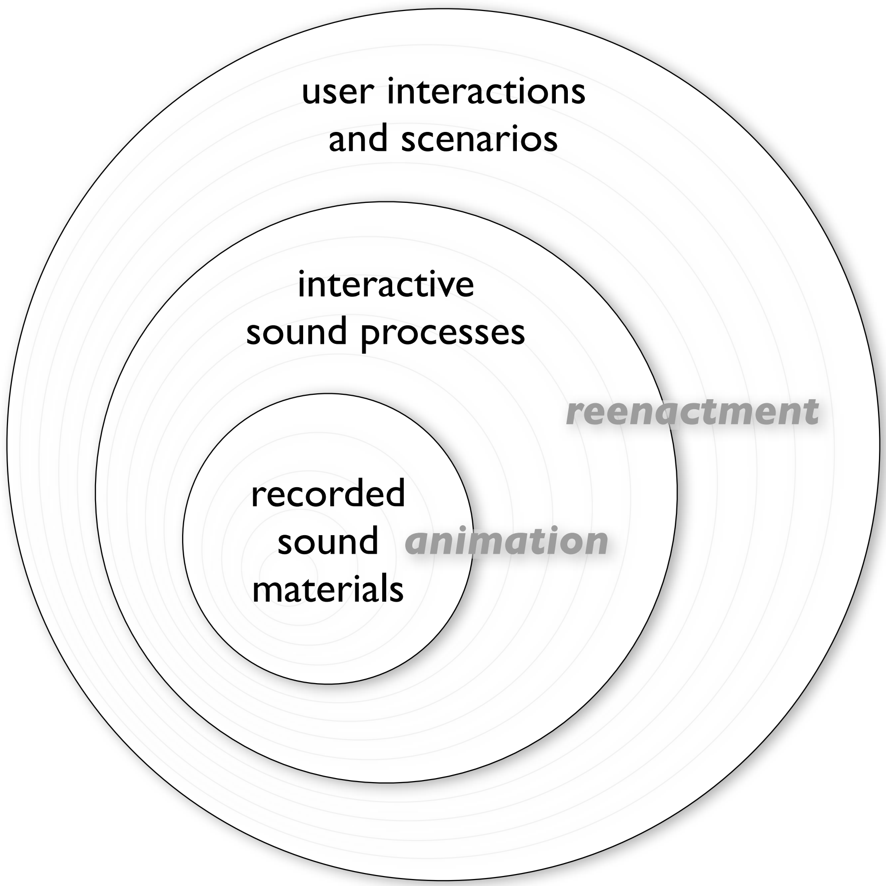


Figure 1.2: Schematic overview of the principal concerns in the design of interactive audio applications based on recorded sound reformulated as elements that are around – and in reference to – recorded sound materials.

application as well as *scenarios* within which these interactions evolve.[14] The principal objects of this dissertation appear from this perspective, *animation* and *reenactment*. Designing interactive sound processes based on recorded sound materials, here becomes an act of *animation*, and the users' actions and their scenarios are seen as a *reenactment* of the recorded sound materials.

The two concepts have been chosen as the principal focal points of investigation because they allow for addressing fundamental aspects of design. They are related to profound *motivations* and *meanings* concerning the process of design as well as the users' interactions with the designed applications. Motivations and meanings that, insofar as they are communicated through design, also constitute the relationship between designers and users. An important ambition of this work is that its argumentation, even where it investigates rather abstract concepts, stays very close to actual concerns and challenges that occur in the practice of design. As the

[14]Above as well as further below, these interactions and scenarios are referred to as *playing techniques* and *playing scenarios*.

choice of the two key concepts, most of the principal arguments of this dissertation originally depart from the introspection of my own design practice. The conceptualization of this practice finally allows for defining it and for relating it to other practices.

In spite of its strong inscription into practice, this dissertation stays within the framework of an interdisciplinary study based on the examination of literature from different fields. In addition to music technology, musicology, and design to which it principally belongs, the fields contributing to this work include philosophy, culture studies, psychology, neurosciences, and linguistics, as well as further fields of science and technology. As it is often the case for studies in music technology, musicology, and design, the actual object of this study consist almost exclusively in nonverbal interactions. These nonverbal interactions include not only the users' interactions with a particular application, but also the relationship between users and designers.[15] One of the principal claims elaborated in this work is that the *nonverbal* interactions through which the users – *listeners* and/or *players* – engage with interactive digital applications actually constitute a form of *reflection*. Digital technologies that allow for constructing these interactions here appear as a means of constructing reflections, which instead of being based on words, rely on nonverbal multimodal interactions. While the principal concern of this dissertation is ultimately nonverbal, its principal object is *words* – such as *animation* and *reenactment* – and its methodology essentially consists in assembling arguments from studies that themselves strongly care about words.

The argumentation of this dissertation would hardly be possible without the foundations that phenomenology has contributed to the investigation of the world as it appears. The essence of these foundations consists in paying particular attention to phenomena (e.g. things, actions, practices, abstractions) as they are experienced by a particular observer from a particular point of view. This idea of considering objects of study actually does not exclude the possibility that phenomena are evoked by recalling former experiences or by imagining fictional experiences. However, this idea does not allow for establishing abstract categories without clearly referring to the qualities that justify these categories. Similarly, it inhibits the observer to isolate a particular aspect of a complex phenomenon without being clear about the intentions and possibilities of the resulting perspective in respect to the phenomenon as a whole. Rather than as guidelines, these teachings from phenomenology have been adopted as an attitude towards examination and reflection that applies to all parts of this dissertation including the exploration of the most abstract concepts as well as the examination of design practice and user interactions. To apply this approach also in reasoning that is not based on actual observations or the introspection of my own practice, the argumentation includes examples of applications designed by others or constructed illustrative examples summarizing further experiences.

[15]Even if interactive audio applications are often accompanied by complementary textual publications and explanations (i.e. announcements, program notes, posters), these publications are rarely considered as necessary for understanding the interaction with the designed application.

Directly related to this approach towards investigation, this dissertation consistently privileges considering interactions and relationships rather than objects and representations whenever possible in its argumentation. This preference is also noticeable in the choice of the term *reenactment* instead of *reembodiment* which had been considered at first place. While *reembodiment* can easily be misunderstood as referring to objectified representations (i.e. by material bodies or the human body), *reenactment* refers much more clearly to action and furthermore can easily refer to concretely *situated* actions such as the users' interactions within the interaction scenario of a particular application. This is especially important in the exploration of digital technologies, as well as the interactions with these technologies, in which the question of what the involved beings and things *are* or *might be* is much less the center of attention than what actually *emerges* from their relationships and interactions. As another consequence of its phenomenology-based approach, this dissertation generally avoids and explicitly argues against established distinctions, oppositions, and categories that are unnecessary for the observation and understanding of particular circumstances or even obstruct it. From this perspective, this dissertation will, for example, critically discuss the opposition of *action* and *sound*, *material* and *cultural* interactions, *inscription* and *embodiment* as well as *experience* and different categories of *knowledge*. In many of these distinctions there still resonates the opposition of *body* and *mind* – or *soul* – that has determined Western thinking over millennia before finally being abandoned through the inventions of phenomenology.

Finally, it is the openness towards phenomena and their experience that sound and interaction design has in common with the methodology that this dissertation applies in its investigation. Beyond the distinction of the verbal and nonverbal, this openness ultimately concerns not only the practice of design and the users' engagement with the designed applications, but also the outside observer reasoning on their relationships and interactions.

The dissertation is essentially divided into two parts. Chapter 2 is entirely dedicated to the in depth investigation of the two key concepts, *animation* and *reenactment*. The principal concern of the investigation of *animation* in 2.1, as well as of *reenactment* in 2.2, is to explore these concepts and to enrich them with manifold perspectives. The argumentation of chapter 2 is further structured through *claims* that regularly highlight certain of its key points. These claims are prefixed and numbered from *A1* to *A12* in 2.1, the first part of the chapter on animation, and from *R1* to *R37* in 2.2, the second part on reenactment. Chapter 3 finally applies and refines these concepts in the description of nine recent applications developed over the past few years. The division of the chapter into two parts allows for focusing on two important aspects, *metaphor* and *play*, even if they ultimately concern all of the presented applications. Each of the four parts constituting the chapters 2 and 3 (i.e. 2.1, 2.2, 3.1, and 3.2) is followed by a brief conclusion. Final conclusions which address the global argumentation and consider further perspectives of the contributions of this work are presented in Chapter 4.

In the making of speech and language the spirit is continually "sparking" between matter and mind, as it were, playing with this wondrous nominative faculty. Behind every abstract expression there lie[s] the boldest of metaphors, and every metaphor is a play upon words. Thus in giving expression to life man creates a second, poetic world alongside the world of nature.

Johan Huizinga in *Homo Ludens*, 1938

CHAPTER 2

Key Concepts

Introduction

Two principal concerns constitute the design of interactive audio applications and musical instruments of virtually any kind. One deals with creating real-time interactive sound processes and the other with the way these sound processes are influenced by the bodily movements and gestures of their players. In this sense, the essence of designing interactive audio applications lies in the creation of meaningful relationships between movement – or gestures – and sound. A particularity of design based on recorded sound consists in the fact that these relationships precede the design in the form of actions that are encoded in the recorded sound materials. In order to better understand the involved design processes, techniques, and interactions, this chapter proposes two key concepts that are investigated in depth. These concepts already appear in the title of this work as *animation* and *reenactment*.

Borrowing the notion of *animation* from cinema, the first part of this chapter considers the designers of interactive audio applications as *animators*. The concept of *animation* not only allows one to confront literature from different fields of culture studies focusing on audio recording, cinematographic animation, and technology, but also to connect sound and interaction design to philosophical concepts such as *soul*, *mind*, and *agency*, which determine many of its aspects. This exploration of profound factors, inspirations and motivations of design through references to culture studies and philosophy is interleaved with historical perspectives on technological developments that are important legacies for the design of interactive audio applications.

The second part of this chapter is dedicated to the concept of *reenactment*. While the notion of *animation* primarily applies to the act of designing, *reenactment* concerns the relationship between the users' interactions with an audio application and the recorded sounds on which a particular application is based. The concept of *reenactment* is grounded in the concepts of enaction and embodied cognition and allows for a unified perspective on the users' engagement into the interaction *with* and *through* sound. The exploration of the concept starts from the investigation of listening as a cognitive process based in the (re-)enactment of action perceived through sound.

While similar investigations have often primarily considered visual perception, the originality of this part of the argumentation lies in its focus on sound. It is the basis for the understanding of interaction through various concepts such as *reproduction* and *congruency*, *knowledge* and *experience*, *metaphor* and *affordance*, as well as *play*. The different perspectives that support and enrich the idea of reenactment include epistemology, semiology, and hermeneutics. Finally, the notion of *playing technique*, borrowed from music, is proposed as a basic concept in interaction design.

2.1 Animation

Introduction

The notion of sound is closely related to the notion of *animation* as the state of '*being full of life.*'[1] In our – human[2] – perception, sound always appears as a witness of movement and interaction. This can refer to the acoustic phenomenon of vibration (e.g. of our vocal cords) as well as to the exchange of arguments in a conversation, the electric drill penetrating the wall nearby while these lines are being written, and the rain falling on a roof in the courtyard.

The fundamental relationship between sound, movement, and interaction is described by Aristotle, in his treatise *On the Soul* (Aristotle and Smith 2006; Polansky 2007), investigating the nature of life and all things living.

> *Sound according to actuality*[3] *always arises of something with respect to something and in something; for that productive of it is a blow*[4]*. Hence indeed, there being but one thing, it is impossible for sound to arise; for the thing striking and that being struck are other; so that which sounds sounds with relation to something; but a blow does not arise without locomotion.*[5]

Animation is evoked by movement and sound very concretely as the side effect of the state of being animated (e.g. of an *animated conversation* or *an animated object*), but also as a constitutional concept. Moreover, its etymology relates it to the latin *animare*, *to give breath to*, *breathing* that represents both, an archetypical movement as well as an archetypical sound.[6]

[1]From the definition of *animation* provided by the Oxford Dictionaries online.

[2]In this chapter, first person (i.e. *we* and *our*) is consistently employed to refer to *us humans*.

[3]Aristotle distinguishes *actual* sound as what we hear from *potential* sound as a property of an object or being that can produce sound.

[4]*Blow* here refers to the propagation of sound in air, *locomotion* (below) to movement

[5]In *On the Soul*, book II, at the beginning of paragraph 8, in a literal translation from Greek provided by Polansky (2007).

[6]Breathing is among the most quintessential examples for the association of bodily movement and sound. Consequently, breath and breathing are repeatedly used as examples and metaphors to support central arguments of this dissertation.

The verb *to animate* denotes the act of *bringing to life* something that initially was lifeless or the act of creating the illusion of life for something that actually is lifeless. Cinematographic animation is described as '*the art of making inanimate objects appear to move.*'[7] Even if other art forms like puppetry claim the term animation, studies of and comments on animation almost exclusively refer to cinematography. Applied to inanimate objects and matter, the notion of animation perfectly fits many practices related to recorded sound that create novel sounds from recorded sound materials.

Nevertheless, *sound materials* and *sound objects* somehow resist their classification as inanimate and lifeless. The inventor of the idea of the sound object, Pierre Schaeffer, closely related the notion of sound objects to movement by studying in his *Traité des objets musicaux* their fluctuation, evolution, and modulation (Schaeffer 1966). One could argue that a digitized sound representing action and movement as a sequence of frames is as least as animated as a motion picture. The animation scholar and theorist Alan Cholodenko even claims that *the digital* was *itself* a mode of animation (Cholodenko 2007). To support a perspective on animation as an essential notion in the understanding of many practices based on digitized sound, it is necessary to further explore circumstances and qualities that relate recorded sound to lifelessness and death.

2.1.1 Sound as a Matter of Death

Probably, the most prominent association of audio recording with death has been immortalized by Francis Barraud's painting *His Master's Voice* from 1898. The painting features Nipper, the dog that, after Friedrich Kittler, '*started sniffing at the bell-mouth of the phonograph upon hearing its dead master's voice.*'[8] The success of this icon is undoubtedly related to the fact that variations of this image became the logo of several branches of a gramophone and record company. After the founder of the British *Gramophone Company* had bought the painting, the US American *Victor Talking Machine Company* had acquired the rights to use *His Master's Voice* as a trademark a few years later (Petts and Andrews 1983; Gelatt 1955). The speculations about whether the polished surface on which the dog sits besides the gramophone depicted the coffin of his master, adds an anecdotal, morbid gloss to this icon.

The association of recorded sound wit death – along with *melancholia* and *erotic nostalgia* – is particularly strong in Allen S. Weiss' book *Breathless* on the origins of sound recording (Weiss 2002). Weiss refers to the period of the invention and exploitation of early sound reproduction techniques as '*the century that spans the death of Poe in 1849 to the death of Artaud in 1948.*'[9] This morbid note is also present in the title *A Resonant Tomb*, the last chap-

[7] From the article on motion picture animation in Encyclopaedia Britannica.
[8] In (Kittler 1986)
[9] In (Weiss 2002).

ter of Jonathan Sterne's book *The Audible Past* that perspicuously investigates the relationship between sound recording and lifelessness (Sterne 2003). In his book, Sterne argues that the development of sound recording has been driven by the same efforts to overcome death that also drove the development of other practices in the same cultural context. He considers the development of sound recording through its contemporary development with the booming of chemical embalming and canned food in the American middle-class society. Sterne insists in particular on the fact that '*early users were fascinated with phonography's potential to reanimate the voices of the death*' well before the technology matured to point of enabling preservation. However, he depicts later sound recording and chemical embalming as processes that '*both transform the interiority of a thing (body, sound performance) in order that it might continue to perform a social function after the fact.*'[10]

Sterne further cites the pejorative term '*canned music*' that the American composer and conductor John Philip Sousa has been credited with. This allegory condenses the resistance that accompanied the introduction of sound recording and clearly opposes it to the live experience of sound, living voices, and traditional practices of music. Lessig (2008) cites another expression of John Philip Sousa's categorical refusal of sound reproduction in his book *Remix: Making Art and Culture Thrive in the Hybrid Economy*. While referring to the phonograph, and particularly to the *Victor Talking Machine Company*, Sousa evokes in this citation the image of '*infernal machines*'. This expression contains a double reference to lifelessness, *machines* – as opposed to the living – and *hell*. Citing further of Sousa's statements, Lessig sees in this attitude the '*fear ... that culture would become less democratic*'[11] in the sense of a decay of musical amateurism. Sousa's vision of the decadence of musical culture culminates in his expression '*the tide of amateurism cannot but recede*'[12] in which he associates sound recording with another very concrete notion of demise – that of existing musical practices and their communities.[13]

The above paragraphs explore different *cables*, as Sterne (2003) puts it, running between sound recording and death. Some of the references already evoke the reproduction of recorded sound as an act of animation (e.g. animating the voices of the dead). Others simply refer to recorded sound and their techniques of reproduction as lifeless and thus fulfilling a necessary

[10] In (Sterne 2003).

[11] In (Lessig 2008).

[12] Cited in (Lessig 2008).

[13] Even though, the idea of animation emerging from this vision, regarding music making as a vivid and vital practice, is not at all unrelated to this investigation of a particular family of practices turned towards digitized sound, a serious discussion of these aspects would largely exceed the scope of this dissertation. However, the motivation for this work draws from the excitement about novel forms of music making blooming on the resonant tomb of the recording industry following its construction with a delay of quasi exactly a century. The arguments on a necessary reform of copyright and the development of alternatives to established business models for the production and distribution of digital content Lessig develops in his book, are perfectly complementary to the investigation of design and techniques creatively applying to this same digital content.

condition for being animated. Finally, each association of recorded sound with lifelessness leads to a new opportunity of animation applied to recorded sound.

A 1 | Sound reproduction is inherently a question of animation dealing with the inanimate.

The basic act of animating recorded sound consists in reproducing it mechanically by recreating the sound waves represented by the recording. However, the reversal of the process of capturing and encoding by decoding and diffusion fits better with the notion of re-animation rather than animation as defined above. The act of *animating* is more clearly present in the constructs of *playing a recording* and *listening to a recording* used in common language.

Taking a closer look at these expressions, the meaning of *sound recording* sensibly shifts from one expression to another. While, more generally, *sound recording* refers to a technology and practice, the act of *reproducing* a recording more precisely refers to the act of interpreting a representation or encoding of sound. In *playing* a recording, the verb *playing* could arbitrarily apply to an apparatus (i.e. gramophone, record player, tape recorder), a support (i.e. record, CD, sound file), or to the recorded content (i.e. sound or piece of music). In all of these cases, the signification of *playing* is similar to the notion of performance and interpretation in the sense of a musician *playing* an instrument and/or a piece of music.

2.1.2 Playing the Phonograph

The idea of considering and using the phonograph as a musical instrument has accompanied its development from the very beginning. Although the quality of early sound reproduction was rather poor, the expressive and poetic possibilities of the gramophone vividly inspired the imagination of poets, musicians, and music critics in the beginning of the 20th century (Battier 2007).

The controversy about whether the phonograph should be considered as a musical instrument becomes palpable in the account of the conductor, composer, and performer Piero Coppola in his account of a meeting with Maurice Ravel and other musicians friends in 1932. Coppola remembers himself '*declaring to [his] friends that the moment has come to admit the phonographic disc to the rank of a musical instrument definitively breaking with the prejudice of some retarded music lovers who insisted on considering it as a simple entertainment machine*', concluding that '*the fight today is absolutely won.*'[14] Even if in Coppola's statement it is not clear who should be considered as the performer of this new musical instrument – the

[14] In (Coppola 1944), cited in (Donin 2004). Translated from the french original: '*[L]e moment était venu d'admettre le disque phonographique au rang d'un instrument de musique, rompant définitivement avec le préjugé de quelques mélomanes retardataires qui s'obstinaient à le considérer comme une simple machine de divertissement. La lutte aujourd'hui est gagnée absolument.*'

musicians or the listener? –, the gramophone is considered as a serious means of musical performance and expression.

However, it was only after the mutation of the gramophone to a portable device in the 1950s, using vinyl discs and integrating electric amplification, that the *turntable* was adopted as an expressive musical instrument by a significant community of performers. Starting within the hip-hop movement of the 1960s and 1970s, DJs developed skillful playing techniques such as *backspinning*, *scratching*, *cutting*, *mixing*, *blending*, and *punch-phrasing* (White 1996).[15] Using these techniques, the performers create new rhythmic patterns from preexisting sound materials – an act of animation that beyond the recorded sound itself aims at engaging with other musicians as well as with an audience. The interpretation of the record player as as expressive musical instrument has been clearly manifested by the DJ and performer Rob Swift who compels performers not only to consider the turntable '*as just this mechanism that you play records on*', but '*apply [themselves] to it as if it were an instrument*' and '*express [themselves] through the turntable.*'[16]

In a different cultural context, the possibilities of the phonograph for the composition and transformation of recorded sound were explored by Pierre Schaffer and Pierre Henry already in the 1940s. Their early experiments up to around 1950, when the first tape recorder arrived in their *Studio d'Essaye*, explored temporal sound transformations and principally relied on turntables. During that same period they successively incorporated into their explorations further technologies which allowed for the manipulation of timbre as well as for the spatialization of sound on multiple audio channels (Battier 2007). Even if musical performance was not at all a priority of Schaeffer's work, movement and gesture have been a primary concern not only in the systematic study of the *typo-morphology* of sound (Schaeffer 1966), but also in his compositional work. Many of Schaeffer's musical studies and pieces as well as the works of his closest collaborators and followers Pierre Henry, Guy Reibel, and Bernard Parmegiani are obviously concerned with something that perfectly fits the notion of animation as the act of creating movement and liveliness via the manipulation of recorded sound materials. Compositions like Pierre Henry'e *Variations pour une porte et un soupir* or Bernard Parmegiani's *L'instant mobile* can be taken as programmatic works in this sense.

As the director of the research department at the French national public broadcasting organization RTF, Schaeffer certainly had excellent insight into the contemporary development of technologies for cinematographic production and animation. Apart from the *Groupe de Recherche Musicale* (GRM), the RTF research department integrated a group that developed technologies for cinematographic animation such as a machine called the *Animographe*.

[15] Cited in (Hansen 2010).

[16] Quoted in (Reighley 2000). Interestingly, in his expression Swift makes an explicit distinction between the turntables and the musical instrument by including '*as if it were an instrument*'.

Both practices, *musique concrète* and hip-hop turntable performance, have laid out the essential elements for the transformation, composition, and performance of recorded sound, imposing new structures, evolutions, and expressions to existing materials. Based on analogue technologies, the techniques that already existed in the 1960s anticipated most of the methods and paradigms for the processing of recorded sound that are still valid today. This concerns not only temporal, timbral, and spatial sound transformations, but also the organization of sound materials and the control of transformations for complex compositions and performances.

A 2 | Based on turntables and tape machines, *musique concrète* and hip-hop have invented the basic paradigms of the animation of recorded sound in composition and performance.

In the 1960s and 1970s, the composition and performance of recorded sound still strongly involved the physical manipulation of sound media and devices. When Pierre Henry states in 1970 that '*[a]nyone who practices Electroacoustic music invents his music, makes it, makes it real, with his fingers,*'[17] it reads like a description of the work of a sculptor or even a puppeteer. The disembodied voice had found a comfortable new body in the phonographic devices installed in recording studios, concert halls, clubs, and private homes. The organs of this artificial body were animated by the gestures of sound engineers, composers, performers, and music lovers. While in the early gramophone these gestures were sustained by spring mechanisms, in later phonographic technologies they are articulated with the force of electric motors driving turntables tape wheels.

Similar mechanisms sustaining the production of sound by mechanical energy were already present in musical automata preceding the development of the phonograph by several centuries (Dohrn-van Rossum 1996). The articulation of human bodily gestures through mechanically sustained energy and automated temporal processes is a central problematic of this dissertation. It connects the design of interactive audio applications with the history of musical automata and with the history of automation more generally. This history is also intrinsically related to the history of horology (Ord-Hume 1973; Koetsier 1997).

2.1.3 Automation

The musical automata described by the Persian Banu Musa brothers in the 9th century and the ensemble of four automatic instrument players created by the Arab or Kurd engineer Al-Jazari in the 13th century were animated by the same hydraulic mechanisms that had been used for

[17] Cited in (Battier 2007).

building clocks since ancient times in Egypt and China (Dohrn-van Rossum 1996; Haspels 1987). Similarly, the dulcimer player acquired by Marie Antoinette at the end of the 18th century was based on a spiral balance spring built by a German clockmaker.[18] The balance spring that had revolutionized horology over hundred years earlier (Usher 1988), still drove commercialized gramophones at the end of the 19th century (Gelatt 1955). The development of clocks and musical automata share the technical requirement of temporal precision and steadily sustained movement with many other machines.[19]

While musical robots, such as the automata mentioned above, imitate the physical movements of instrument players, other mechanical musical automata such as music boxes and player pianos, only generate the necessary movements to directly drive the actuators and resonators of musical instruments. The phonograph ultimately is limited to the reproduction of sound waves. Yet the turntable and tape wheels still inherited a trace of macroscopic movement from their ancestors which is not directly related to the production of sound but to the progression of time.[20]

The notion of *automation* is semantically very close to *animation*. The word automat etymologically derives from the Greek *automatos* meaning *self-acting*, composed of *autos* for *self* and *matos* for *animated*, *willing*, or *thinking*.[21] While the notion of animation focuses on the idea of creating *liveliness*, automation further insists on *autonomy*. From this perspective, automation appears as a special case of animation by mechanical means. On the other hand, an automated process, in opposition to human movements and gestures, is lifeless by definition. In this sense, automation seems rather to signify the opposite of animation as the act of creating a lifeless process which replaces or simulates actions that originally were lively and embodied. While animation applies to an object by creating a process unfolding in time, automation, in the contrary, abstracts a temporal process into an object.

[18] The *Joueuse de tympanon* constructed by the German clockmaker Peter Kintzing and cabinetmaker David Roentgen is preserved in the Musée des Arts et Métiers in Paris.

[19] In fact, it is only in the period when sound reproduction mainly relied on turntables and tape machines driven by electric motors coinciding with the invention of *musique concrète* and hip-hop, that the mechanisms of driving timepieces and sound playback diverge for a few decades. After this, from a historical point of view very short period, chronometry and sound reproduction converge again in the development of digital technologies relying on crystal oscillators.

[20] The gestures of music concrète composers and hip-hop turntable performers are actually articulated against the resistance of electric motors designed to steadily drive forward the temporal progression of the playback of a sound recording. This association of force and inertia to the progression of time and to the consequent transformation of pitch and timbre is a remarkable artifact of this generation of technology that has been often simulated by digital applications since.

[21] From the Online Etymological Dictionary.

A 3 | As the creation of autonomous temporal processes, automation is a form of animation.

As the objectification of a lively temporal process, automation is the opposite of animation.

The *self*, *will*, and *thought* of the automata described above is basically represented by a sequence of preprogrammed actions. Al-Jazari's automatic musicians and Marie Antoinette's dulcimer player, as well as many other musical automata, were driven by camshafts (Maggiore and Meneghetti 2011). Although they encode discrete events rather than continuous movements, the cylinder of pegs in a music box as well as the music roll of the barrel organ and player piano are almost identical from a mechanical point of view. Apart from the dimensions and generated force, the functioning of all of these mechanisms is very similar to that of Edison's cylindric phonograph and later the gramophone records. In each of these mechanisms, a temporal process is encoded sequentially as the shape of a rotating object and decoded through a steady movement of rotation that translates the shape into physical movements which cause the generation of sound. Since ancient times, the sequences of movement and action generated by camshaft mechanisms are programmable by changing the rotating cams. The rhythms played by Al-Jazari's automata could already be composed this way. A more recent example of composition by programming the sequences of musical automata is the work of Conlon Nancarrow who created his compositions for player piano by directly punching holes into the piano roll of a player piano determining the pitches, durations, and dynamics of notes and chords (Gann 2006).[22]

With the musical robots mentioned above, the gap between the macroscopic actions of the mechanical musicians encoded into the camshaft and the generation of sound is completed by two elements. One is the simulation of the musicians' gestures actuating the instrument and the other is the musical instrument itself. Complementary to the mechanical representations of actions and movements, these elements can be seen as objectified representations formalizing temporal processes.[23] The musicians' gestures are represented by mechanisms simulating the functioning of mechanical limbs, fingers, and even facial expressions. In a similar sense, musical instruments formalize the modes of the production of musical sounds by the materials and mechanical topologies of their components. This includes the geometries of vibrating cords, plates, and air tubes, the mechanisms of exciters, resonators, and dampers, as well as the disposition of finger holes, fingerboards, and keyboards. The material and dimensions of a

[22]The player pianos used by Nancarrow were actually designed to play back piano performances that had been recorded on special recording pianos.

[23]The notion of *formalization* here refers to the stabilization of defined abstractions that are associated with reproducible phenomena.

dulcimer string for example, have been precisely adjusted to determine required qualities such as its frequency, timbre, and sustain.

This brief analysis shows that early musical automata already juxtapose multiple elements of formalization for the production of complex articulations of sound. Those elements include preprogrammed sequences of actions and movements, as well as models of bodily movements and acoustic features. The potential for animation appears here as the result either of constructing, programming, or recording, through the simulation of bodily actions and, more generally, through the formalization of temporal processes. However, in this unified view, and as mechanical formalizations of temporal processes, the vibrating strings and the camshaft mechanisms are fundamentally different in the way that they represent a temporal process and, ultimately, time itself. While the sequence of movement represented by the camshaft is a spatial representation that is potentially reversible by turning the camshaft into the opposite direction, the strings implement an irreversible process of resonance.[24]

A 4 | The animation of sound is based on formalizations of action, movement, and sound as temporal processes.

The automata considered so far are limited to the simulation of actions and movements that are predetermined by sequential programs and recordings. Regardless of their obvious limitations, the reproduction of the pre-composed actions, movements, and sounds evoke intention, behavior, intelligence, and emotions, even if these notions largely exceed their programs.

In an introductory chapter of his book *Cybernetics or Control and Communication in the Animal and the Machine* (Wiener 1961) first published in 1948, Norbert Wiener condenses the history of automation in a few lines which include interesting allusions to animation.

> *At every stage of technique since Daedalus or Hero of Alexandria, the ability of the artificer to produce a working simulacrum of a living organism has always intrigued people. In the days of magic, we have the ... concept of the Golem, that figure of clay into which the Rabbi of Prague breathed life with the blasphemy of the Ineffable Name of God. In the time of Newton, the automation becomes the clockwork music box, with the little effigies pirouetting stiffy on top. In the nineteenth century, the automation is a glorified heat engine, burning some combustible fuel instead of the glycogen of the human muscles.*[25]

[24] Interestingly, also the balance spring driving the rotation of the camshafts into the automata of the 18th century belong to these irreversible representations of time – once they are wound up.

[25] In (Wiener 1961).

The fact that these automata do not actually act autonomously but are obviously driven by inanimate mechanics, ultimately confirms a worldview that separates the materialities and mechanics of bodies from an entity that brings them to life. Depending on the cultural or philosophical context, epoch, and translation, this entity, manifesting itself as *intention*, *intelligence*, and *free will*, is referred to as *soul*, *spirit*, *psyche* or *mind*.

2.1.4 De Anima

Forged by mystic thinking, religion, philosophy, and science over millennia, the idea of the animate, is related to the idea of souls inhabiting nature and the bodies of living beings. Animism, as '*the attribution of a living soul to plants, inanimate objects, and natural phenomena*' or '*the belief in a supernatural power that organizes and animates the material universe,*'[26] reaches far into history and into the prehistory of mankind. The 18th century Scottish philosopher David Hume states in his *Natural History of Religion* (Hume 1757) that '*[t]here is an universal tendency among mankind to conceive all beings like themselves, and to transfer to every object, those qualities, with which they are familiarly acquainted, and of which they are intimately conscious.*'[27]

A generalized idea of animism that persist in the image of a divinity animating the universe, has been perpetuated not only by religion and philosophy scholars, but also by scientists of all centuries. For example, the astronomer Johannes Kepler, before proposing his revolutionary interpretation of the movements of the planets around the sun, starts at the end of the 16th century from the idea of a divine force that he calls '*anima motrix*' (Kepler et al. 1937).[28] About 400 years later, probably not without irony, Albert Einstein makes reference to a divinity as the driving force of the universe in his critique of quantum mechanics that is condensed in the famous question about whether *God* would play dice (Isaacson 2007).

Hume, together with philosophers like Descartes, Berkeley, and Kant, is one of the protagonists of the *philosophy of mind* in the 17th and 18th centuries. This current of philosophy inherits the idea of a soul – or mind – as a separate entity from ancient Greek philosophers like Socrates, Plato, and Aristotle (Morton 2010). Aristotle's treatise *On the Soul – De Anima* in latin – is an essential reference for all currents of Western philosophy treating this subject. Even if Aristotle's observations and reasoning on motion, for example, is extremely precise,[29] from today's perspective many parts of *On the Soul* read like a myth featuring metaphorical events and characters. The idea of a soul – *psyche* in Greek – in the time of Aristotle has its ori-

[26]From the definition of *animism* provided by the Oxford Dictionaries online.

[27]In (Hume 1757).

[28]While already Aristotle mentions the relation between the distance of the planets and the duration of their orbit in his work *On the Heavens*, Kepler finally introduces a purely mechanical and quantified explanation.

[29]A complete summary of Aristotle's idea of motion is given in (Sachs 2005).

gin in the *breath-soul* escaping death in Homer's epic poems about 500 years before Aristotle (Polansky 2007).

Aristotle's investigation in *On the Soul* starts from the discussion of the work of his predecessors who, according to Aristotle, distinguished the *ensouled* from the *not-ensouled* primarily by the faculty of movement and sensation. Aristotle himself associates the presence of soul to five principal faculties that can be approximately translated as *movement, nutrition*, *perception*, *imagination*, and *intellect*.[30] He conscientiously explores the faculties of the soul and their mutual relationships before employing them to consider the ensoulment of plants, animals, and human beings. In one passage of the treatise, Aristotle describes the soul as the origin of *voice*, a characteristic capacity of human beings that he also attributes to certain animals. Aristotle describes the production of voice as involving the agency of the soul impacting on the organ of respiration he calls *windpipe*.

> *Not every sound . . . made by an animal is voice . . . ; what produces the impact must have soul in it and must be accompanied by an act of imagination, for voice is a sound with a meaning, and is not merely the result of any impact of the breath as in coughing; in voice the breath in the windpipe is used as an instrument to knock with against the walls of the windpipe.*[31]

Here the soul expresses itself through the voice, that can be perceived and recognized by other souls. Whereas for his predecessors, movement was a sufficient indicator of ensoulment, Aristotle insists on *imagination*.[32]

The investigation of the *animated* may appear as a distant echo of Aristotle's examination of what constitutes the *ensouled*. By the time Réné Descartes transforms the concept of *soul* into the concept of *mind* – renewing its clear distinction from the body –, movement and especially the movement of the *heavens* have been sufficiently explained. And as if Aristotle's list of faculties of the soul had been taken as a program of a vast of scientific project, even nutrition was no longer listed among the faculties of the mind, but had become a function of the body. The automation of a defecating duck presented by Jacques Vaucanson in 1738, alongside an automatic breathing flute player, was the public demonstration that, at least theoretically, [33] the mystery of the transformation of the body by nutrition had been unveiled (Riskin 2003).

[30]The English terms actually used to describe these faculties differ in different translations. This list combines terms that can be found in (Aristotle and Smith 2006) and (Polansky 2007).

[31]In *On the Soul*, book II, at the end of paragraph 8, in the translation provided by (Aristotle and Smith 2006).

[32]Instead of the term *imagination* taken from (Aristotle and Smith 2006), in (Polansky 2007) systematically the term *phantasia* is used. Aristotle discusses in detail this faculty of soul and its relationship to others, especially perception and intellect, in book III of *On the Soul*.

[33]The digestion Vaucanson claimed as a chemical elaboration in the duck's stomach was demonstrated to be fraudulent. Actually, the grains devoured by the duck were falling into a reservoir and the duck's rear was loaded before the demonstration with fake excrement.

As the above citation of Wiener summaries the developments from the point of view of the 20th century, the invention of the steam engine around the beginning of the 17th century had definitely demonstrated man's mastering of the relationship between nourishment and the faculty of movement as they had been investigated by Aristotle. These developments mean that at the beginning of the 20th century, only *perception*,[34] *imagination*, and *intellect* remain of the Aristotelian faculties of the soul.

> A 5 | The notion of animation reveals the development of technology as a questioning of the qualities and boundaries of the living.

The technical conquest of the last faculties of the soul are the domain of what Wiener lays out as the '*new science of cybernetics*' (Wiener 1961) – together with updated versions of Newton's models and complex natural phenomena such as meteorology. By cybernetics, Wiener predicts the emergence of multidisciplinary approaches bringing together the knowledge of different fields including engineering, mathematics, physics, biology, psychology, linguistics, and sociology. Relying on mathematical formalization and computational simulation, these disciplines undertake the investigation of complex dynamic systems including natural phenomena, animal behavior, and the human nervous system (Wiener 1961).

Before cybernetics, automation was basically limited to the physical reincarnation of movement patterns, but Wiener's new science of automation does not stop before any intelligible phenomena including human intelligence itself. According to Wiener '*the newer study of automata*' has completely established itself from power engineering as '*a branch of communication engineering, and its cardinal notions are those of message, amount of disturbance or "noise" . . . , quantity of information, coding technique, and so on.*'[35] When Wiener wrote *Cybernetics*, digital computers were just starting to outperform analog calculators. In parallel, the description of continuous signals, symbolic information, and automata of any complexity had been unified into digital representations that can be transformed, one into the other, by relatively simple numeric operations. These representations further unify the means for the formalization and simulation of temporal processes that in early automata had been realized mechanically.

2.1.5 Sound as Digital Matter

The transitional electro-mechanic bodies hosting the disembodied voice in the electronic studio of the 1960s, including turntables, tape machines, and mixing desks, manifested their own bodily utterings. The imprecision and aging of their mechanical, chemical, and electric com-

[34] Even if the mechanisms of perception had been the object of many investigations such as Descartes and Newton's optics, and Helmholtz' acoustics, these works explain only the mechanical parts of perception.

[35] In (Wiener 1961).

ponents interfered with the increasing expectations regarding a perfect and sustained fidelity of the reproduced sounds to their recorded originals.[36]

The transitional reincarnation of recorded sound by electro-mechanic machines has been practically eradicated by the development of digital audio technologies. Although, the first experiments with digital sound synthesis and digital musical representations are conducted in the late 1950s, the quality of digital technologies started to be comparable to existing analog sound reproduction technologies only towards the end of the 20th century (Collins 2010). While early digital audio technology still produced undesirable artifacts especially due to the limited precision of analog-digital conversion, today the complete disembodiment of digitized sound can be considered as achieved. As with other media, digitalization has largely liberated sound from any residual body.[37]

A 6 | Digital sound is a disembodied inanimate matter calling for its re-embodiment and re-animation.

Nowadays, virtually all representations of sound and articulations of sound have been embraced by the digital revolution. Many further formalizations and models of temporal processes have been developed since.[38] Representations of audio signals, coexist on today's computers with symbolic music representations, audio descriptors, motion capture data, and images, as well as with a multitude of algorithms for the generation, transformation, and articulation of digital sound. Each of these representations produces a different point of view on sonic phenomena and multiple representations can produce complementary points of view on the same phenomenon. Being constructed from the same digital substance, one representation can be easily transformed into another.

However, without being animated, digital representations of temporal processes are a completely bodiless and lifeless matter. Again, the primary mode of animating digital content resides in playing it back. A sophisticated version of this mode has become known as *digital sampling*. As for sound recording in general, also for digital sampling[39] the connotation of ani-

[36]In addition, especially when considering the immateriality of sound, these devices can retrospectively appear massive and clumsy.

[37]Evidently, still today the reproduction of digitized sound depends on electroacoustic apparatuses at least for its projection into physical spaces. Nevertheless, the sequence of numbers constituting a digitally represented sound has a completely immaterial identity that radically can change their incarnation without suffering any loss.

[38]Especially in the digital domain and applied to temporal processes, the notions *representation*, *model*, *algorithm*, *program*, and *formalization* finally refer to very similar objects. But even if these objects have the same function of digitally representing a temporal process, the terms may refer to different actions like *model building*, *place holding*, and *formalizing*.

[39]The terms *sampling* and *sample* can be misleading, since their original technical meaning differs from their meaning in music. Below, they systematically occur in their "musical" sense, *sample* signifying a digitally represent segment of recorded sound.

mation can vary significantly between different practices and approaches. As the compositional practice of including fragments of existing music recordings into live performances and studio productions, *sampling* has its origin in hip-hop culture that precedes the digital revolution by at least two decencies. Digital sampling, as a widely used technique in music production since the 1980s, inherited many aspects from *musique concrète* (Warner 2003). The musicologist Joanna DeMers has created the term *transformative appropriation* to describe the act of referring to older works in the context of newer works by sampling in hip-hop and later pop music (Demers and Coombe 2006).

In the age of digital sound databases and online music publishing services, the total disembodiment of digital sound turns into the promise of perpetual reincarnation of digital sounds through their permanent exchange and transformation. Sterne (2006) advocates in favor of digital recordings against the ongoing criticism regarding their reputed lack of liveliness and embodiment. He argues '*that their liveliness should be judged by the degree to and manner in which the recordings themselves circulate*' concluding that '*[j]udged by their social lives, rather than by a dubious metaphysics, digital recordings are at least as lively as analogue recordings ever were.*'[40]

A 7 | The disembodiment of digital sound facilitates its animation.

A more technical notion of animation is involved in applications of sample-based techniques that seek for the simulation of acoustic musical instruments or environmental sounds in the production of music and multimedia applications such as games. As literal reproductions, digital samples can easily convince with rich and lively details of the captured sound, while many details of synthesized sounds reveal their synthetic nature. However, it is sufficient to trigger a given sound sample multiple times to prove the lifelessness of a sampled sound object. The art of letting the repetitions, modulations, and transitions of sampled sound snippets appear lively, perfectly fits the idea of animation.

Physical modeling, as a successful alternative to sample-based techniques for sound simulation, actually simulate the dynamics of vibrating masses and propagating sound waves. From the perspective of this discussion, an essential difference between sample-based techniques and physical modeling lies in the different notions of animation they imply. Independently from the employed synthesis technique, the obvious object of animation in sound simulation is the simulated sounding object or process. This could be, for example, a saxophone sound played on a digital keyboard or the breathing associated with an avatar in a video game. But apart from these ultimate objects of animation, each technique is constituted of representations (i.e. models, algorithms, programs) that produce actual sound only through their animation.

[40]In (Sterne 2006).

In this sense, sample-based techniques are concerned with the animation of digitized sound materials and sound synthesis by physical modeling consists primarily in an animation of the underlying mathematical model. Finally, any model of sound and sound articulations, first of all enables the animation of itself, whether it focuses on the reproduction of captured sound waves, physical dynamics, statistical distributions, or geometrical structures.

A 8 | The representation of temporal processes underlying a particular technique of animating sound is a primary object of animation.

The examples of saxophone performance and breathing may illustrate how different notions of animation appear in the simulation of sounds that can be associated with bodily interactions. A basic simulation of a saxophone could be limited to the rendering of the different timbres produced by the saxophone. This would leave all aspects of performance, such as modulations and phrasing, to the performer (see for ex. Vergez and Tisserand 2006). A keyboard performer could use the modulation and pitch bend wheels – possibly together with a breath controller – to animate the virtual saxophone like a puppeteer. More sophisticated saxophone simulations could include models of expressive modulations and coarticulations that automatically complete the limited note information provided by the keyboard controller (see for ex. Lindemann 2007). For the automatic performance of digital music representations (i.e. MIDI scores), a convincing simulation would have to generate expressive tempo changes, phrasing, and ornamentation (see for ex. Friberg et al. 2006). Automatic improvisation systems, ultimately attempt to simulate the musical intelligence of a human performer (see for ex. Pachet et al. 2013). A similar scenario can be created for the hypothetical example of simulating the sounds of breathing in and out associated with a game avatar. The frequency, regularity, duration, and timbre of the avatar's breathing would represent aspects of its corporal and emotional state. Analog to the example of the saxophone, a first version of this simulation could provide a control of each individual breathing cycle, before abstracting more and more the interaction with the avatar by integrating complex behavioral patterns.

Again, in these examples, animation and automation go hand in hand as complementary aspects of simulation. While the notion of *animation* highlights *liveliness*, *automation* insists on *autonomy*. Each step of automation abstracts a lively temporal process into a model, finally enclosing the entire process of saxophone improvisation into multiple strata of digital representations. At the same time, each level of automation creates a new opportunity of interaction and expression. In the virtual saxophone example, the progression of automation transforms, step by step, the instrumental interaction with the saxophone into the musical interaction with an automatic saxophone player. The interaction with the avatar, starting from a breathing marionette, ends up as a character that autonomously expresses its fatigue and anger through breathing.

As in these examples, the interdependence of aspects of automation and interaction can be observed in the evolution of traditional musical instruments based on mechanical formalizations of bodily action. In this sense, turning the handle of a barrel organ is an automation of organ playing, that itself is an automation of flute playing. Today, computer controlled pianos act as autonomous partners in musical accompaniment and collective improvisation using technologies such as scorefollowing (Vercoe 1984), models of musical style (Dubnov and Assayag 2002), and improvising systems (Lewis 2000). In the design of novel digital instruments, the instrumental interaction with simulations of instrumental mechanics and acoustics can coexist with digital realizations of different aspects of musical performance, composition, and listening. The infinite possibilities of combining arbitrary models of automation and interaction in the digital domain have created an inexhaustible field of experimentation in music, as well as in many other domains.

The animation of digital sound has two interdependent modes, automation and interaction. On one hand automation appears as replacing interaction, on the other hand automated processes create new possibilities of interaction. From these two modes, emerges a possible definition of animation as a factor in the design of interactive audio application. In this definition, animation appears as the articulation between elements of automation and elements of interaction.

A 9 | The animation of sound consists in an articulation of automation and interaction.

By improvising music in interaction with other musicians, machines have shown their capacities of *perception*, *imagination*, and *intellect*. Within the past decades, engineers have conquered the last three entries on Aristotle's list of faculties which distinguish the animate from the inanimate. Already before, the conceptual separation of mind and soul, inherited from classical Greek philosophy, has created major contradictions that over several centuries have been discussed by philosophers and theologists as the *mind-body problem* (Morton 2010). Already vividly criticized in the 18th century by philosophers like Berkeley and Kant, this separation has been ultimately omitted in the beginning of the 20th century in the stream of philosophy initiated by Edmund Husserl as *phenomenology*.

2.1.6 Actors and Interactions

In his doctorate thesis form 1896, *Essai sur les données immédiates de la conscience*,[41] Bergson develops a revolutionary conception of time in the notion of *pure duration*[42] (Bergson 1908). By *pure duration* Bergson means a '*succession of qualitative changes, which melt into and permeate one another, without precise outlines.*'[43] He strictly distinguishes *pure duration* from *spatialized* notions of time that are omnipresent in the way we operate with time in our imagination, as well as through externalized spatial representations such as clocks and calendars. Since *representation* and *imagination* (i.e. creating an image) fatally introduce a notion of spatiality, it is difficulty to represent and imagine *pure duration*. However, Bergson describes it as an *ongoing progress* that is neither defined by what happened in the past nor by the anticipation of the future.

> *Can time be adequately represented by space? To which we answer: Yes, if you are dealing with time flown; No, if you speak of time flowing. Now, the free act takes place in time which is flowing and not in time which has already flown. Freedom is therefore a fact, and among the facts which we observe there is none clearer.*[44]

The concept of *free agency* that Bergson defines as a logic consequence of *pure duration* appears in the history of philosophy as the ultimate *state of aggregation* of what Aristotle had called *psyche* and what had become *sprit* and *mind* later on. While the presence of a *soul* distinguishes the living from the dead or mechanic, *free agency* introduces the idea of the *free* and *unruled* in opposition to the *predetermined* and *controlled*. As a consequence of Bergson's precise reasoning, *free agency* is not only an abstract or metaphorical concept like *soul*, *spirit* and *mind*, but also a function of the body that directly emerges from the concept of *pure duration*. While Bergson himself maintains the idea of a *spirit* separated from the body, nothing in his reasoning appears as an objection to drop the idea of this distinction.[45] Other

[41] The first english translation of Bergson's thesis appeared 1910 under the title *Time and Free Will: An Essay on the Immediate Data of Consciousness* (Bergson and Pogson 1910).

[42] The translation of the french term *durée*, used by Bergson, into *duration* is justified insofar it also translates the proximity of *durée* to its verb *durer* also translating to *endure*.

[43] In (Bergson and Pogson 1910), chapter II, translated from the french original: '*[L]a pure durée pourrait bien n'être qu'une succession de changements qualitatifs qui se fondent, qui se pénètrent, sans contours précis.*' (Bergson 1908)

[44] In (Bergson and Pogson 1910), conclusion of chapter III, translated from the french original: '*« le temps peut-il se représenter adéquatement par de l'espace ? » - A quoi nous répondons : oui, s'il s'agit du temps écoulé ; non, si vous parlez du temps qui s'écoule. Or l'acte libre se produit dans le temps qui s'écoule, et non pas dans le temps écoulé. La liberté est donc un fait, et, parmi les faits que l'on constate, il n'en est pas de plus clair.*' (Bergson 1908)

[45] Merleau-Ponty and Deprun (1997) have analyzed the contradictions generated by the maintenance of the separation of spirit (*esprit*) and body (*corps*) in Bergson's work.

philosophers like Husserl and Merleau-Ponty have since then proceeded in abandoning this conception.

Norbert Wiener summarizes the consequences of Bergson's ideas from an engineer's point of view by stating that '*there is no reason in Bergson's considerations why the essential mode of functioning of the living organism should not be the same as that of the automation*' – based on cybernetic models. Wiener's expression is carefully composed around a double negation. The question whether the soul belongs to the domain of mechanics or whether mechanics can be ensouled, he concludes with the remark that '*the whole … controversy has been relegated to the limbo of badly posed questions.*'[46]

The idea of considering humans, animals, and machines from a common perspective, allows for introducing a notion of symmetry into their interactions. While before, the interaction between a human and a machine was defined as an ensouled and mindful human controlling a mechanical device, now the same interaction appears as a complex relationship between two interdependent systems. Even if in many applications, the machine stays a tool that is largely inferior to the complexity of the human body, complex digital systems may display complex patterns of behavior from which emerge notions of emotion, intelligence, and creativity.

The player's interaction with a traditional musical instrument is obviously not the same as the dialog between two improvising musicians. However, both cases can be seen as balanced couplings. The musical instrument has been adapted to the player's capacities as well as the player's capacities have adapted to the instrument in the process of learning it. Even if the symmetry of this interaction is less obvious when focusing on certain aspects of physical interaction, their mutual influence can be observed from a larger perspective. In this sense, Suchman (2007) proposes the idea of *mutuality* that better allows for reasoning on the coupling between humans and machines into common processes without denying obvious asymmetries appearing from many perspectives.

In contemporary theories, anyone or anything can be considered as an *actor*. This includes not only humans and machines, but also any other kind of phenomenon. No distinction is made between *material* and *social* interactions or relationships of any other kind (see for ex. Dewey and Bentley 1949). *Actor-network theory* (see for ex. Latour 1996) unifies human and non-human actors into '*a semiotic definition – an actant – that is, something that acts or to which activity is granted by others.*'[47] In *actor-network theory*, an interaction is always situated within a network of further actors and interactions. In addition, any interaction can be decomposed to a network of further interactions. Instead of considering interaction as something happening between preexisting actors, actors are *defined* through their network of interactions with other actors. The basic ideas of *actor-network theory* are grounded in John Dewey's concept of *trans-*

[46]In (Wiener 1961).
[47]In (Latour 1996).

action (Dewey and Bentley 1949). Defending a similar perspective on interaction, Suchman (2007) has advanced the concept of *situated action*. The concept of *intra-action* proposed by Barad (1998) goes as far as questioning the existence of actors prior to their interaction. Barad's idea of *intra-action* is deeply inscribed in quantum mechanics, but the basic concept is to consider not only the mutual involvement of two actors, but a virtually infinite entanglement of interactions that constantly reconfigure – and materialize – around their interaction.[48]

An example of the interaction between a trumpet player and his trumpet may illustrate the idea of entangled networks of interactions and actors. The musician is walking down the street in a Sicilian village with a marching band during the celebration of the local saint. He plays the trumpet that he inherited from his grandfather several years ago. From a certain point of view, this interaction between the trumpet player and his instrument is determined by their participation in the marching band that started the usual course through the village already some hours ago. From here depart – and here arrive – the traces that connect the trumpet player's meeting with the marching band to his childhood and birth in the neighbor village, the history of his family, and ultimately of mankind all together. On the trumpets side, the origins are not only entangled with the trumpeter's family, but also with the genesis of the trumpet as a musical instrument and its musical practice. Even the material of the trumpet comes with its own part of the history that is deeply rooted in the history of earth, human culture, manufacture, and instrument making which involve a great number of recognized, as well as anonymous actors. In addition, the coincidence of the trumpet and the musician are determined by the history of the marching band that itself is related to that of the celebrated saint. Another perspective, focusing on the manual interaction of the trumpeter with his trumpet, reveals the entangled development of the player's hands and the trumpet's mechanics. When further focusing on the interaction of the player's fingers with the valves while playing a particular note, another trace of interaction leads through the genesis of the currently played piece deep into the musical practices of harmony and counterpoint. Finally, the interaction between the trumpet player and his trumpet is entangled with the interactions between the musicians of the marching band and their interactions with their instruments. Moreover, there are the marching band's interactions with the other people of the village standing along the way – and so forth.[49] This example, which could be infinitely extended, illustrates how an interaction intertwines into complex entanglements of material and non-material (e.g. social, cultural or discursive) interactions with further actors.

Obviously, also the design of interactive digital applications is always situated within manifold interactions that *constitute* the application technically as well as culturally. Designing

[48]To some extend, the idea of *structural coupling* developed by Humberto Maturana and Francisco Varela (Maturana and Varela 1980) also strongly relates to these ideas.

[49]The example does not mention any of the actions and causalities related to the interaction of the author who has observed or imagined it before writing it down. Moreover, the reader may add his or her own elements of entanglement related to his or her interaction with the example.

interactions here becomes the act of situating and re-situating interactions within complex entanglements.

A 10 | Interactions are always situated within and constituted by a virtually infinite entanglement with further relationships and actors.

Designing interactions consists in situating and re-situating interactions within these entanglements.

This perspective on interaction, intentionally avoids any notion of *agency* insofar as it could reintroduce the idea of an ultimate animating instance like a *soul*. However, the identification of actors is not only an important aspect of interaction design, but more generally of our interactions with the world. This includes our own identification as actor, as well as the identification of other actors we engage into interaction with. These actors and our interactions with them may be more or less *material* or *cultural*, *real* or *imaginary*, *mechanical* or *intelligent*, and all of it together. Our interactions ultimately involve all parts of our existence. Beyond the question of whether they are animated or not, a legitimate preoccupation in our engagement with other actors concerns whether they are dangerous or not, whether we can rely on them, can be creative with them, or which other actors they remind us of – physically and socially. Questions that, rather than qualities of those actors, concern our relationship and engagement with them.

Leaving aside the inquiry of the animate, in a metaphorical sense, animation finally describes in the act of exploring different aspects of such engagements by creating interactions.

A 11 | Animation = interaction design.

The design of a musical instrument particularly fosters the musician's physical engagement with the instrument and – through the instrument – with an audience. In the contrary, an audio player radically reduces the users' physical interaction to a few buttons and favors the appearance of other actors such as a singer or an orchestra the user engages with through listening. In the scenario of improvising music with *creative machines* (Lewis 2006), the design aims at the emergence of digital actors that a musician can engage with through musical dialog.

2.1.7 Animation as Narrative

Another interaction scenario that largely eliminates the physical engagement of the spectator to maximize the appearance of imaginary actors is that of cinema. In the interaction of watching a movie, the spectators largely renounce on their own bodily activity by sitting down in the dark and immersing into the audiovisual stream of a movie.[50] Their attention is totally turned towards the actions and actors evoked by images and sound.

Recurrent ingredients of the narratives of movies and especially cartoons explore the boundaries between the animate and inanimate. The forces of nature, the behavior of animals, the intelligence of humans, and the mechanics of machines here become interchangeable characteristics of animated characters. Animated movies feature singing and dancing animals, plastic toys caring for their owner, as well as machines falling in love and declaring war. Due to the rapid development of digital animation techniques, the unlimited imagination of 20th century cartoons enters into the realistically staged feature films of the 21st century. These often amusing and more or less profound products of imagination comment and anticipate the evolution of science, technology, and ethics that continually shift the understanding of the boundaries between animate and mechanical, human and nonhuman, intelligent and rote (Riskin 2003).[51] The maintenance of human life by medical machines, the defense of animal rights, the participation of chat-bots in social networks, and the employment of robots in the care for the elderly are examples from real life that illustrate how these evolutions permanently challenge our intuitions.

The cartoon series *Out of the Inkwell* produced in the 1910s and 1920s features Koko the clown emerging from an ink spill (Stacey and Suchman 2012). This example from the early years of cartoon animation[52] illustrates several elements that are omnipresent in animation, even if they may appear more or less explicitly in the storyline of a particular movie. One of these elements is the *mutability of matter* which refers to the play with primary elements of color, form, movement, and behavior that closely relate graphical animation to music composition and sound design. The Disney movie *Fantasia*, released in 1940, perfectly demonstrates this analogy through a choreography of mutating characters and forms that are synchronized with the permanent evolution of sound and music throughout the whole movie.[53] Another aspect of animation that is illustrated by *Koko the clown* concerns the appearance of animation as a narrative element. While in *Out of the Inkwell* it is explicitly part of the story, animation itself is often an implicit narrative element in the scenario that it supports. Illusion, deception, and the *suspension of disbelief* are important factors not only of the '*magic of the cinema*' (Stacey

[50] The actual affordances of a movie theatre obviously exceed this focused view on the interaction of *watching*.

[51] Also cited in (Stacey and Suchman 2012).

[52] The creator of the series, Max Fleischer, invented the rotoscope technique for the production of the cartoons of the series. The technique allowed Fleischer for creating part of the animations by copying frame by frame his brother's performances as clown previously recorded on film.

[53] In sound design, the mutability of matter and motion can for example occur as sound morphing.

and Suchman 2012),[54] but also of other animation crafts such as stage magic and puppeteering. The maintenance of the spectator's distance to the created illusion is finally part of the art.[55]

Huizinga (1949) insists on the etymological origins of *illusion* from *inlusio*, *illudere* or *inludere* (literally *in-play*) revealing an element of *play* which requires the complicity of the deluded. In this sense, the fascination of puppetry performance is nourished by the fact that the illusion of lively characters emerges from the coarse shapes of the puppet's faces that along Their clumsy movements driven through obvious mechanism permanently remind the spectator of the reality behind the illusion without spoiling it. While the most advanced 3D animation technologies in cinema, may pretend to further improve the refinement of illusion and immersion, usually they are all but hidden and, moreover, often strongly featured in the promotion of the movie.

A 12 | Animation itself is a latent self-referring element of narration wherever it occurs.

Even if this brief exploration concerning cinematographic animation is not directly related to sound and interaction design, it is certainly interesting to consider the latent self-referring narrative element of animation also in the design of interactive audio applications.

Conclusion

In this section, animation has been investigated as a key concept in design that largely exceeds interactive audio applications. This investigation included different connotations of animation such as *liveliness*, *bringing into live*, *creating the illusion of life*, and *ensoulment*. It has visited profound motivations in the development of technology that concern the qualities of the living and free agency. At the same time, it has hardly left the domain of sound reproduction and music technology.

After a brief introduction insisting on aspects of *liveliness*, animation has been examined in terms of *bringing into live*. Through insights into the cultural history of sound reproduction technologies, several connections of sound recording to lifelessness and death have been explored. After these explorations, the argumentation has turned towards different modes of animating recorded sound. Especially, the practices developed in hip-hop culture and *musique*

[54] Stacey and Suchman (2012) cite the idea of *suspension of disbelief* coined by Coleridge to underline significant differences in the conception and perception of research projects on artificial life and cinema productions, suggesting that projects in artificial life could '*take cinema as a model for their own practice, in ways that embrace and celebrate their reliance on artifice.*'

[55] Walton (1978) proposes that instead of actually suspending our disbelieve and decreasing our distance to fictional worlds, we rather '*extend ourselves to their level*' arguing that '*we do not stop actually existing when it becomes fictional that we exist*'.

concrète have been considered from this point of view. It has been concluded that the essential elements of sound animation already existed in the 1960s based on analog technologies.

The idea that sound reproduction technologies belong to the domain of automation, led to an investigation of mechanical elements that formalize temporal processes in early musical automata and musical instruments. This investigation has been continued with the examination of digital audio technologies concluding that sound animation consists in an articulation of automation and interaction.

These reflections have been interleaved with an exploration of animation in terms of *ensoulment*. Hereby, Aristotle's *faculties of the soul* guided through an examination of parallel developments in philosophy and technology. This examination showed how the faculties that Aristotle had identified for distinguishing the animate from the inanimate, one after the other, have been conquered by technological developments over several centuries. Over the same period, in philosophy, the idea of a *soul*, inherited from ancient times, was replaced by Descartes' *mind*, and ultimately by Bergson's concept of *free agency*.

The conclusion of this historical overview coincided with the introduction of phenomenology at the beginning of the 20th century, giving up the conceptual separation of mind and body. A citation from Wiener's *cybernetics* that explicitly refers to Bergson's ideas, has confirmed this worldview by proposing a common perspective in the study of the behaviors of humans, animals, natural phenomena, and machines. These ideas have been further consolidated through a brief summary of contemporary concepts of interaction (i.e. *trans-action*, *actor-networks*, *situated action*, *intra-action*) that abandon the distinction between material and social interactions or actors. From this perspective, animation appeared as a metaphorical concept that is ultimately synonymous with interaction design.

Finally, an excursion into graphical animation introduced aspects of animation that are particularly obvious in the context of cinema, but also inform sound and interaction design.

2.2 Reenactment

Introduction

In contrast to the exploration of *animation* in *2.1*, this section introduces the notion of *reenactment* as a specific concept in the design of interactive audio applications based on recorded sound materials. After its introduction, the concept will be extended by further connotations. The idea of *reenactment* is grounded in the concepts of *enaction* and *embodied cognition* (see for ex. Varela et al. 1992) that, drawn out of phenomenology, have strongly influenced studies on perception and cognition since the end of the 20th century. Recently these ideas have been introduced as key concepts in the design of interactive audio systems (Armstrong 2006; Essl and O'Modhrain 2006; Wessel 2006).

The concept of reenactment investigated in this dissertation, proposes a particular perspective on the relationship between the bodily actions of a user interacting with an interactive audio application and the actions evoked by the recorded sounds the application is based on (see figure *2.1*). These evoked actions may be movements and gestures, concrete events, but also practices and further connotations of action.

R 1 | *Reenactment* is a perspective on the relationship between the actions of a listener or performer interacting with an interactive audio application and the actions evoked by recorded sounds the application is based on.

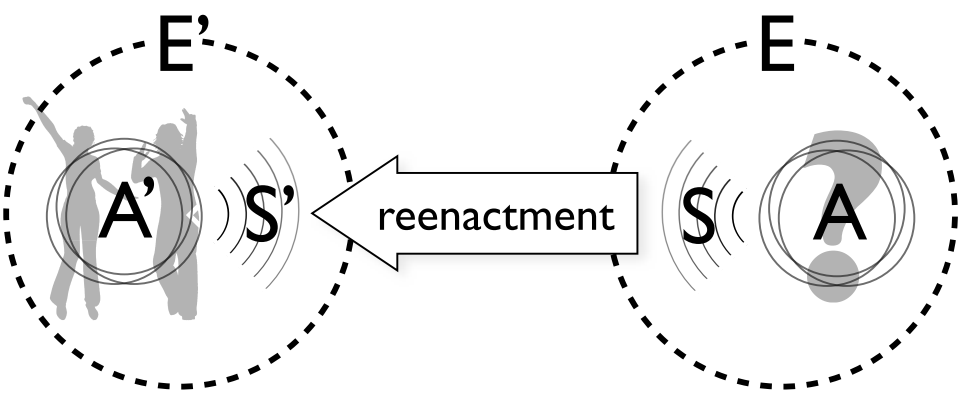


Figure 2.1: Reenactment E' of the original sound S associated with the action A. The recorded sound and actions E are reenacted by sound S' and the actions A'.

Interactive audio applications based on recorded sound implement causal relationships between the users' actions and the transformed rendering[56] of the recorded sound materials. Assuming that the digitized sound materials are *disembodied* action representations (see *2.1.5*), the interaction scenario of an application can be seen as a *reembodiment* of these actions or, more precisely, their *reenactment*. An interaction scenario includes the actions of an individual or a group within the applications environment that generally implies a physical space, objects manipulated by the users (i.e. props), and the projection of sound. Optionally, a scenario can include visual and robotic elements. In addition to these material elements, an interaction scenario may further imply a set of conventions (e.g. in a concert or museum installation) and rules (e.g. of a game). The notion of reenactment can apply to any and all of these elements.

R 2 | The interaction scenario of an audio application based on digitized sound materials can be seen as a reenactment of these materials.

As an example one may consider an application allowing for conducting a virtual orchestra where the movements of a single user captured by a motion capture system control the interactive rendering of an orchestra recording. This scenario consists basically of a space with a motion tracking system around a conductor's podium and music stand. The user enters the podium to conduct a virtual orchestra with a baton or free hand gestures. The interactive audio system of the application realizes the rendering of the orchestra recording as a function of the motion capture data.

Figure *2.2* shows the general schema of an interaction scenario allowing for the reenactment of sound associated with action. In the case of the virtual conducting scenario, the recorded sound (S) of the orchestra that can be, for example, associated with the gestures of a conductor (A).[57] The recorded scene is represented by digital representations (e) that include the recorded sound materials and, optionally, further digital action representations. The user's conducting gestures (A') are captured as the control streams (a'_i)[58] that are the control inputs of the interactive audio system. The system produces the sound streams (s'_i) the user perceives as the sound of a virtual orchestra (S') that he conducts. The scenario as a whole is a reenactment of the recorded orchestra performance.[59]

[56]The term *rendering* here is used as a generalization of *playback*. While *playback* usually refers to the untransformed delivery of recorded sound, *rendering* may include the transformation of recorded sound and their projection within the interaction scenario.

[57]As it will be argued below, the attribution of the action to the gestures of the conductor is an arbitrary choice that not necessarily corresponds to the reality of the recorded sound materials.

[58]The subscripts i and n indicate that these streams are represented by multichannel time series.

[59]The fact that the used recordings could originate from multiple different performances again points to the idea that the attribution of the digitized sound materials to an original sound and action is an arbitrary choice.

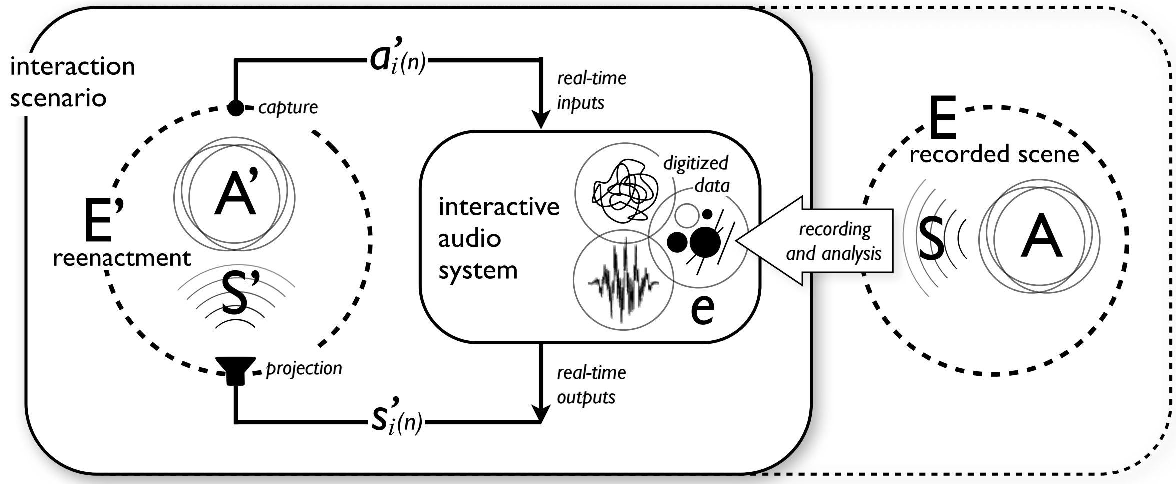


Figure 2.2: Reenactment of a recorded scene E within an interaction scenario E'. The recorded scene E corresponding to sound S associated with action A is represented by the digitized data $e(n)$. Based on $e(n)$ and as a function of the control inputs $a'_i(n)$ captured from the actions A' performed within the environment of the application, the interactive audio system renders the sound streams $s'_i(n)$ that are projected into environment of the application.

The idea of reenactment implies that the interaction scenario of any audio application based on recorded sounds articulates a point of view on these sounds and the actions they evoke. Reenactment can equally apply to the above conducting scenario as to a hip-hop turntable performance that reenacts recorded drum patterns. Even the scenario of somebody putting on a record in his living room can be seen from this point of view. The reenactment of recorded sounds may consist in an imitation of the actions that are evoked by these sounds or as completely different actions. In any case, reenactments restage recorded sound materials within arbitrary causal relationships between bodily actions and sound. These relationships are produced by an interactive audio system through the transformed rendering of the recorded materials.

R 3 | Reenactments restage recorded sound within arbitrary action-sound relationships.

In the virtual conducting example, an interactive audio system creates particular causal relationships between the gestures of the user and the sound produced by the application. Independently of the question whether these relationships actually corresponds to those between the movements of a real conductor and the sound produced by the orchestra he or she conducts, this reenactment creates a particular point of view on the recorded scene. Alternatively, the same sound materials could have been reenacted, for example, through actions resembling those of instrument players, a recording engineer moving sliders, or a dancer.

2.2.1 The Art of Causing Sound

The capacity of producing and modifying sound with our body is part of our experience as well as the capacity of identifying the actions that have produced or modulated a particular sound. In accordance with Aristotle's observations cited in *2.1.4*, our everyday experience permanently confirms sound as the product of an interaction, such as striking an object that causes the propagation of sound towards our ears. Within the same kind of experience, sound may appear sustained by a permanent interaction, such as wind blowing through a tube or an object scratching a solid surface.

In the design and performance of traditional musical instruments, the preservation of the perception of these basic causalities seems to be a major concern. The performance of a piano concert, for example, leaves no doubt about the fact that the pianist produces every detail of the sound reaching the ears of the audience. This energy is transmitted from the movements of the pianist's arms, hands, and fingers to the keyboard, from where it is translated into the movements of the hammers striking the strings. The vibrations of the strings caused by the hammer strokes are transmitted through the air of the concert hall towards the listeners' ears.[60] The complete transparency of the physical causalities within this process is supported by all of its elements. In order to augment this evidence for the listening public, the pianist may exaggerate his or her gestures, or the public may draw nearer and occupy additional seats around the pianist on stage. The listener's musical experience ultimately emerges through a chain of causalities. Even though invisible, these causalities are identified as clearly as the apparent strings of a marionette in a puppet theater (see *2.1.7*).

In *2.1.3* musical instruments and automata have been presented as mechanical formalizations of sound production. More precisely, musical instruments formalize action-sound relationships (see also Jensenius 2007). This applies not only to the entire instruments, but also to their recurrent components such as strings, membranes, tubes, and resonators. Inherited from the long tradition of instrument making and producing sound, also interactive audio applications are based on such formalizations. Moreover, they formalize themselves the production of sound as a function of their users' actions.

R 4 | Interactive audio applications formalize the production of sound as a function of their user's actions.

From this perspective, the history of instrument making appears as the systematic exploration and exploitation of any possible way of causing sound through different materials and

[60] Even the sophisticated hammer and damper mechanisms of the piano, although factually removing some immediacy from the relationship between the pianist's actions and the resulting sound, can be interpreted as a means to clarify this causality that otherwise could be lost in the superposition of too many sounding chords.

interactions within different musical practices. Their freedom from most of the mechanical constraints that substantially determined the design of their ancestors, allows digital applications to explore novel relationships between action and sound (see also Jordà 2005). While the interaction with traditional instruments has been strongly restricted by their mechanical components, many instruments constructed within the past hundred years have reiterated this exploration in the domain of electronic and digital technologies. Often relying on electric circuits and digital processes, the causalities created with these new instruments often have upset – and continue to upset – established knowledges about relationships between action and sound. The relatively short history of electronic and digital musical instruments includes many developments that surprised their audience with the experience of radically new causalities of which we still can imagine the astonishment they produced. These experiences include the phonograph's voices coming out of nowhere, as well as the sounds of a singing saw that is magically induced by Leon Theremin's gestures. The image of Jimmy Hendrix, entertaining the vast public of an open-air concert by causing powerful sounds with the sensitive touch of his tongue, is inscribed into this history, as well as the secret actions of the *Kraftwerk* performers hiding behind their machines.

In his investigation of concepts and methodologies for the study of action-sound relationships in the context of music, Jensenius (2007) proposes to keep a distinction of '*natural action-sound couplings*' from '*artificial action-sound relationships*'. His argumentation reposes on the idea that the experience of '*natural*' action-sound couplings is more stable than the artificial ones. Even if this argument is as difficult to defend as similar arguments aiming at the distinction between *the natural* and *the artificial*, it implies a fundamental idea. It relates our perception of the causality between action and sound to previous experiences of such causalities, as well as their recurrence and reproducibility that constitute our knowledge about our environment.

In this sense, the material interactions that have formed our ears over centuries and millennia strongly resonate in the design of many recent applications. The reference to physics is particularly strong in the work of Cadoz and Luciani et al. (2003). Their interaction scenarios are based on haptic input devices, as well as software modules for the synthesis of action and sound entirely based on physical models.

R 5 | Interactive audio applications explore familiar action-sound relationships and allow for creating unprecedented ones.

Apart from these material interactions, applications can rely on the study and formalization of existing relationships in other domains such as language, music, and dance. Freed from the mechanical constraints of pre-digital technology, explorations in sound and interaction design may concern sound-based interactions beyond those of *causing* sound. Finally, the design of interactive audio applications can openly turn towards the exploration of *listening* as the creative interaction with our environment.[61]

R 6 | Interactive audio applications allow for exploring listening as an active engagement into interaction.

2.2.2 Action and Listening

Sound is an element of complex networks of entangled interactions and actors (see *2.1.6*). In fact, sound itself *is* an action and its vibrating medium can be observed as an actor.[62] However, sound is an action of rather microscopic dimensions and temporal scale. Only sound of high energy and relatively low frequency is perceivable by other senses than audition such as sight, touch, and proprioception. Beyond the microscopic level, sound is an element in interactions of larger dimensions and timescales such as speaking, music making, and other macroscopic bodily actions that also may coincide with sight, touch, and proprioception within multimodal interactions. Listeners may be themselves the actors causing the sound or perceive the actions of another actor through sound. In many situations the actor and action producing a particular sound may not be clearly identified.

The exploration of the manifold relationships between action and listening is a fundamental element in the investigation of reenactment. This investigation departs from a set of assumptions concerning the notion of action. These assumptions include that action is always also interaction and that consequently both terms can be considered as synonymous. Since an action generally can be decomposed into multiple actions, *action* can also denote a set or sequence of actions as well as *activities* in a more general sense.[63] Different qualifiers of actors and actions such as *physical*, *material*, *discursive*, and *cultural*, will be used only to help identifying particular actors or actions and not as categories constituted by essential differences (see *2.1.6*). Moreover, action is assumed to have a consequence for all involved actors.

Finally, an action is always defined within a particular observation and discourse. It is this observation and discourse that creates a particular point of view in which a particular action exists as such. Having said this, it is obvious that *action* and *listening* are related to each other in at least two ways. Without any doubt, listening is an action. At the same time, listening is the observation of action.

[61] Evidently, listening has always been an important factor in musical instrument performance. Nevertheless, the presentation of musical instruments as instruments of listening is rather rare.

[62] While in purely auditory perception a vibrating medium is rarely perceived as an actor, this perception can be reinforced by touch and sight.

[63] Activity, action, and interaction are used quasi synonymously in different grammatical constructs depending on whether a particular argument concerns one or multiple actors such as the *action of* somebody or the *interaction between* somebody and something else.

R 7 | Listening is an action as well as the observation of action.

This postulate reflects the enactive approach to perception and cognition as it has been founded by Varela and Maturana together with Thompson and Rosch (1992). This approach is substantially based upon the *phenomenology of perception* as it has been investigated by Merleau-Ponty (1962). The basic idea of enaction and embodied cognition[64] is that perception and cognition are a common process that involves the whole body. This bodily involvement essentially consists in the ongoing action of exploring our environment by permanent palpation from changing points of view. Together with the resulting experience of perception, these actions constitute themselves the cognitive representations of material, form, space, and movement. Following this idea, the representations of our cognition are nothing else than this experience, even if, in Merleau-Ponty's words, '*instead of attending to the experience of perception, we overlook it in favor of the object perceived.*'[65] Similar approaches linking perception and cognition to sensorimotor activity have been proposed by the psychologists James J. Gibson, as part of his ecological approach to perception (Gibson 1966), and Jerome Bruner, who was the first to introduce the idea of enactive representations in cognition (Bruner 1966).

Noë meticulously defends the enactive approach to perception and cognition claiming that '*our ability to perceive not only depends on, but is constituted by, our possession of … sensorimotor knowledge*' (Noë 2005). As a variation of the illustrative examples Noë gives in his book *Action in Perception*, one may imagine a white ceramic vase sitting on a table. The form and color of this vase are constituted by the way its apparent shape and color changes from different points of view and with different inclinations of the light that it reflects. While inspecting the vase on the table, its *vaseness* emerges from the particular perceptual experience of shapes that coherently change with our movement. Hereby, it is particularly the perfect congruency between the proprioception of our own actions and the changes occurring in the perception of our environment that allows us for relating the involved perceptual experiences across different perceptual modalities. Taking the vase off the table and manipulating it in our hands, its *vaseness* and *whiteness* is encoded in the congruent multi-sensorial patterns implying our vision, touch, audition, and proprioception. Already when the vase was still placed on the table, the particularly regular changes of reflections in respect to our movement seemed to indicate a certain rigidity of the material and a polished surface. These constituting qualities of the vase are confirmed by our touch experiencing a characteristic resistance, as well as a characteristic

[64] As standing for a particular idea of perception and cognition, *enaction* and *embodied cognition* can be considered as synonymous. Nevertheless, by referring to *action*, *body*, and *cognition*, each of the terms puts a specific accent on the concept they commonly represent.

[65] In (Merleau-Ponty 1962), translated from the French original: '*au lieu d'être attentif à l'expérience perceptive, on l'oublie en faveur d l'objet perçu*' (Merleau-Ponty 1945).

adherence when moving the fingers over the surface. The surface appears relatively cool and does not warm up very quickly under the influence of our touch.

In addition to the visual experience, the vase's characteristic qualities are constituted by the experience of sound that, as colors and shapes, permanently changes in congruency with our actions. Characteristic resonant sounds appear in coherence with the vase's interaction with our hands. They are particularly confirmed when striking the surface of the vase with a fingernail revealing its ceramic material and hollowness. Apart from the vase's acoustic responses, our movements around the vase induce subtle changes of the sound reaching our ears – similarly to the visual distortion of shapes. These subtle changes inform us about the position of the vase, as well as particular qualities of the surrounding space.[66] For the experienced explorer of vases, a single strike on the object in absence of sight is enough to gather a considerable amount of information about the vase and the space that surrounds it. This simple *multimodal* example illustrates how the knowledge about an object and its environment is constituted by the actions of exploring it. In this sense, listening is part of the activity and experience of exploring the world or – as Merleau-Ponty employs the term coined by Heidegger – the experience of *being-in-the-world*.

While listening is an action itself, it is also an important element of the perception of action. The above example of a white ceramic vase can be easily modified to a situation in which we listen to the actions of somebody else manipulating the vase in the dark. In this situation, our former experiences of action and sound allow us for deducing somebody's interactions with the vase from the sound they produce. This capacity has been studied in different domains. Studies in psychoacoustics have investigated the mechanisms and capacities of spatial hearing (Blauert 1997), the perception of sounding objects (Davide Rocchesso 2003; Giordano and McAdams 2006), and the analysis of auditory scenes (Bregman 1994). Neurosciences have extended the investigation of hearing and listening to the exploration of cross-modal interactions (Ghirardelli and Scharine 2009) and cognitive action representations. Particular studies in neuroscience have assembled evidence that, independently from the perceptual mode, '*perceived events and planned actions share a common representational domain*' (Prinz 1997). This *common coding principle* has been supported by findings in the domains of movement perception and imitation as well as speech and language perception and production (Prinz 1997). In the domain of sound, this means that the perception of distal action through sound and our own actions producing sound have common cognitive representations that are activated in both cases, making sound and listening to others making sound.

[66]The immediate translation of these subtle changes in timbre as well as interaural intensity and time differences into the perceived movement of a sound source relative to our head, are a perfect example of how we overlook part of our perceptual experience '*in favor of the object perceived*' (see Merleau-Ponty citation above). Even, if it is very difficult to render these changes to our consciousness, the reader may try to listen with only one ear – shutting the other with a hand – to a constant source of broadband noise, such as running water. Attentive listening can reveal subtle changes of the noise timbre as a function of the heads orientation.

The *motor theory of speech perception* defends the idea that speaking and decoding speech by listening share a specialized neuronal module commonly representing phonetic units by *motor primitives* and *articulatory gestures*. While earlier experiments meticulously studied the response of subjects to synthesized speech stimuli (Liberman 1985), recent studies use brain imaging techniques to show the activation – or *resonance* – of the cortical centers involved in the motor control of speech production (Fadiga et al. 2002; Wilson et al. 2004). Similar coincidences of the brain network involved in the execution, perception, and understanding of action have been found for different sound producing actions such as non-speech vocalizations, hand clapping, door knocking, the use of tools, and music performance.

Summarizing a considerable corpus of studies particularly from the 1990s and 2000s, Fadiga et al. (2009) propose that this network is '*tuned to detect and represent complex hierarchical dependencies, regardless of modality and use*' postulating that '*language, action, and music share a common syntactic-like structure*'. From this point of view, the hierarchical structuring of language and music appear as the upper part of larger hierarchical action structures based muscle activations. The similarity between speech and other hierarchically organized sequential actions has also been revealed by Greenfield (1991) crossing studies of children's development of language with studies of their use of objects and tools.

Auditory-motor interactions in the perception and production of music have been studied by many authors (Zatorre et al. 2007; Haueisen and Knösche 2001; Lahav et al. 2005; Chen et al. 2008). These studies confirm that the findings concerning the perception and production of speech and language also apply to music. The relationship between language and music in terms of the perception and cognition of timing and rhythm, pitch and melody, as well as syntax and meaning has been comprehensively investigated by Patel (2010). His *shared syntactic integration resource hypothesis* (SSIRH) postulates that '*linguistic and musical syntactic representations are stored in distinct brain networks . . . , whereas there is overlap in the networks which provide neural resources for the activation of stored syntactic representations.*'[67]

Further studies have given insights into the development of individuals, as well as of the evolution of the human brain (Fadiga et al. 2009). The comparison of specific networks in the brain of monkeys and humans seems to suggest that the networks, specialized in the treatment of human language, could have been developed out of the areas that are involved in the treatment of bodily action. For what concerns individual development, it has been shown that the capacity for constructing hierarchical cognitive representations of action and language is already available to newborn infants (Gervain et al. 2008).[68] Particular skills and knowledge are considered as acquired by learning. As a study of novices who learn to play melodies on a

[67] In (Patel et al. 2009b).

[68] Cited in (Fadiga et al. 2009).

piano has shown that cognitive representations associating sound with action can be acquired by short-term learning (Lahav et al. 2007).

R 8 | Listening and producing sound in language, music, and other actions share common cognitive resources processing the hierarchical and sequential organization of action.

The above example of the exploration of a white ceramic vase kept a particularly analytic distance and avoided mentioning the emotional aspects of action. Listening to the sound produced by somebody else manipulating the vase could evoke for example the fragility of the vase. The perceived sound may evoke danger through the recognition or anticipation of actions that could break the vase. Additionally, the actions of the person manipulating the vase may express – consciously or unconsciously – other emotional states such as calmness or nervousness.

Already this very simple example shows how action representations encode emotion as the *style* in which actions are performed. Moreover, it suggests a connection between emotion and the anticipation of the consequences of a particular action. Giving a comprehensive overview of the role of emotion in cognition, Damasio (1994) emphasizes the crucial role of emotion in the estimation of consequences and the planning of action. As Damasio and many other neurobiologists, psychologists, and philosophers over the past decades, Zhu and Thagard (2002) underline the supportive role of emotions in reasoning especially in social interactions. They particularly insist on the automatic character of emotional response that '*penetrates into almost every aspect of our mental life, including perception, social cognition, motor performance, the setting of behavior goals and motivations, and subjective evaluations and judgements*'.

Emotions, as an important aspect of our – embodied – cognition, affect not only the planning and performance of deliberate actions, but also can be the origin of automatic reactions such as laughter and facial expressions, as well as changes of physiological states such as respiration and heart rate.[69] In all of these cases, action is perceived as expressing emotion. Moreover, our perception of emotions expressed by our own actions seems to reinforce our emotional states.[70] Reviewing neuroimaging and behavioral studies contributing to the understanding of the relationship between motion and emotion, Molnar-Szakacs and Overy (2006) claim that '*humans may comprehend all communicative signals ... in terms of their understanding of the motor action behind that signal, and furthermore, in terms of the intention behind that motor*

[69] Krumhansl has shown how music induces changes in breathing rate, respiration, temperature, and heart rate (Krumhansl 1997).

[70] Following the *facial feedback hypothesis* already expressed by Charles Darwin and William James at the end of the 19th century, Strack et al. have shown how the proprioception of facial expressions supports and reinforces emotional states (Strack et al. 1988).

action', concluding that '*[t]he expressive nature of any human action or vocalization sends a signal of the intentional and emotional state of the executor, such that even footsteps can be correctly interpreted as conveying simple emotions*'.

The role of expectation and anticipation, especially in the perception of music, but also in language and action more generally, has been investigated by Huron (2006). His *ITPRA theory* distinguishes five '*expectation-related emotion response systems*', including *imagination*, *tension*, *prediction*, *reaction response*, and *appraisal*. Repeatedly referring to the evolution of human cognition and the function of emotional responses for survival, Huron gives many examples to illustrate different aspects of expectation. Apart from music, these examples include the behavior of animals, speech, and narration.

R 9 | Listening and producing sound share common encodings of emotion and intentionality based on action and expectation.

According to enactivism, our knowledge of the world is constituted by the experience of our interactions. Our perception and cognition are based on the cognitive *re-enactment*[71] of this knowledge. From this perspective, what has been thought of as separated domains merges into a common process of enacted cognition. This process includes our actions exploring the environment, our perception of actions and actors within this environment, as well as our operation with knowledge about ourselves and our environment. The phenomenological concepts underlying these ideas have been embraced by epistemology and fundamentally changed the understanding of knowledge. Dewey's *knowing* (Dewey and Bentley 1949), Bruner's concept of *enactive knowledge* (Bruner 1966), and Polanyi's *tacit knowledge* (Polanyi 1966) are products of this evolution, as well as *embodied knowledge* (Johnson 1989), *bodily knowledge* (Parviainen 2002), *sensorimotor knowledge* (Noë 2005), and *thing knowledge* (Baird 2004).

Listening to action consists in the decoding of sound patterns into the actions that could have produced them. Moreover, it implies the evaluation of different qualities associated with these actions and the involved actors. Sound, here serves primarily as a sequence of cues evoking the motor primitives and articulatory gestures involved in the production of the actions projected onto sound by the listener. The mechanisms of listening match and anticipate complex acoustic signals in the same way that actions are planned and concretized in a continuous flow that is hierarchically organized on multiple levels of temporality.[72] Depending on the category of perceived actions (e.g. a certain language or style of music) this flow may comply with a particular syntax and correspond to particular qualities that are associated with emotions

[71] Fadiga et al. use the term *re-enactment* in this sense (Fadiga et al. 2009).

[72] The idea of different *levels of temporality* has been introduced and comprehensively discussed by Fraser (1990).

and intentionalities. From this perspective, sound is an auditive representation of action and listening is the act of decoding intentions, emotions, and possible consequences encoded into actions.[73] This decoding is based on the re-enactment of previous experiences.

R 10 | Listening is the decoding of intentions, emotions, and consequences of actions perceived through sound based on the re-enactment of previous experiences.

This notion of re-enactment refers to the internal process of listening that does not require any externalized action of the listener. However, as already mentioned above, listening generally implies further bodily action such as physiological changes, movements that support listening, and actions entrained to the perceived sound.

The entanglement of our listening with further actions may be illustrated by the example of a rustling sound occurring in a dense and dark forest. Perceiving the sound, suggesting the presence of another being, we would probably immediately enter a state of excitement or anxiety that affects, among others, our heart rate and breathing. Completely immobile and silent we re-enact former experiences of actions that could have caused the rustling in our nervous system, trying to understand what or who causes the rustling and which intentions are involved. Anticipating the evolution of the perceived sound and action, we control our movements to avoid disturbing our perception and automatically adjust them to the rhythm of the rustling. We move our head to enhance our audition. Sudden sound events scaring us or causing our particular attention lead to an additional synchronization of our actions to the perceived sound. These actions, very immediately related to the perceived sound, may be interwoven with our planning and execution of further actions that aim at assuring our security or a better perspective of observation.

The cognitive processes, physiological states, and actions described in this example, are very similar to the bodily processes and actions that are involved in music listening. Similar to the rustling sound in this scene, also music induces the change of physiological states and automatic expressions of emotion. It entrains the listener's movements to the perceived music, while reacting to the flow of sound events and anticipating their evolution.

R 11 | Within the process of listening, perceived action is extended through and entangled with performed action.

[73]Insofar as listening can be described as the act of decoding action from sound, making sound becomes an act of encoding. Especially when making sound is the actual aim of a particular action such as in music performance and speaking, this encoding of intentions and emotions into sound is what generally is referred to as *expression*.

Godøy et al. (2005) have studied the extension of listening through gestures within an experimental setup featuring the performance of '*air-instruments*'. For this experiment, subjects were asked to reenact recorded piano performances of different music styles with gestures in the air. The air-performances have been assessed through a protocol of observations regarding particular correspondences to '*movements necessary for any real performance of the excerpts*' on the piano. The assessment considered various levels of detail including the density of gestures, relative locations of on the keyboard, synchrony of movements to notes and phrases, extent and velocity of gestures, articulatory details such as accents, staccato, and legato. The experiment clearly shows that the correspondence of the air-performance to real piano performance gestures increases with the subjects own musical experience and in particular their training in piano performance. Even if the study focused on gestures imitating the performance of musical instruments, the authors also observed '*more vague sound-tracing gestures, such as in following melodic contours, rhythmical/textural patterns or timbral/dynamical evolutions with hands, arms, torso, or whole body.*'[74] *Sound-tracing* gestures have been the object of another study by Godøy et al. (2006), for which subjects were asked draw gestures with a pen on a digital tablet in response to recorded sound fragments. In the same series of experiments, Haga (2008) explored the correspondence of free dance movements to music.[75] Godøy (2009a) refers to the sound- and music-related gestures explored in these studies as *gestural renderings* that translate perceived sound features to gestures and reveal '*amodal gestural images of musical sound*'.

These *gestural images* may correspond in terms of their geometry to particular morphological aspects of sound,[76] such as melodic contours, rhythmical patterns, the evolution of timbre and texture, or the dynamics of the gestures. These morphological aspects generally can be represented by the evolution of corresponding sound features such as the density of events, speed, and effort. While the sound-tracing experiment (Godøy 2009a) focuses on geometric congruencies (i.e. trajectories), the study of spontaneous dance movements (Haga 2008) mainly examines congruencies of movement dynamics. However, both cases concern congruencies of gesture and movement features with perceived features of sound and music.[77] In each of these experiments (Godøy et al. 2005, 2006; Haga 2008), gesture and movement are regarded

[74]In Godøy et al. (2005).

[75]Further studies of congruencies between dance movements and music using motion capture and sound analysis has been conducted by Naveda (2011). While in (Haga 2008) the participants were asked to spontaneously improvise movements on very short musical excerpts of twentieth-century Western music without steady beat, Naveda's studies focus on Samba explicitly referring to Afro-Brazilian tradition.

[76]Without actually applying the distinctions proposed in Schaffer's *typo-morphology* of sound (Schaeffer 1966), *sound morphology* here is used in the sense of *morphological aspects* of sound referring to abstract features and forms perceived in sound – including but not limited to musical features and forms – as it is proposed by Schaeffer's idea of *reduced listening* (Chion 1983) and opposed to other modes of listening such as when aiming at the identification of actions, intentions, and relationships occurring in a particular sound scene.

[77]While the cited experiments are mainly based on the transcription of observations several authors have also evaluated these congruencies quantitatively using statistical modeling (Caramiaux et al. 2009, 2010b, 2011; Nymoen et al. 2011, 2012; Caramiaux 2012).

as externalized and displayed enactments of listening. They show how listening extends into displayed externalized action and that this action reveals how sound and music are perceived and understood by the listener.[78]

The idea of listening as entangled action is supported by the concept of *entrainment*. Phenomena of entrainment, such as the synchronization of movements to a regular beat, have been studied by several authors (see for ex. Leman 2008). Underlining the importance of entrainment in the study of music, Clayton et al. (2004) propose to apply the mechanical model of entrainment, first identified by Huygens, to describe actions occurring in music making, music listening, dance, and conversation as phenomena of resonance. Their study provides a comprehensive summary of physical, physiological, biological, and social processes that can be described as rhythmic phenomena that imply the entrainment of plants, bacteria, animals, and humans. The examples of '*endogenous or naturally occurring rhythms within the human body*' include neuronal activity and brain waves, heart beating and blood circulation, respiration, eye blinking and rapid eye movement, swallowing and sucking, yawning as well as daily and monthly cycles (i.e. menstruation). Also chewing, locomotion and virtually all human movements, can be seen as '*inherently rhythmic*'. While insisting on the importance of *self-entrainment* for the production of rhythm, Clayton et al. (2004) describe entrainment as an '*interpersonal or social*' phenomena claiming that '*[e]ntrainment to and through music needs to be seen as a particular case of entrainment in social interaction.*'[79]

While these investigations of entrainment essentially stay in the framework of direct exchange between humans, Leman (2008) has investigated entrainment in the framework of *Embodied Music Cognition and Mediation Technology*. He includes *cultural objects* such as musical instruments and sound reproduction technologies into the reasoning on entrainment and concludes that '*action/perception processes of the subject realize the link between energies and cultural objects through resonances (entrainment)*'.

Even if the capacity of entrainment to regular oscillation has been observed for animals such as crickets, frogs, and fireflies, Patel et al. (2009a) insist on important differences between these synchronized acoustic displays and the human ability of synchronizing movements to music. Human beat perception and synchronization '*involves a periodic motor response to complex sound sequences..., can adjust to a broad range of tempi, and is crossmodal.*'[80] Moreover, it is strongly related to auditory-motor interactions in human cognition.[81]

[78] Although if in each of these experiments subjects are explicitly asked to translate their perception of sound and music into movements, (Godøy 2009a) also relates this work to movements that occur spontaneously in listening to music.

[79] In (Clayton et al. 2004).

[80] In (Patel et al. 2009a).

[81] The discovery of a parrot's ability to synchronize movements to music, has recently confirmed the *vocal learning and rhythmic synchronization hypothesis* according to which the ability of rhythmic synchronization is related to vocal learning, an ability humans share with parrots, dolphins, and seals (Patel 2006).

With these findings it becomes clear that the human beat perception and synchronization relies on complex auditory-motor interactions. Mechanical models of entrainment like the synchronization of pendulums apply to these interactions at most metaphorically. Descriptions of entrainment into music regularly use the metaphors of *resonance* and *tuning-in*. The understanding of entrainment in terms of *tuning-in* implies the listener's active involvement into the process of entrainment. While studies of entrainment mostly refer to the synchronization of periodic actions, the engagement into physical and social interactions often requires the coordinated recognition, anticipation, and production of much more complex temporal structures.

Even if the understanding of the cognitive mechanisms involved in complex interactions is far from complete, the studies of perception and cognition seem to suggest that the notion of entrainment could be extended to interactions that largely exceed the synchronization to regular beats. Engaging into a conversation, collectively interpreting or improvising music, controlling complex machines, and navigating a boat in stormy weather, are examples of interactions that require the adjustment of actions to the actions of other actors on multiple levels of temporality.[82] Some of the studies cited above suggest the existence of hierarchical action representations that are involved in the perception, planning, and performance of actions. Picking up the metaphors cited above, it seems that we can *tune into* interactions with other actors, whereby their actions *resonate* in our cognitive processes of observing, predicting, and performing action.

In the light of the above discussion, the act of listening includes the understanding and performance of action. Both aspects are closely coordinated and mutually influential. This process of listening is not limited to the simulation of action as a mute cognitive process, but also implies the active engagement into material and social interactions. While when driving a car, for example, the role of listening is to support the performance of bodily action, in the audience of a concert our bodily actions support our listening. However, in general it is difficult to isolate both aspects, entangled into a common cognitive process.

R 12 | In listening, the observation, understanding, and performance of action are entangled into a common cognitive process.

In the interaction with audio applications, listening is mediated through technology that not only transmits and defers action (i.e. recorded sound) in space and time, but also re-situates it into a new context. Technology here may appear itself as a sonic object or environment the user interacts with or as an *instrumentation* of the user's listening. While in the first case technology itself is perceived an actor, in the latter case it acts as mediator for the interaction with other

[82]Staying in the beat, here refers to one of many parallel aspects of action on one of many levels of temporal structures involved in music performance.

actors. The concept of reenactment includes both of these aspects that generally are entangled in the interaction scenarios of interactive audio applications.

R 13 | Interactive audio applications may appear themselves as sonic objects or environments the user interacts with or as an instrumentation of listening mediating the user's interactions.

This perspective connects audio applications to sound reproduction technologies in the legacy of early applications such as the telephone and the phonograph that since have been reincarnated into teleconferencing, home cinema, and mobile music systems. Even if the faithful reproduction of sound is not a major concern of this investigation, the perspective of sound reproduction contributes to the understanding of interactive audio applications based on recorded sound materials and the concept of reenactment.

2.2.3 Sound Reproduction and Interaction

An important aspect of sound reproduction is that the actors that have produced the reproduced sound are largely absent from the scenario of reproduction.[83] As many other sound installations produced within the last decades, Janet Cardiff's piece *The Forty Part Motet* perfectly illustrates this absence and, moreover, includes it as an apparent element into her artistic discourse (Berwick 2006). The piece consist in the reproduction of a choir performing *Spem in Alium* by Thomas Tallis on forty loudspeakers that are set up in a circle in a reverberant room. Each loudspeaker reproduces the voice of a single singer of a forty-voice choir.[84] The recordings are captured very close to each singer so that, in addition to singing, each of the sound channels includes breathing and other utterings by the signers. This creates a strong physical presence of the – otherwise absent – singers, especially when the listener approaches one of the loudspeakers more closely.

Following the discussion of perception and cognition above, insofar as the actors and actions of a particular scene are reproduced by the playback of a recording, the listeners can cognitively re-enact the recorded scene of actors and actions. Hereby, they complete the reproduced sound with their tacit knowledge of the singers and singing. However, the above argumentation suggests that it is difficult to clearly delimit the cognitive re-enactment of sound by the immobile listener from affective responses and further engagements with the repro-

[83]The visual absence of the source of sound also defines *acousmatic music*. Pierre Schaeffer and the French novelist Jérôme Peignot derived the term *acousmatic* from the greek *akousmatikoi* referring to pupils of Pythagoras who followed their teachers readings behind a veil to enhance their concentration.

[84]The piece is actually written for eight choirs of five voices each.

duced sound scene and the environment of its reproduction. Within any scenario that involves the reproduction of recorded sound, the actions evoked and induced by the sound reproduction, interweave with further actions of the listener that are not necessarily related to listening.

Many studies have investigated the impact of music listening on interactions that not directly concern the reproduced sounds or the devices for their reproduction. Bolivar et al. (1994), for example, have studied the effect on music on the perception of unrelated visual actions. Other authors have studied how music affects the behavior of customers in supermarkets (Milliman 1982) and work performance (Lesiuk 2005). In summary, our interaction with reproduced sound ultimately includes interactions with sound reproduction technologies (e.g. slipping a coin into a jukebox and choosing a song), actions directly related to sound (e.g. listening and dancing), and actions – observed or performed – that are arbitrarily concurrent with sound. The latter may refer to the rich activities of an urban environment observed while listening to music through the headphones of a mobile music player, or the actions of dishwashing while listening to the radio.

The development of sound reproduction technologies like high-fidelity stereo or surround systems have for a long time focused on the representation and reproduction of timbre and basic aspects of spatiality. The reproduction of these aspects generally allow the listener to understand a reproduced sound from the point of view of a blind and totally immobilized listener emerged in the recorded scene.[85] In reproduction scenarios based on these technologies, any movement of the listener rather reveals the details of the reproduction technology itself (i.e. the loudspeakers) than the reproduced scene of actions.[86]

More recent sound reproduction technologies, however, include the representation and reproduction of spatial qualities that enable the listeners to explore arbitrary sound scenes by moving their head and walking freely through the sound environment generated by a particular audio application. These technologies are, for example, based on binaural techniques including head tracking (Begault 1994) or the synthesis of sound fields using techniques such as *Wave Field Synthesis* (WFS) (Berkhout et al. 1993) and *Ambisonics* (Gerzon 1985). They allow for adapting the recorded sound scene to the scenario of its reproduction, taking into account the listeners' interaction moving within the projected sound space. In addition, these technologies may by used to create distinct sound sources and their trajectories that are explicitly represented within the audio rendering system.

In this sense, interactive audio applications represent a logical evolution of sound reproduction technologies in which the reproduction of sound is extended to further aspects of interac-

[85]The virtual immobility of the listener listening to a stereo recording, after all, can be seen as a fair approximation of a listener's situation in a concert hall.

[86]The term *reproduction scenario* here is used as a particular case of *interaction scenario*, as defined above, focusing on sound reproduction. In the following, the term *scenario* will be distinguished from the term *scene* describing the situation in which a recorded sound has been originally produced or in which its production has been imagined.

tion. These extensions allow the listener to further engage into the interaction with a reproduced sound environment and particular sound sources.

R 14 | Interactive audio applications represent a logical evolution of sound reproduction technologies, allowing the listener for further engaging with the reproduced sound environments and sound sources

This idea of interactive audio applications as sound reproduction technologies has to be further clarified. In fact, sound reproduction technologies are *always* interactive and *always* support listening as an interaction. While *nothing but* listening already is an interaction, a record player, for example, additionally allows the listener for choosing and starting a record, as well as for adjusting the volume and exploring the sound space created through the loudspeakers. The so-called *interactive* applications, extend these existing interactions.[87]

Glenn Gould has anticipated the possibilities of future sound reproduction technologies extrapolating the interactions provided by the record players of the 1960s.

> *Dial twiddling is in its limited way an interpretative act. Forty years ago the listener had the option of flicking a switch inscribed "on" and "off" and, with an up-to-date machine, perhaps modulating the volume just a bit. Today, the variety of controls made available to him requires analytical judgment. And these controls are but primitive, regulatory devices, compared to those participational possibilities which the listener will enjoy once current laboratory techniques have been appropriated by home playback devices. ... There is, in fact, nothing to prevent a dedicated connoisseur from ... exercising such interpretive predilections as will permit him to create his own ideal performance*[88]

Moreover, insofar as somebody somewhere and sometimes recorded a particular sound scene that is reproduced to a listener, listening is always an interaction with the actors that produced the reproduced sound. Even if the sound is transmitted and deferred over long distances and times, from a transactional point of view, the actual process of mediation still appears as a complex mutual interaction between both actors and their respective environments (see *2.1.6* and *2.2.5*). Generally, the listeners and their environments influence as much the recording situation as the recorded sound affects the situation in which it is reproduced. From the listeners' perspective, the interaction with the reproduced scene is entangled with many other material

[87]Nevertheless, this argumentation will continue to distinguish *interactive* applications from others. This distinction will be further investigated in *2.2.5* below.

[88]From the section *The Participant Listener* in *The Prospects of Recording* (Gould 2004).

and social interactions involved in recording and reproduction. The actors presence may be tangible, like the record player and loudspeakers, or revealed by the artifacts they produce (e.g. the recording equipment). Moreover, they may be present through their description in the booklet of the recording.

The idea that any reproduction of sound and any audio application are already interactive, suggests that the adjective *interactive* is not a distinctive quality of particular applications, but a particular perspective that can be applied to any application. This perspective explicitly includes all interactions of the listener into the reasoning on the applications' design.

R 15 | *Interactivity* is not a distinctive quality of particular applications, but a perspective that can be applied to any application.

Whether a sound has been mediated by sound reproduction technologies or not, our listening is highly attentive to congruencies between actions within the same and across different modalities on virtually any level of temporality. Our sensibility for congruencies allows us to integrate simultaneous stimuli of different modes in the perception of action (Calvert et al. 2004). Moreover, it enables our recognition and distinction of simultaneous and differed occurrences of the same or different actions. Hereby, we skillfully recognize the identical reproduction of actions by their reflection in a mirror,[89] by an echo or through technology, from imitations or similar actions performed by independent actors.[90] In the former case, the comparison between multiple occurrences of the same action may be expressed in terms of *amplification*, *attenuation*, *transformation*, *distortion*, and *delay*. For the latter case, temporal congruencies are understood in terms of *coincidence* and *synchronicity*, while *asynchronicity* and *variation* refer to incongruencies. The congruency and incongruency between actions can be *temporal* and *spatial* – or *geometric* –, but also *semantic* and *formal* (Bolivar et al. 1994).

The understanding of recurrent action includes notions like *regularity*, *rhythm* and *tempo*, *frequency* or *period*. Over multiple occurrences of a recurrent action, a listener may establish notions of *evolution* or *behavior*. Congruencies between recurrent actions can be expressed in terms of *synchronization*, *tuning* and *consonance*. Moreover, different recurrent actions can be *in phase* and *in rhythm*, *beat*, or *groove*. Depending on their congruencies and incongruencies, the polyphony of multiple actions may be perceived as *harmony* or *cacophony*. In the same

[89] *Mirror self-recognition test* (Gallup 1970) show the capacity of humans and certain species of animals (Nielsen and Dissanayake 2004) to recognize themselves in a mirror. Most animals take the mirror image for the appearance of another animal or do not react at all – which also can be taken as a sign for the ability to distinguish the image of their actions from the actions of other animals.

[90] The success of Edison's *tone tests* (Welch and Burt 1994), and similar experiments that compared the perception of the live performance of a singer to a gramophone reproduction, show how the capacities of distinguishing human actors and from reproductions may evolve with the development of technology and listening.

sense, the superposition of multiple similar sounds may appear as decorrelated *textures*, such as the superposition of multiple voices of a choir or instrument group[91] and the noise of leafs moving in the wind.[92] But when they are strongly correlated, they also can blend into the perception of timbre and the spatial qualities of a particular sound source and its environment (Blauert 1997).

While the above explorations mainly focus on temporal aspects of action, the perception of congruencies also concerns spatial aspects. Many efforts in the development of sound reproduction technologies in the context of audiovisual systems concern the ability to spatialize sound sources coherently with visual cues. These developments find their application in movie theaters (Kerins 2011) and teleconferencing systems (Kilgore et al. 2003; de Bruijn 2004), as well as sound reinforcement systems for live performance (Hoeg et al. 1983). Studies of mutual influences of acoustic and visual cues in the localization of sound sources have shown how existing congruencies between auditive and visual cues reinforce precision and speed of the localization of objects (Bolia et al. 1999; Gondan et al. 2005). The *ventriloquist effect* (Howard and Templeton 1966; Bertelson and Radeau 1981) shows how in the case of spatially incongruent but temporally congruent visual and auditive cues, the listener may create a consistent perception of a located sound source. The influence of sound on the perceived temporality of visual cues often is referred to as *temporal ventriloquism* (Morein-Zamir et al. 2003). Another phenomenon revealing how we resolve cross-modal incongruencies in the perception of speech is known as the *McGurk effect* (McGurck and MacDonald 1976).

Moreover, the notion of congruencies can be related to anticipation and expectation. Surprise, for example, can be seen as an incongruency of the actual evolution of a perceived action with the expected evolution. Finally, the discussion of entrainment in *2.2.2*, proposing a generalization of the concept of entrainment to more complex interactions, can be related to the idea of congruency. In this sense, the process of *tuning into* an interaction consists in creating congruencies between our own actions and those of other actors within a process that is based on our capacities of anticipating action (Huron 2006).

In summary, the perception of the interdependence of actions and the identification of actors – including ourselves[93] – relies on the observation of congruencies. Our sensibility to congruencies of different qualities, within different timescales, and across perceptual modes, is a fundamental factor in the perception of interactions and in our engagement into interactions with other actors.

[91]The possibility to synthesize multiple – decorrelated – voices of a choir by multiple transformations of a single sound has for example been explored Schnell et al. (2000).

[92]The perception of sound textures as statistical auditory phenomena has been studied for example by McDermott and Simoncelli (2011).

[93]The recognition of our own actions and, through these actions, of ourselves has for example be investigated by Knoblich and Flach (2003) (see also Flach et al. 2004).

R 16 | Our sensibility to congruencies of different qualities, within different timescales, and across perceptual modes, is a fundamental factor in our engagement into interactions.

In sound reproduction, congruencies reveal and create relationships between all involved actors. These actors include those which have produced the recorded sounds, the involved technologies, the listener, and further actors of the listener's environment. Sound reproduction technologies can be considered as a medium for transmitting and deferring the actions of a recorded scene into the listening situation. Within this perspective, the design of reproduction scenarios seeks to maximize the interaction with the actors of the recorded scene and to minimize the listener's attention to the mediation technology itself. *Virtual reality* and *telepresence* (Riva et al. 2003) are paradigms that foster such approaches in the design of mediation technologies and interactive systems. In the light of the above discussion, the design of reproduction scenarios following these approaches would aim at rendering a faithful image of the recorded scene in the listening situation. Here, sound reproduction could be reduced to a question of perfecting the congruency between the actions of the original recorded scene and those reaching the listener in the reproduction scenario.

However, apart from the problematic of *faithful* sound reproduction through the capture, transmission, and rendering of sound, a major concern in design of sound reproduction technologies is the fact that the recorded scene and the listening situation are fundamentally incongruent. The factual absence of the recorded actors from the listening situation is confirmed by the absence of certain modes of perception and expected responses to the listener's actions. In consequence, the rendering of the recorded scene has to be adapted to the listening situation. Examples that illustrate this challenge are the rendering of an orchestra concert transmitted from the *Musikvereinssaal* in Vienna on an old kitchen radio and the reproduction of an open-air rock concert on a car stereo.

The congruency between a recorded sound source and the listener's perception is mediated by many actors in the chain of audio reproduction. This chain can be arbitrarily decomposed into different human and non-human actors. Each decomposition creates a different point of view on mediation. A basic scenario of sound recording and reproduction from a technical point of view could involve, for example, a microphone, a recorder, a storage media, a player, and loudspeakers. While this chain focuses on technical devices, other points of view could include the actions of boom operators, recording engineers, producers, and DJs. On the recording side, all involved actors are required to contribute to the capture of signals that are congruent with certain actions, excluding many others that are not desired on the recording. On the other end of the reproduction chain, the challenge of reproduction is to preserve certain of these congruencies while creating others that are coherent with the listener's interaction. These congruencies include a particular perspective of perception, as well as the listener's actions and expectations. Sound reproduction, here is defined by mediated congruencies of the actors and actions constituting a recorded scene with the scenario of its reproduction.[94]

R 17 | Sound reproduction is defined by mediated congruencies of the actors and actions constituting a recorded scene with the scenario of its reproduction.

The different levels of temporality in our interaction with sound and music range from the level of sound waves to the macro structure of musical works. Moreover, they may involve the evolution of sound environments over larger periods of time. On the micro-level, sound reproduction technologies congruently capture, preserve, and transform sound signals in form of air pressure waves, electrical signals, and digital streams of information. On larger time-scales, the evolution of recorded sound features coincides and interferes with macroscopic actions such as human body movements, machine motion, and natural phenomena.

The congruencies on different levels of temporality involved in sound reproduction may be illustrated by the example of listeners interacting with a record player. The record player itself becomes in these examples the embodiment of an interactive process that mediates between the actions recorded through sound and the listeners' actions. The listeners here can interact with different components of the record player that they perceive as actors in this process. The listeners may perceive the micro-level interactions of the record player's needle with the cavities of the record grooves that are further congruent with the vibrations of the loudspeaker membrane. On larger dimensions, may appear the congruencies between the motor driving the rotation of the turntable, the turning vinyl disc, and the arm moving towards the center of the record.[95] Any interaction of the listeners with the record player – intentional or accidental –, is inevitably articulated with the recorded sound.

[94]A perfect example of how reproduction technologies may represent the listener in the recording situation is the *dummy head*. Interestingly, the better the dummy head corresponds to the listener's head the better the recorded sound scene can be reproduced by the sound played back to listener by head phones. Theoretically, a perfect geometric and *acoustic* congruency of the dummy head with the listener's head should result in a complete congruency of the signals entering the dummy heads microphone with those entering the ears of the listener. In consequence, the sound reproduction chain would vanish from the listener's auditive perception as a completely transparent medium. Unfortunately, also in the case the slightest incongruency of the listener's movements with those of the dummy head, reveals the medium as an independent actor in the listener's auditive interaction. These ideas will be further developed below.

[95]Interestingly, even though the turntable turns with constant speed, the relative speed of the needle in the record groove is not constant but decreasing with the needle advancing from the outer edge of the record towards its center. Nevertheless, since the record grooves have been written into the record in a way that is perfectly congruent with the way that it is read, this change of speed is usually inaudible when using modern record players.

An example that illustrates interactions on a larger timescale may be the following description of a music lover who appreciates the reproduction of an orchestra concert on a home stereo system. The different movements of the composition give a structure to the concert that also determines the partitioning of its recording on multiple records. This temporal structure further coincides with the actions of the listener putting up and turning the vinyl records to start each part of the recording. Within this interaction, the listener may sit down in front of a pair of loudspeakers and stand up to operate the record player. Additionally, the listener may synchronize – consciously or unconsciously – the performance of many other actions to the temporal structure of the recorded concert. He or she may wait for the end of a movement to prepare a drink or to visit the bathroom. Moreover, the listener may transform and adapt the reproduced orchestra performance to generate congruencies with other interactions. For example, he or she may pause the record player while picking up the phone. Through the interactions with the record player, the listener can not only modulate the intervals between the movements and interrupt the concert at any moment, but also arbitrarily shorten and repeat any of its parts. The act of reproducing sound, here becomes a negotiation between congruencies with the recorded scene (i.e. the orchestra performance) and congruencies with the situation of its reproduction.

Audio processing and editing techniques rely on temporal congruencies throughout all levels of temporality to preserve certain sound features and their temporal evolution while transforming others. Similar to the way manual tape editing privileged splices that coincide with silences and transients, audio processing techniques that allow for the temporal transformation of recorded sound (i.e. pitch transposition and time-stretching) have to respect temporal congruencies with structures such elementary waveforms and articulations to minimize audible artifacts (Peeters 2001; Roebel 2003, 2010). Beyond sound reproduction, any sound transformation is defined by a negotiation of congruencies between manifold aspects of the original sound and the transformation process. Both together ultimately determine the result of the transformation.

The music lover's interactions described above and the performance of a hip-hop DJ, scratching the recording of a drum performance, are different instances of this negotiation. In common turntable performance techniques, the performer controls with one hand the movement of the turntable to select, playback, and transform single events and sequences of the recorded sound. With the other hand, the performer operates a crossfader increasing the congruency of the produced sound with the desired rhythm (Hansen 2010). In the resulting sound, the actions of the recorded drummer and the actions of the DJ are entangled. While certain aspects and passages of the drumming are reproduced as they are recorded, other aspects of the produced sound are congruent with the actions of the performing DJ.

In all of these processes, the faithful reproduction of recorded sound and its adaptation within the listeners' – and performers' – interactions go hand in hand as coexisting and partially contradictory challenges.

R 18 | The faithful reproduction of recorded sound and its adaptation to the listener's interactions go hand in hand as coexisting and partially contradictory challenges.

Concerning spatial aspects of sound reproduction, congruencies may be produced between identified sound sources in the original scene and sound sources in the reproduction scenario. A common practice in the production of sound installations consists in distributing loudspeakers in a physical space where they coincide with the sound sources of the (re-)created sound scene. Many sound installations by artists like Berhard Leitner and Robin Minard consist in sound environments where the listener can freely explore the reproduced sound scene by walking through the created space (Blume 2008; Minard 1993). Also the piece by Janet Cardiff, *The Forty Part Motet*, introduced at the beginning of this section, belongs to this kind of sound reproduction and interaction scenarios. In these scenarios, the loudspeakers spatially coincide with the sound sources that they reproduce at fixed positions in space. Usually each sound source is represented by a separated audio channel. To overcome the limitation of immobile sources, in many of Berhard Leitner's and Robin Minard's sound installations, multiple loudspeakers form trajectories that allow realizing moving sound sources by successively assigning the source signals to different loudspeakers along a given trajectory.

An additional congruency here concerns the acoustic properties of the loudspeakers that have to fit the acoustic properties of the sound sources that they reproduce in order to create a coherent interaction with the listener. For example, it is difficult to convincingly display the sound of a powerful machine with a single small loudspeaker.

In his sound installation *A World Beyond the Loudspeaker*, Edwin van der Heide extends the reproduction of single sound sources and trajectories to the reproduction of a sound field that the listener can interact with through the metaphor of a *sonic window*[96] (van der Heide 1998). Instead of reproducing each of the sound sources by an individual loudspeaker, the installation uses a matrix of 40 loudspeakers to reproduce an acoustic wave field captured by a matrix of 40 microphones. Following the principles of wave propagation, that are also the foundation of wave field synthesis (Berkhout et al. 1993), the congruency of the geometric setup of microphones and loudspeakers allows for reproducing the acoustic wave field captured by the microphones.

[96]The expression *sonic window* appears in an interview with Edwin van der Heide in (Gusev 2009).

2.2.4 Beyond Reproduction

In the examples above, the listeners' interactions are limited to the exploration of acoustic spaces composed of static sound sources, trajectories, and sound fields represented by loudspeakers. Examples of so-called *interactive* sound installations illustrate how more complex representations of actors and actions extend the listeners' engagement with and through sound.

Similar to the sound installations by Leitner and Minard (see *2.2.3*), the technology that has been developed in the framework of the *LISTEN* project, gives the possibility to create sound spaces that ca be explored by listeners walking through a gallery space (Eckel 2001). As Leitner's and Minard's installations, the *LISTEN* project is based on the idea of identifiable sound sources that appear at certain positions in the gallery space. But instead of coinciding with the positions of loudspeakers, the sound sources in *LISTEN* are rendered through a binaural spatialization system and can appear at virtually any position in the space around an individual listener. The tracking of the listeners' position and orientation allows not only for adapting the spatial rendering of the sound sources to the movements of each listener, but also for creating further congruencies between the reproduced sound and the listeners' actions. Beyond the spatial properties of sound sources, the *LISTEN* system selects and controls the playback of recorded sounds as a function of the listeners' trajectories in the gallery space. The action representations used by the system to create the necessary congruencies primarily consist in individual sound streams, the captured motion data,[97] and the rules controlling the playback of pre-recorded sound materials as a function of the listeners' motion.

Another example of an interactive sound installation based on advanced action representations, is the piece *Global String* by Atau Tanaka and Kasper Toeplitz (Tanaka and Bongers 2001). In this installation, visitors of gallery spaces at different geographic locations can commonly interact with a virtual string as if it was running through all of the participating galleries. In each gallery space, the string is embodied by a steel cable that diagonally crosses over the space stretched between the floor and the ceiling. The vibrations caused by the visitors action on the strings in the different gallery spaces are captured and transmitted via a an internet connection to the computer simulation of a vibrating string. The resulting sound is transmitted by loudspeaker systems in each gallery space.[98] It is interesting to notice how all elements in this interaction scenario contribute to the reenactment of the string the visitors interact with. The steal cables, the loudspeakers, and the video monitors, transmitting a visual representation of

[97] Considerable efforts in the framework of the *LISTEN* project have been dedicated to the research on head related transfer functions (HRTF) that allow for easily adapting the system as good as possible to any listener's audition (Warusfel and Eckel 2004).

[98] While before this example, the discussion for the concept of reenactment strictly focused on audio applications based on recorded sounds, the interaction scenario of *Global String* relies on sound synthesis by physical modeling. This allows for extending the investigation of the concept of reenactment to sound reproduction technologies that rely on other action representations than digitized sounds.

the visitor's actions from one gallery to the others, efficiently display the necessary congruencies to create the idea of a single string running through all galleries. These congruencies confirm the visitor's interactions with the string, as well as his or her engagement with the visitors of other galleries through sound and vision.

This perspective on reenactment provided by the idea of congruent actions, can also be applied to the example of virtual conducting already introduced above. This example can be compared to the examples of the music lover listening to the recording of an orchestra concert and the turntable performer scratching the recording of a drum beat. The example of the music lover focused on congruencies between the macro structure of the reproduced concert and the listener turning the records between the movements of the recorded piece. The scratching techniques in the example of turntable performance require the skillful alignment and synchronization of the performers gestures with the details of the recorded sound. In the light of these examples, the interactive audio system realizing the virtual orchestra through the interactive rendering of sound recordings appears as a sophisticated record player. Other than the turntable performer and the music lover who are themselves responsible for most of the congruencies between their actions and the recorded sound, the congruencies between the gestures of the conducting listener and the actions of the recorded orchestra have to be produced by the interactive audio system that renders the recorded sound materials. This may concern tempo variations, but also other expressive variations such as the dynamics and the balance of the different instrument groups. From the interacting listener's point of view, the interactive system in this example aims at creating congruencies between his or her conducting gestures and the produced sound. However, these congruencies are imagined as being additionally mediated through their interpretation by the musicians of the virtual orchestra following the listener's conducting. Therefore, the listener expects characteristic incongruencies of the produced sound with the conducting gestures. These incongruencies are consistent with the idea of interacting with independent players and instrument groups interpreting a common score.[99]

As already mentioned above, to realize a particular scenario, the involved interactive system has to formalize the listener's actions, as well as the actions of the virtual actors the listener engages with, as digital processes. Some of these actions are already represented by the recorded sound materials. These representations generally have to be completed by actors represented within the hard- and software of the interactive audio system. The examples of interactive applications given above show how these actors are composed of multiple processes based on different techniques and formalisms.

[99]While this particular reproduction scenario evokes virtual instrument groups, it could also envisage an interaction evoking individual instrumentalists or other virtual actors that may refer to known human or non-human actors. As mentioned above, the exercise of designing such interaction scenarios is a play with tacit knowledge involving the imagination of designers and users.

Many of the actors that technically compose a given system from the designer's point of view, may disappear or appear differently from the listener's point of view. For example, in the recording, transmission, and reproduction of a news show, the listener may primarily engage with the news speaker and the radio (i.e. switching it on and off, adjusting its volume and position, etc.). Other actors present in the recorded scene and reproduction scenario may disappear from the perception of the listener because they are simply out of the scope of the listener's perception or because they are strongly congruent with the perceived actors. In this sense, the actions of the new speaker's microphone, the radio transmission, and the radio loudspeaker, for example, are usually masked by the listener's perception of the news speaker's voice. However, they may suddenly appear as actors as soon as something would bump into the news speaker's microphone or disturb the radio transmission. The head tracking system used to generate localized virtual sound sources in the *LISTEN* project illustrates how particular actors of an interactive system are designed to disappear from the listener's point of view. To allow the engagement with a sound source appearing at a given position, any of the listener's movements is compensated for by a contrariwise movement of the virtual source. Hereby, the transfer functions used by the binaural rending have to be precisely congruent with the listener's physiology in order to disappear from the listener's perception. Another example of how perfectly congruent actions are used for compensating others that should not be displayed to the listener is *active noise canceling* (Kuo and Morgan 1999). In this case, these actions are sounds themselves.

Many approaches in industrial sound design and sonic interaction design (Rocchesso 2011) aim at enhancing the engagement into the interaction with objects and electronic devices by adding mechanical or electronic sonic actors. Due to their strong congruencies with other components of the same objects, these enhancements do not appear as independent actors from the user's point of view. A classical example of such applications is the sonic feedback generated when pushing a button of an electronic device such as a mobile phone or an automated teller machine. When providing the necessary congruencies, the user associates the produced sound to the buttons rather than an additional component of the device. Another often-cited example of sonic interaction design, mostly based on mechanical elements, is the enhancement of car sounds. These enhancements concern virtually all parts of a car, and especially the engine and the doors. Similar approaches, aiming at augmenting the evaluation of the overall quality of consumer products, can be found in the sound design of vacuum cleaners (Bodden and Igsleder 2002) and many other electrical supplies (Fog and Pedersen 1999).

The sound design of graphical user interfaces can also be seen from this perspective. While in the above examples sound design aims at enhancing existing sound, in the sound design of graphical user interfaces, sounds are generated to emphasize actions that do not produce sound themselves (Gaver 1986, 1989). Sound here is created in spatial and temporal as well as semantic congruency with graphical actors and hidden processes the user engages with.[100]

Even though they all contribute to the listener's engagement with particular actors emerging in the created interaction scenarios, the congruencies explored in these examples actually belong to different categories. While in the first example, the effect of congruencies between the elements in the chain of sound reproduction technologies has been interpreted as an effect of *masking*, the following two examples of binaural spatialization and noise canceling evoked the notion of *compensation*. The examples featuring different applications of industrial sound design can be described through the notion of *emphasis*. These three notions, *masking*, *compensation*, and *emphasis*, generally coexist within the designer's composition of *actors* the listener finally engages with. From this perspective, the work of the designer consists in creating adequate congruencies among the components of the designed system so that particular actors can appear from the listener's point of view. In the design of interactive audio applications, these actors generally include, in addition to the actors evoked by recorded sounds, a multitude of digital processes, physical devices, and objects manipulated by the listener. The congruencies of their actions with the listener's actions finally allow the listener to engage into particular interactions intended by the designer.

R 19 | The design of interaction scenarios consists in creating congruencies among the involved elements that allow for the emergence of actors the listener engages with.

The examples of interaction scenarios cited in this section illustrate the concept of interaction design departing from the idea of listening as an engagement into bodily interaction. Especially the example of turntable playing techniques illustrates how technology mediated listening may extend into musical performance. Insofar as the interaction scenarios of interactive audio applications support listening as an active bodily engagement with sound, all actions present in a given scenario contribute to the reenactment of sound. But other than dance or similar actions exteriorizing – or *enacting* – the process of listening, the interaction scenarios of interactive audio applications restage recorded sounds as the listener's active involvement in its reproduction and transformation.

In this sense, the notion of *reenactment* extends the notion of sound *reproduction* by integrating the listeners' involvement. It creates a perspective on sound reproduction technologies that reconciles aspects of listening with aspects of performance that have been strongly separated into two different categories by earlier sound reproduction technologies.

[100] Picking up on semiotic and cognitivist concepts, Gaver refers to these congruencies as *mappings* distinguishing *symbolic*, *metaphorical*, and *iconic* mappings (Gaver 1986).

R 20 | The concept of reenactment extends the notion of sound reproduction by integrating the listeners' interaction.

It reconciles aspects of listening and performance that have been separated into two different categories by earlier sound reproduction technologies.

Most of the examples cited above, rather than faithfully reproducing scenes that actually existed, are concerned with producing imaginary scenes featuring more or less identifiable actors and actions. Electroacoustic music pieces and sound installations, as well as movies and game sound design are examples of applications that generally aim at the production of fictitious and in many cases even completely abstract sound scenes. Especially, movie sound design and Foley art illustrate the blurred boundaries between faithfulness and fiction (Ament 2012). Taking a closer look to established practices in audio recording, reveals that these boundaries are even not clear for productions that pretend to reproduce the physical realities of traditional music performances. Weinzierl and Franke (2002), for example, have shown to what extent a commercialized audio recording of a Beethoven symphony is composed of – partially very short – sound snippets from different performances of the same piece, even though the recording creates the illusion of a coherent continuous performance.

The examples given in this section, as well as the applications presented in chapter 3, show how the blending of realistic details with abstract elements, as well as faithfully reproduced aspects with fictitious elements, is an important aspect in the design of interactive audio applications. This blending concerns not only the sound and action produced by the application, but also the way these actions are related to the listeners' interactions.

2.2.5 Epistemic Tools and Phenomenological Essays

Phenomenology and the consequent development of concepts like enaction and embodied cognition have contributed to the evolution not only of the understanding of cognition, but also the understanding of knowledge, and thus to an evolution of epistemology. As already mentioned above, the enlarged notion of knowledge emerging from this evolution integrates *enactive* (Bruner 1966), *tacit* (Polanyi 1966), *bodily* (Parviainen 2002), *embodied* (Johnson 1989), *personal practical* (Clandinin 1985), and *material* (Baird 2004) aspects (see 2.2.2). It includes different modes of operating with knowledge other than written or spoken language and formal systems such as those used in mathematics.[101]

[101] The idea of distinguishing different *aspects* of knowledge tries to avoid opposing different *kinds* of knowledge. Polanyi (1966) distinguishes *tacit* from *explicit* knowledge and Parviainen (2002) uses the term *articulated*

Arguing for the understanding of digital musical instruments and other interactive audio applications as *epistemic tools*, Magnusson (2009) comprehensively summarizes and discusses how different aspects of knowledge are manifested[102] in the relationships between designers, performers, and applications. The notions of *embodiment* and *incorporation* here mainly appear in Merleau-Ponty's sense as the result of the '*acquisition of habit as a rearrangement and renewal of the corporeal schema.*'[103] But it is also applies to artifacts in the sense that they may be seen as *embodying*, *incorporating*, and *encapsulating* knowledge. Magnusson cites in this context Hayles (2008) who, in her account of *cybernetic technologies* occurring in fiction and science, introduces the distinction between *incorporation* and *inscription*. In her analysis, she refers to Connerton (1989) who comprehensively explores incorporation and inscription *practices* in his investigation of '*How Societies Remember*'. Following this reasoning, we *incorporate* knowledge as well as we *inscribe* knowledge into representations such as writings and movies.

When applying this distinction to the design of interactive digital applications, it seems to inform to some extent the relationships between designers, users, and applications. In the light of *incorporation*, it becomes clear how the design and use of applications imply the designers' and users' incorporated knowledge. An often-cited instance of an incorporation practice is the performance of musical instruments in which instruments become extensions of their performers' body.[104] The notion of *inscription* here applies in the sense that the designers' encodes knowledge and cultural practices into the designed application. The application be seen as an artifact that formalizes, memorizes, and transmits such aspects of knowledge. In this sense, Magnusson (2009) argues that a musical instrument '*can be seen as a text, something we have to read in our use of it.*' Recorded sound materials and program code are obvious *inscriptions* of interactive audio applications. On the other side, designers and users primarily operate with

as opposed to *bodily* knowledge in her account on knowledge involved in dance. Nevertheless, while both distinctions – tacit/explicit and bodily/articulated – are understandable in the context of the discourse they support, they both seem actually refer to the distinction of complementary aspects of knowledge that is memorized and transmitted by written or spoken language and aspects that are *articulated* – often even *explicitly* – within other interactions. The underlying distinction of nonverbal from verbal aspects is also manifested by Polanyi's statement in the introduction of his book that '*we can know more than we can tell*' (Polanyi 1966). As an example of an explicit but nonverbal articulation of knowledge, one may imagine a dancer or musician teaching a particular movement or phrasing by articulating it in an exaggerated manner. Another example, that also implies cross-modal representation, could be the description of a particular musical phrasing through a manual gesture. Latour (1987) more clearly applies the notion of tacit knowledge to knowledge that is implied in the understanding of a text without being explicitly transmitted by it.

[102]The term *manifested* here stands for all aspects of operating with knowledge such as *acquisition*, *memorization*, *transmission*, *representation*, and *discussion*.

[103]In (Merleau-Ponty 1962) also cited in (Magnusson 2009).

[104]This kind of interaction is often referred to as *embodied interaction*. This term is avoided in this dissertation since it is composed of the transitive verb *to embody* and *interaction* suggesting interaction as a direct object of embodiment. The difficulty of employing this term clearly appears when trying to apply in constructs like '*the embodied interaction between the performer and the instrument*' or '*the performer has embodied the interaction with the instrument*' or '*the interaction between performer and instrument is embodied*'.

embodied knowledge in their interactions with the application. Finally, it is difficult to clearly distinguish inscription and incorporation in the design of such artifacts since they only appear together within a complex entanglement.

R 21 | In the design of and interaction with digital applications inscription and incorporation practices are entangled.

To better understand this entanglement without being distracted by metaphorical notions of inscription, it can be useful to have a closer look on how inscription and incorporation appear in the process of writing and reading. Connerton (1989) explicitly insists on their entanglement in the process of writing.

> *It is certainly true that writing, the most obvious example of inscription, has an irreducible bodily component. We tend to forget this; writing is a habitual exercise of intelligence and volition which normally escapes the notion of the person exercising it because of this familiarity with the method of procedure.*[105]

Regarding writing, we may intuitively separate the (*incorporated*) knowledge *of* using a pencil, typewriter, or word processing system from the (*inscribed*) knowledge *about*[106] facts and artifact, places, actions, and experiences that are described in the written text. However, the distinction of these two levels is less clear for the knowledge concerning *argumentation* or *storytelling* that are involved in writing.[107] Having a closer look at these aspects, it is difficult to decide whether, for example, the *style of writing* would belong to the technique *of* writing, or to what the text is *about*.[108] Moreover, studies in semiotics that investigate the relationship between symbols and meanings, show the difficulty to draw a clear line between words and understandings they evoke.[109] On the reader's side the knowledge inscribed in the text and the knowledge the reader employs to decrypt and understand it are also strongly entangled. While interpreting the text, the reader operates with incorporated aspects of knowledge in the way he or she *lives through* – and re-enacts – actions and situations described in the text. Hayles

[105] In (Connerton 1989).

[106] Grammatically, the proposition *of* indicates knowledge in a construct using genitive/possessive case, while the proposition *about* puts the known into the function of a direct object.

[107] One could also refer to this knowledge as writing *skills*. The relationship between knowledge and skills is clarified by Polanyi (1966) who considers skills as the exhibition of (tacit) knowledge.

[108] An excellent example of an essay in which writing style becomes the actual object is *Exercices de style* de (Queneau 1963).

[109] See for example (Peirce and Moore 1998) on *sign processes*, (Wittgenstein 1965) on the relationship between writing and thinking, and (Dewey and Bentley 1949) on language and *knowing*.

(2008) finds a '*double entanglement of the textual corpus and the physical body*' referring to Wills (1995) who regards writings as bodily extensions comparing it to physical protheses.[110]

Connerton (1989) discusses cinema as a practice of inscribing, insisting on how the reception of a movie by the spectator relies on incorporation practices. The filmed scenes are *inscribed* into the movie, but the spectators understand the displayed action through their *incorporated* knowledge. Apart from written language and movies, the idea of inscription fits various technologies discussed above such audio recordings, as well as the piano rolls of the player piano and the camshafts of early musical automata. The distinction between aspects of inscription and incorporation becomes even more difficult when considering digital technologies that, instead of reproducing action by the playback of a recording, generate action through different models of temporal processes. Evident examples of such technologies are the models generating the behavior of characters featured in recent movies and computer games, as well as the generators of sound and movement employed in interactive digital audio applications (see *2.1*). While the act of creating these actors and actions by programming undeniably belongs to the category of inscription practices, the created digital actors incorporate knowledge of acoustics, motion, music, and perception.

Especially machine learning techniques, applied in multimedia and robotics for modeling aspects of movement and behavior on different levels of spatiality and temporality, illustrate the complete entanglement of inscription and incorporation practices in recent technologies. In the design of digital audio applications, machine learning techniques may be applied in form of generative statistical models for the analysis and resynthesis of sound textures (Kersten and Purwins 2010, 2012) and musical structures (Assayag and Dubnov 2004), as well as techniques of computer vision (Szeliski 2010), audio and motion analysis (Müller 2007), and gesture recognition (Bevilacqua et al. 2009). While these techniques allow for inscribing and/or incorporating knowledge by machine learning, also explicit rule systems such as the *KTH Rule System for Musical Performance* (Friberg et al. 2006) have shown their efficiency in modeling complex aspects of musical interpretation on the boundaries of inscription and incorporation.

Even if it is difficult to maintain the distinction of *inscription* and *incorporation* as different *practices* it is clear that artifacts like books, tools, musical instruments, and digital applications allow for operating with various aspects of knowledge within the practices they support. From the point of view of interaction design, the question of whether these knowledges are *inscribed* or *incorporated* is much less important than the question of who and what is involved in these practices and what are the relationships and interactions[111] between these actors.

[110] Another instance of this double entanglement is the French translation of *The embodied Mind* by Varela et al. (1992) that can be translated as '*The Corporal Inscription of the Mind*' (*L'inscription corporelle de l'esprit*).

[111] In this investigation, the distinction between *relationship* and *interaction* is coherently used to distinguish the *interactions* within a particular interaction scenario that is the object of design from all other *relationships* such as, that between the designer and the user.

Magnusson (2009) copes with the entangled distinction of inscription and incorporation in his investigation on digital musical instruments and audio applications by proposing the notion of *epistemic tools*. His investigation focuses on the design of digital musical instruments for expert-users (i.e. musicians) and shows how the design and performance of instruments unfolds an open space of reflection *on* and *through*, music, science, and technology that includes virtually all aspects of knowledge. The construct *epistemological tools* clearly displays the abolition of the distinction between inscription and incorporation. Inverting Magnusson's construct, composed of an adjective referring to knowledge which applies to incorporation practices, into an adjective referring to experience applying to inscription practices, produces *phenomenological essays*. While the idea of *epistemic tools* focuses on the manifestation of knowledges through the design and use of tools, the idea of *phenomenological essays* insists on the authoring of experiences. The notion of *tools*, suggesting expertise and function, hereby is replaced by *essay*, referring to exploration and reflection. From this point of view, interactive digital applications become nonverbal digital essays that invite their users into experiences. Instead of words, these essays are composed of nonverbal multimodal interactions.

R 22a | Interactive digital applications invite their users into exploration and reflection through nonverbal multimodal interactions.

When accepting the idea that the term *multimodal interactions* could refer to any of our senses,[112] the image of an author who invites an audience into an experience does not only describe interactive digital applications, but also many other productions of art and entertainment such as dance, music, cinema, and many forms of plastic art, as well as cooking, city walks, and roller coasters. Even if interaction design and each of these practices are based on their own codes and representations, their similarity from this very general point of view is not surprising. The notion of *essay* additionally implies, that the readers in a certain sense follow the author's narration and reflection, whether the subject of the essay is clearly identified or not. Moreover, it highlights that the user discovers whatever the essay evokes within an entanglement of the relationship to the author and the essay itself – as an artifact.

However, multiple aspects distinguish the design of so-called *interactive* digital applications from other products of authorship calling for the attention of an audience. First, interaction design not only implies that users interact, but it is explicitly concerned with the question of *how* users' interact. Second, the users' interactions are not only multimodal in the sense of audiovisual, but additionally involve the users' proprioception.[113] Similar to the way certain forms of

[112]In *2.2.2*, listening has been investigated as interaction. Watching obviously should be considered as interaction in the same manner.

[113]Proprioception refers to the perception of our own movement and orientation, through our inner ear, as well as our own muscle action (Sherrington 1907).

art and entertainment focus on particular senses (e.g. audition, vision, taste), interactive digital applications particularly address their users' proprioception.[114] Last but not least, the design of interactive applications is concerned with the users' experience of themselves as actors.[115]

R 22b | Interactive digital applications explore their users' experience of themselves as actors.

The users' experience of themselves as actors is essentially based on their observation of the consequences of their actions through the congruencies between their proprioception and other perceptions (see for ex. Knoblich and Flach 2003). In this sense, the above definition completes the idea of interaction design as consisting in creating congruencies (see *2.2.4*). The idea that the designer not only authors experiences the user engages with, but scenarios in which the user is staged as the active explorer of an environment provided by the application, finally complies with the notion of *phenomenological essays*. The world to be explored, the designer creates for the user – its interaction scenario –, assembles elements of the world the user already knows (i.e. objects, behaviors, and other *phenomena*) with elements of fiction, comment, and cross references that emerge in the users' interactions with the application. This *essay-world* may fit into a mobile phone or a gallery space, or it spans over a whole city or the globe (e.g. networked games). The distinction – and latent dichotomy – between inscription and incorporation vanishes in this account. What remains are authors connected to their audiences through practices and artifacts – manufactured, written, and programmed.

The switch from *epistemic tools* to *phenomenological essays* obviously has not eliminated the idea that interactive digital applications are concerned with knowledge. In the contrary, as *essays*, these applications remain artifacts that foster the operation with knowledge and experience.[116] Dewey and Bentley (1949) have investigated knowledge in relationship to our interactions with the world. Regarding knowledge from within the process of inquiry rather than a result outside or beyond this process, they propose the notion of *knowing* (and its plural *knowings*). Instead of an object of memorization and transmission, *knowing* is defined by an interaction – actually a *trans-action* – between the *knower* and the *known*. The procedural and interactional aspects of knowledge cumulate in the notion of the *knowing-known* that is part of the precise terminology Dewey and Bentley define for their epistemological inquiries. Apart from language-based forms of *knowing-knowns* – here called *namings-nameds* – these inquiries

[114]At this point of the argumentation, aiming at defining the specificity of interactive applications, *roller coasters* are still included into the definition. While already the first point did not support them very much, they will be definitely eliminated by the next.

[115]In psychology and neurology, the experience of ourselves as actors is investigated as *agency* (see for ex. Repp and Knoblich 2007).

[116]Even if they regularly appear separately in this argumentation, the distinction between knowledge and experience actually becomes obsolete within the phenomenological approach (Merleau-Ponty 1945). However, it is convenient to distinguish the act of *experiencing* from *knowledge* in more abstract terms.

also explicitly refer to '*other forms [that] include not only the full range of the perceptive-manipulative (Signal), but also those of non-naming linguistic processes such as mathematics (Symbol).*'[117] Independently from the referential system used in a particular practice, *knowings* are defined within the relationship of the *knower* and the *known* that itself evolves within to a larger context of relationships (see *2.1.6*).

Dewey and Bentley reserve the notion of *inter-action* to the inquiry of actors[118] that are defined independently of each other and of their relationship as in Newtonian physics. From this concept of *inter-action* they distinguish *trans-action*. The notion of *trans-action* supports the understanding of interactions and relationships '*without attribution of the aspects and phases of action to independent self-actors, or to independently inter-acting elements or relations.*'[119] Whatever interacts, is first of all defined *within* and, moreover, *by* its interaction.

Informed by the concept of *trans-action*, the design of interactive digital applications, rather than creating artifacts, is a matter of creating interactions. Moreover, the designers and users, rather than as self-acting[120] creators and audiences, are recognized as actors in a context that includes the application and its users, but also further practices and *fields*.[121] The notion of *field* hereby applies to fields such as art, science, and technology, but also refers more generally to the horizon in which interactions are considered, such as the interaction scenario of an application. While the design undoubtedly produces an artifact, the *transactional* account of design emphasizes the interactions and practices that it supports as well as their relationships to other existing interactions and practices.

The concept of reenactment, as it is investigated in this dissertation, establishes a transactional view on the design of interactive digital applications that includes at least three horizons. First of all, it turns towards the scene of interaction between the users and the application. Secondly, referring to *enaction*, it acknowledges the intersubjective experience and knowledge shared by the designers and users of the application. Finally, by its *re*-prefix, *reenactment* refers, beyond the personal and intersubjective experience of interacting, to further elements that are known and can be experienced independently from these interactions. While the in-

[117] In Dewey and Bentley (1949).

[118] Apart from the term *actors*, Dewey and Bentley (1949) arbitrarily use various other terms such as *elements*, *entities*, *essences*, *events*, *inter-acting constituents*, *objects*, *organism*, *things*, and *realities* – also '*little reals*' – stating at some point that they '*employ no basic differentiation of subject vs. object, any more than of soul vs. body, of mind vs. matter, or of self vs. not-self*'.

[119] In (Dewey and Bentley 1949).

[120] Dewey and Bentley (1949) associate the notion of *self-actor* to archaic approaches to inquiry such as found with Aristotle.

[121] Dewey and Bentley (1949) explicitly acknowledge the potential of the term *field* in transactional inquiry that '*[o]n physical analogies … should have important application*'. Nevertheless, they remark that '*[t]he physicist's uses, however, are still undergoing reconstructions, and the definite correspondence needed for behavioral application can not be established*' and that '*[t]oo many current projects for the use of the word have been parasitic*' concluding with the claim that '*transactional studies of behaviors on their own account are needed to establish behavioral field in its own right*'.

vestigation of the concept of reenactment above has particularly emphasized the reference to phenomena, actions, events, and practices that are evoked by recorded audio materials, ultimately any representation to action and reference to action is concerned by this idea.

R 23 | The notion of reenactment implies a transactional account that embraces the interaction between the user and the application, the intersubjective relationships between users and designers, as well as references to the world outside the application.

What remains to be explored, are the hermeneutic and semiotic aspects of the reflection through multimodal nonverbal interaction. While a comprehensive investigation of these aspects largely exceeds the scope of this dissertation, at least some elements, found in the writings of Latour and Heidegger, are given in the following.[122]

Strongly influenced by Dewey's ideas, Latour proposes the concept of *actor-networks* (see for ex. Latour 2005) and, more recently, *monades*,[123], to overcome the idea of individual and independent actors that preexist their relationship to other actors. These concepts acknowledge that whatever is perceived as an actor, appears as such actually and only through its interactions with other actors (see *2.1.6*). In his investigation of '*the construction of facts and machines [as] a collective process,*'[124] Latour (1987) introduces the concept of *blackboxing*. His investigation of *blackboxing* examines in detail how knowledge is revealed and encapsulated in the process of developing and communicating about scientific knowledge and technologies. It is illustrated by examples from biology and computer science. The notion of *blackboxing* here stands for the process of how scientists and technologist encapsulate knowledge within the construction of artifacts and communications in a way that it becomes a tacit element in the functioning of a machine or a particular reasoning.

> *[W]hat they have done is visible in the machines we use, the textbooks we learn, the pills we take, the landscape we look at, the blinking satellites in the night sky above our head. How they did it, we don't know.*[125]

Applying the concept of blackboxing to political and scientific writings, Latour proposes the notion of *positive* and *negative modalities* to distinguish '*sentences that lead a statement away*

[122]Some further keys for understanding of the hermeneutics and semiotics of interactive digital applications are given in the next section on *metaphors* and *affordances*.

[123]Latour (2005) refers to the term *monade* in Gabriel Tarde's *Monadology and Sociology* that refers to the concept, originally introduced in Gottfried Wilhelm Leibniz' *Monadology*.

[124]In (Latour 1987). In his investigation of *blackboxing*, Latour clearly refuses any distinction between material and social interactions.

[125]From the introduction of (Latour 1987) entitled '*Opening Pandora's Black Box*'.

from its conditions of production' – closing the blackbox – from those that '*lead a statement in the other direction towards its conditions of production*' – opening the blackbox.

In the idea of Latour's blackbox, opening and closing through positive and negative modalities in the construction and sharing of knowledge, resonates Heidegger's account of the dynamics in our use of tools (Heidegger and Stambaugh 1996). Following Heidegger's account, our tools may appear *ready-to-hand* ('*zuhanden*'), transparently extending the capacities of our body, or break down to a *conspicuous*, *obtrusive*, and *obstinate* object that becomes *present-at-hand* ('*vorhanden*').[126] Heidegger illustratively describes how we may switch from one perspective to the other, varying our distance and attitude towards one and the same object.

> *The modes of conspicuousness, obtrusiveness, and obstinacy have the function of bringing to the force the character of objective presence in what is at hand. What is at hand is not thereby just observed and stared at as something objectively present . . . Useful things become "things" in the sense of what one would like to throw away. But in this tendency to throw things away, what is at hand is still shown as being at hand in its unyielding objective presence.*[127]

Latour's *blackboxing* and Heidegger's *breakdown* illustrate how different perspectives – or *modes* – may be intended or spontaneously occur in our interactions with ideas or artifacts and modulate our *reflective distance* (see also Voegelin 2000). Heidegger's breakdown strongly resonates in design concepts such as *parafunctionality* and *enstrangement* (Löwgren 2006). Change of perspective and mode is part of many interactions and it often appears as a major motivation for designing particular interactions. Such interactions include for example *learning* and *exploration*, *construction* and *deconstruction*, *reasoning* and *explanation*.

Interestingly, with *blackboxing*, Latour has chosen a visual metaphor that particularly informs the design of interactive audio applications. Audition and proprioception, are modes of perception that are used for the exploration of objects and environments especially in the absence of vision, for example when shaking a closed box to discover its content. As mentioned above, these senses are particularly addressed by interactive audio applications. Moreover, the blackbox metaphor is an instance of the more general metaphor of *seeing* for *knowledge*.[128]

[126]Exploring the notion of *embodiment interaction* in the context of human-computer interaction, Dourish (2004) cites Heidegger's distinction of *ready-to-hand* and *present-at-hand*. Armstrong (2006) has also applied this distinction in his examination of the design of digital musical instruments following an enactive approach.

[127]From *Time and Being* (Heidegger and Stambaugh 1996). Translated from the German original (Heidegger 1967): '*Die Modi der Auffälligkeit, Aufdringlichkeit und Aufsässigkeit haben die Funktion, am Zuhandenen den Charakter der Vorhandenheit zum Vorschein zu bringen. Dabei wird aber das Zuhandene noch nicht lediglich als Vorhandenes betrachtet und begafft . . . Das Zeug wird zu »Zeug« im Sinne dessen, was man abstoßen möchte; in solcher Abstoßtendenz aber zeigt sich das Zuhandene als immer noch Zuhandenes in seiner unentwegten Vorhandenheit.*'

[128]Seeing and hearing have been comprehensively discussed as metaphors for knowledge by many authors (see for example (Celermajer 2006)).

In this sense, the interaction with the blackbox is a double metaphor for the exploration of the boundaries between the tacit and the explicit as well as between the known and the unknown.

R 24 | Interactive audio applications explore the boundaries between the tacit and the explicit as well as of the known and the unknown.

2.2.6 The Interplay of Metaphor and Affordance

Similarly to the distinction of inscription and incorporation practices discussed above, Ihde (1990) identifies *hermeneutic* and *embodiment* relationships in our interactions with technology. Even if the underlying distinction of inscription and incorporation tends to generate more confusion than clarification, as perspectives on relationships of users with technology, the notions of *hermeneutic*s and *embodiment* inform the design of applications and interactions. From the perspective of *hermeneutics*, technologies are media that convey knowledge and meanings which are interpreted by their users. From the perspective of *embodiment*, technologies appear as extensions of their users' body that support their interactions with the environment. The metaphorical character of Ihde's relationships is confirmed by two further notions he proposes. In *alterity* relationships, technologies are accepted as an equal other, and in *background* relationships they seamlessly integrate into in their users' environment (Ihde 1990).

Applying these metaphors to the relationships users establish with an interactive digital applications, leads to the idea that applications may be read like books, played like instruments, dialoged with like equal others, or seamlessly integrate into their users' environment like elements of nature or architecture. These fundamental interaction metaphors describe the users' relationships in terms of interactions with other artifacts and environments.

R 25 | Interactive digital applications may be read like books, played like instruments, dialoged with like equal others, or seamlessly integrate into their users' environment.

The concepts and descriptions of the interactive digital applications above also make extensive use of metaphors. On one hand, these metaphors apply to the designed artifacts, such as van Edwin van der Heide's *sonic window* and Atau Tanaka's *global string*. On the other hand, they apply to the involved interactions, such as *conducting*, *listening*, and *playing*. Moreover, metaphors can apply to the role of the designer who may appear for example in the role of a *luthier*, an *architect*, or a *creator* with traits of Rabi Loew and Frankenstein.[129]

[129] Tomayko-Peters (2006), for example, uses the Frankenstein metaphor for his project *Maestro Frankenstein* and calls the application he designed *monster* – probably not without irony.

Burke (1969) has defined metaphor in terms of *perspective*[130] on '*whatever can be thought of as distinct*'.[131]

> *Metaphor is a device for seeing something in terms of something else. It brings out the thisness of a that, or the thatness of a this. If we employ the word "character" as a general term for whatever can be thought of as distinct (any thing, pattern, situation, structure, nature, person, object, act, role, process, event, etc.,) then we could say that metaphor tells us something about one character as considered from the point of view of another character. And to consider A from the point of view of B is, of course, to use B as a perspective upon A.*[132]

Even if Burke's writings exclusively refer to the use of metaphor in language – and particularly literature and poetry – his definition does not explicitly evoke language and easily can be applied to other interactions without any modification. Henle (1981) evokes Peirce's distinction of symbolic and iconic modes of signification noticing '*that there is clearly an iconic element in metaphor*' which consists in its relying on *analogy*. Inspired by Peirce's ideas on iconic signification he proposes to distinguish *qualitative similarities* and *structural similarities* between the literal signification of a metaphor and that what it applies to. He suggests that there may be other kinds of similarities, but requires that in metaphor '*similarity must be noticed and used as a means of signifying.*'[133]

Black (1954) defines metaphor as *interaction*. He introduces this idea through a citation from *The Philosophy of Rhetoric* by Campbell from 1750 proposing that '*when we use a metaphor we have two thoughts of different things active together and supported by a single word, or phrase, whose meaning is a resultant of their interaction.*'[134] According to Black, by creating a metaphor, such as *man is a wolf*, we actually connect two '*systems of things*'. These systems consist is a *principal* subject (i.e. man) and a *subsidiary* one (i.e. wolf), along with their characteristics and associated implications. The connection between the two systems is bidirectional. While the wolf-metaphor '*organizes our view of man*', it also '*makes the wolf seem more human than he otherwise would.*'[135]

Lakoff and Johnson (1980) claim that '*communication is based on the same conceptual system that we use in thinking and acting, language is an important source of evidence for what*

[130]Henle (1981) remarks that the notion of *perspective* that Burke proposes as a substitution for *metaphor* is itself a metaphor.

[131]Many investigations of metaphor also include other kinds of expressions using figurative language such as *metonymy* and *synecdoche*. Burke (1969) calls metaphor, metonymy, synecdoche, and irony the '*master tropes*'.

[132]In Burke (1969). In this definition also resonates Aristotle's often-cited definition of metaphor as '*the application of a word that does not belong: either from the genus to the species, or from the species to the genus, or from the species to the species, or according to what is analogous*' (cited in Müller 2009).

[133]In Henle (1981).

[134]Cited in (Black 1954).

[135]In (Black 1954).

that system is like'. In their account, metaphors in language are verbal expressions of profound concepts that structure and govern our thought, perception, and relationships – otherwise said, our cognition.

> *Our ordinary conceptual system, in terms of which we both think and act, is fundamentally metaphorical in nature. ... [T]he way we think, what we experience, and what we do every day is very much a matter of metaphor.*[136]

The idea that metaphors ultimately refer to experiences perfectly corresponds to the definition of reenactments as actions that refer to other actions (see *2.2.5*). From this perspective, *reenactments* are based on metaphorical relationships between actions or, more precisely, on metaphorical relationships between experiences of action.

R 26 | Reenactments are based on metaphorical relationships between experiences of action.

While the investigation of metaphors conducted by Burke (1969) mainly relies on examples from literature and poetry, Lakoff and Johnson have shown that also everyday language is largely based on metaphorical concepts. Apart from language, the notion of metaphor has been investigated in many domains including gesture (Cienki and Müller 2008; McNeill 2008), music (Zbikowski 2008), and design (Blackwell 2006). These investigations rely on practically identical definitions of metaphor and confirm the idea that metaphors are an essential element in the understanding of our interactions throughout different practices and technologies.

The use of metaphors has been comprehensively explored is the design of graphical computer interfaces. Metaphors in graphical user interfaces are often restricted to graphical objects such as the *desktop*, *windows*, *menus*, *buttons*, and various icons the user interacts with through input devices such as the mouse and touch screens. Gaver (1986; 1989) has extended the idea of icons in graphical user interfaces to *auditory icons* that convey information about sources of data and ongoing process through sound. More recently, a set of gestures has been introduced in the interaction with graphical interfaces through touch screens. These gestures instantiate metaphors which refer to the interaction with solid objects such as *tapping*, *dragging*, *flicking*, *swiping*, *pinching*, and *spreading* (see for ex. Gruman and Hattersley 2009). However, the use of metaphors for the design of graphical user interface has been associated with – and strongly criticized as – the mapping of objects and functions to images and animations that represent everyday environments such as office spaces. Cooper et al. (2012), for example, denounce metaphor-based design as '*masquerading [the interface] as an office filled with desks, file cabinets, telephones, and address books, or as a pad of paper or a street of buildings*'.

[136] In (Lakoff and Johnson 1980).

As a better alternative to *metaphoric interfaces* he proposes *idiomatic interfaces* arguing that '*[w]indows, title bars, close boxes, screen-splitters, hyperlinks, and drop-downs are things we learn idiomatically rather than intuit metaphorically*'.

In their critical discussion of the use of metaphor in graphical interfaces, Cooper et al. (2012) explicitly refer to the distinction of *metaphor* and *idiom*. In language, idioms often have been referred to as *dead metaphors*. Black (1979) even claims that a dead metaphor, such as it occurs in an idiom like '*falling in love*', '*is not a metaphor at all, but merely an expression that no longer has a pregnant metaphorical use*'. The commonly accepted idea – especially in linguistics – that metaphors could be *dead* and *alive* has been challenged by many authors in the past few decades. Lakoff (1987) calls the distinction of metaphors in *dead* and *alive* '*a holdover from a traditional folk theory of language that has turned out not to be workable*'. Lakoff and Turner (1989) see in the distinction of metaphors in *dead* and *alive* a misunderstanding that '*derives from a basic confusion*' that consists in assuming '*that those things in our cognition that are most alive and most active are those that are conscious*'. In the contrary, they claim that '*those [metaphors] that are most alive and most deeply entrenched, efficient, and powerful are those that are so automatic as to be unconscious and effortless.*'[137] Müller (2009) critically discusses the *dead-alive* distinction and gives a comprehensive overview over studies on metaphor in different domains, including language, image, and gesture. Instead of a static dead-alive dichotomy, she proposes to consider a *degree of activation* of *metaphoricity*. This degree is '*a dynamic property which critically depends on cognitive activation*' in a speaker or listener. It depends on multiple factors of the speaker's and listener's relationship to an expression or act. A particular metaphor can be understood, or not, depending on personal knowledge and experience, cultural background and context. We may be particularly receptive to the metaphorical dimension of a given expression or action at one occurrence, while ignoring it at others.[138]

The dichotomy of *dead* and *alive* metaphors ultimately disappears when, instead of thinking of metaphor as referring to *things* such as *wolf* or *desktop*, one focuses on *experience*. The metaphor of '*falling in love*', for example, directly refers to the experience of *falling*. From this perspective, even *wolf* and *desktop* can be understood as the *experience* of a wolf or a desktop. Other metaphors may refer to spatial relationships such as '*I am down*' or '*I am out of my mind*'. For Lakoff and Johnson (1980) '*such metaphorical orientations are not arbitrary*' since '*they have a basis in our physical and cultural experience*'. Consequently, their definition

[137] Cited in Müller (2009).

[138] Lakoff (1987) explains the metaphoric dimension of the word *pedigree* coming from the Old French *pied de grue* (*foot of a crane*), referring to the graphical shape of former family-tree diagrams having the shape of the foot of a crane. Even if Lakoff mentions the word as an example of a metaphor that could be reasonably called dead, nothing but this explanation may completely change the reader's relationship to metaphoricity of the word.

of metaphor does not refer to *words* or *things*, but to *experience*. According to this definition metaphor '*permits an understanding of one kind of experience in terms of another.*'[139]

From the perspective of experience, it becomes obvious that the *windowness* of an desktop interface element actually consist less in particular attributes of the displayed object that would remind of the wall opening of a house, but in the fact that we *open* and *close* it, and that it permits us to *view* the content of a file within the boundaries of its frame. Respectively, a menu permits us to *choose* and a button to *activate* a particular action. This notion of metaphoricity applying to action is evident in the case of the gestures already mentioned above that are commonly used in the interaction with multi-touch interfaced such as *pinch* and *swipe*. It is important to notice that the metaphoricity of these gestures does not only consist in the fact that their name and verbal descriptions refer to similar actions applied to physical objects. In addition to this linguistic aspect – and congruently – their metaphoricity lies in the actual action of pinching and swiping on the surface of our mobile devices that we understand in terms of our previously acquired experiences of actions that are part of our everyday interactions with our physical environment. The metaphoricity of these actions becomes obvious in the difference between the experience of interacting with physical objects and that of interacting with the graphical objects of the user interface. While designers are particularly conscious of this difference due to the effort they put into generating the graphical objects' behaviors, for the user the metaphoricity of these interactions may loose its evidence in the everyday interaction with touch screen interfaces.[140] From this perceptive, any acquired experience, concrete actions or categories of actions, actually lived or imagined, may be metaphorically activated in the understanding of another.

In this sense, the *sonic window* and *global string* metaphors in Tanaka's and van der Heide's artistic projects are not only linguistic metaphors used in the titles of these works or in other verbal descriptions such as in program notes, publications, and conversations. Beyond language, they are verbal translations of the metaphorical concepts that constitute these works. In Tanaka's work the technical setup of the installation clearly suggests the metaphor of a string running through multiple gallery spaces separated by hundreds of kilometers. The concept of *A World Beyond the Loudspeaker* is not as directly related to the metaphor of a window that van der Heide evoked in an interview about the installation. However, a strongly metaphorical character remains in the interaction that is afforded by the installation. This interaction consists in listening – to the world – *through* the exhibited loudspeaker array and evokes the boundaries between inside and outside, close and distant. This metaphorical character exists independently

[139]In (Lakoff and Johnson 1980).

[140]Obviously, the interactions with graphical objects via touch screens might become metaphors for other interactions with physical objects. Nothing prevents us from activating the pinch experiences we made in with our mobile devices when pinching others. Blackwell (2006) reports that in cognitive psychology research closely linked to computer science in the1960's and 70's, computing had become a '*metaphor for mental processes*' that convincingly replaced '*earlier philosophies of mind based on metaphorical clockwork or hydraulics*'.

of its description by the window metaphor. It lies in what is evoked by the actual interaction of moving in front of the loudspeaker array and listening to the emitted sound.

Similarly, the metaphorical character of the *Global String* installation is not limited to its setup. The interactions between the visitors and the shared string are themselves a metaphor for shared interactions connecting remote spaces over long distances that are characteristic for many Internet applications. It is important to insist that this metaphor is not only a matter of verbally describing the installation, but that it is actually instantiated by the visitor's interactions. The metaphorical concept of the installation equally concerns the design, the visitors' interactions, and any account of the installation including that of the designer and the visitors, but also that of anybody who has not actually interacted with the installation.

R 27 | The metaphorical concepts underlying an interactive digital application equally concern the application's design, the user's interactions, and any description of the application.

This notion of metaphor applied to interactions is closely related to the notion of *affordance*. The concept of *affordance* has been originally introduced by Gibson (1977) and appears in his ecological approach to visual perception (Gibson 1979) before being adopted by technologists and designers (Gaver 1991; Norman 1999). Gibson defines *affordances* as the possibilities or opportunities of action and behavior offered by an environment. He argues that '*an affordance is neither an objective property nor a subjective property ... an affordance cuts across the dichotomy of subjective-objective and helps us to understand its inadequacy*' adding that it is '*equally a fact of the environment and a fact of behavior.*'[141] For Gaver (1991) affordances are defined by the interaction of an artifact with the human motor system.

It is interesting to notice the similarity between Black's – and Campell's – idea of metaphor and Gibson's definition of affordance, as well as Gaver's actualization of it. Beyond the formal resemblances of the transactional accounts of their authors, the concepts *metaphor* and *affordance* overlap significantly. Both concepts can be depicted as a transaction of knowledge between two distinct domains. While in the case of affordance these two domains refer to an subject and its environment, in the case of metaphor they refer – in Black's words – to two '*system[s] of associated implications.*'[142] The similarity between *metaphor* and *affordance* becomes even more obvious, when applying the notion of metaphor to interactions. In this case, both the operation with metaphor as well as the operation with affordances consist in the creating, perceiving, and understanding interactions in reference to previously acquired experiences

[141] In (Gibson 1979).
[142] In (Black 1954).

of interaction. However, a fundamental difference between both concepts persists in the perspective they create on a particular interaction. While *metaphor* primarily describes to an act of imagination (i.e. the *understanding in terms of*), *affordance* is defined by a pragmatic perspective concerning possibilities of interaction. While the former may be understood as the question ‘*what does this mean?*’, the latter translates to ‘*what I can do with this?*’.

The notions of metaphor and affordance become synonymous in the reasoning on multi-modal interactions, such as the interactions with a turntable and a musical instrument. This idea is particularly manifest in the declaration of the DJ and turntablist Bob Swift that already has been quoted above in *2.1*:

> *Don’t look at the turntable as just this mechanism that you play records on. Apply yourself to it as if it were an instrument, and you can express yourself through the turntable as if it were an instrument.*[143]

Swift’s incitation is more than a play with words. He proposes to understand the experience of interacting with a turntable *in terms of* playing a musical instrument. While the construct ‘*as if it were an instrument*’ explicitly indicates the presence of metaphor, Swift also refers to the affordances of a musical instrument such as the possibility to *apply* and *express* oneself, as opposed to the turntable’s original affordance of *playing records*. By encouraging performers to look at the turntable ‘*as if it were*’ an instrument, he suggest to play the turntable *as* a musical instrument. However, the musical instrument metaphor for the turntable ceases to exist as such in the act of performing music with it, in which it actually *becomes* a musical instrument.[144] When leaving the explicit evocation of the musical instrument metaphor beside, the act of performing music with a turntable implies about the same metaphoricity as that of performing music with any other instrument. If, nevertheless, the act of performing music with a turntable appears more easily as a metaphor than the act of hammering – even with an object that has not been conceived for hammering – this probably does not lie in the nature of the turntable or the hammer, but in the metaphorical character of making music more generally.[145]

As many other authors (see for ex. Cook 1998; Clarke 2005), Zbikowski (2008) has examined occurrences of metaphor in music composition and analysis, as well as in the experience

[143] Quoted in (Reighley 2000).

[144] This example also perfectly illustrates the difficulty revealed by Müller (2009) to define metaphoricity as a static attribute of an object or action instead of considering it within the process of understanding and creating action and interaction.

[145] *The Hammer Song* by Pete Seeger and Lee Hays seems to suggest that also the concrete action of hammering actually could be understood metaphorically (i.e. referring to the civil rights movement). Other than metaphorical expressions like ‘*hammering argumentation*’, the lyrics of the song evoke the action of hammering literally. Even if it occurs in a conjunctive construct (‘*if I had a hammer, I’d hammer...*’, the song clearly adds an idea of metaphoricity to the action of hammering. A similar metaphoricity could also be found in turntablism as part of the hip-hop movement that has created and promoted novel practices of music performance and composition together with other forms of artistic expression.

of music. Metaphor can be found in descriptions of music, as well as in the way '*the language of music*' (Cooke 1959) refers to extramusical symbols and actions in composition and performance. Eitan and Granot (2006) have empirically investigated how listeners associate changes in musical parameters with physical space and bodily motion. In their experiments, subjects are asked to associate simple and controlled musical stimuli with images of motion in space. Similarly, the experiments on gesture-sound relationships conducted by Godøy et al. cited in *2.2.2* (Godøy et al. 2005, 2006; Haga 2008) can be seen as explorations of metaphor. From this point of view, the gestures that the participants of these experiments perform in correspondence to recorded excepts express their understanding of sound and music in terms of gestures and movements.

More generally, Small (1998) summarizes the metaphorical character of music performance and listening in the introduction of his book *Musicking*:

> *The act of musicking establishes in the place where it is happening a set of relationships, and it is in those relationships that the meaning of the act lies. They are to be found not only between those organized sounds which are conventionally thought of as the stuff of musical meaning but also between the people who are taking part, in whatever capacity, in the performance; and they model, or stand as a metaphor for, ideal relationships between person and person, between individual and society, between humanity and the natural world and even perhaps the supernatural world.*[146]

In this explanation of the neologism of *musicking* that he introduced in his book, Small directly derives the meaning of music performance and listening to its metaphorical character. Music here appears as a play with relationships that are metaphors for other extramusical relationships. Said otherwise, music consists in interactions that are metaphors for interactions – *social*, *physical*, and *metaphysical*.

On the other side, music making permanently operates with affordances. Affordances that concern possibilities of producing and articulating sound, for example, are a fundamental concern in instrument making and instrument performance. But also compositional aspects of music such as rhythm, melody and harmony, as well as different practices of music making, could be understood in terms of their affordances. These affordances concern the exploration of different modes of musical discourse and expression, as well as social and cultural opportunities, such as integration – and exclusion –, entertainment, ritual, and learning. According to Small's view, metaphoricity is a fundamental intention in music making. Consequently, many of the implied affordances directly or indirectly concern the potential of creating metaphoricity.

[146] In Small (1998).

In interaction design, the vector of intentionality in the interplay of metaphor and affordance usually points in the opposite direction. An often-cited motivation for creating interaction metaphors in design is the idea of communicating the affordances of an artifact to its users. Blackwell (2006) describes metaphor as a '*visual communication channel via which the designer achieves the rapid transfer of an effective mental model into the user's head*'. In the design of interactive digital applications, this interplay of metaphors and affordances is a central concern regarding particular interactions as well as whole interaction scenarios. In the interaction with digital applications, the question of '*what I can do with this?*' and question of '*what does this mean?*' ultimately converge in the users' attitude of curious exploration.

R 28 | The design of interactive digital applications is characterized by the interplay between affordances and metaphors concerning multimodal interactions.

Blackwell (2006) observes how interaction metaphors in the design of graphical user interfaces define the user's role. A metaphorical office or typewriter describes the user '*as an office worker or typist*'[147] in the same way as the musical instrument metaphor implies the DJ' role as a musical instrument performer. Even if for many applications this role is much less clearly defined as for a musical instrument or an office desktop, this transaction of metaphor between the applications, the user's role, and actual and potential actions exists in all scenarios. The play with metaphorical connections between these elements and their transformation is part of many artistic projects. In this sense, the visitor performing Tanaka's string becomes a global communicator and the one listening to van der Heide's loudspeakers becomes an observer contemplating the world through a sonic window.

2.2.7 Playing Techniques, Strategies, and Scenarios

In his book *Homo Ludens*, Johan Huizinga (1949) comprehensively investigates *play* as fundamental element of human – and animal – life. In this investigation he also examines how the notion of *playing* applies to music. Summarizing the etymology of words which correspond to the English *play* in different languages, he remarks that neither in Greek and Latin nor in today's Italian and Spanish, the performance of musical instruments is referred to in terms of *playing*. However, he finds that '*the Arabic la'iba bears this sense in common with a number of European languages, namely the Germanic (and some of the Slavonic) which, as far back as their mediaeval phase, designate instrumental skill by the word "play"*.'[148] The fact that the

[147] In (Blackwell 2006).
[148] In (Huizinga 1949).

French *jeu* and *jouer* apply to musical instruments, Huizinga takes as an ‘*indication of Germanic influence.*’ In a particular insightful passage of his book, he supposes that the notion of *playing*, as it refers to the performance of a musical instrument, may have its origin in the idea that playing an instrument has strong similarities to the art of a juggler.

> *The fact that "Spielmann" in German ("Speelman" in Dutch) has taken on the connotation "musician" need not be directly connected with the playing of an instrument : "Spielmann" corresponds exactly to joculator, jongleur, the original wide meaning of which (a performing artist of any kind) was narrowed down on the one hand to the poetic singer and on the other to the musician, and finally to anybody who did tricks with knives or balls. ... Further, bearing in mind that the term "playing" is never applied to singing, and to music? making only in certain languages, it seems probable that the connecting link between play and instrumental skill is to be sought in the nimble and orderly movements of the fingers.*[149]

Playing here appears in the sense of *movement* and *technique* rather than through aspects of representation and ceremony on which Huizinga comprehensively insists in his inquiry of many other domains of play. Nevertheless, he acknowledges that there is no doubt that music is ‘*lying within the sphere of play*’ since it ‘*begins and ends within strict limits of time and place, is repeatable, consists essentially in order, rhythm, alternation, transports audience and performers alike out of "ordinary" life into a sphere of gladness and serenity, which makes even sad music a lofty pleasure*’. Fundamental characteristics which Huizinga associates with play ‘*as a special form of activity*’, ‘*as a "significant form"*’, and ‘*as a social function*’In Huizinga (1949)..

The dimensions of *play* that can be found in music, equally concern the design of interactive audio applications. As music performance, these applications imply the users’ physical involvement that in some cases may appear similar to the ‘*tricks with knives or balls*’ Huizinga evokes in his book. As in music, these interactions occur along with elements of *poetry*, *narration*, and *ceremony* which are closer to dimensions of play as they are manifested in theatre, for example. In this sense, the design of interactive audio applications generally implies competences that can be compared to the competences of a luthier together with competences of composition (Jordà 2005; Magnusson 2009). In the design of interactive audio applications, these two aspects occur as the design of *playing techniques* and *playing scenarios*. While scenario-based design is common within *user-centered* and experience-centered approaches (Carroll 1995; Buxton and Buxton 2007; Wright and McCarthy 2010), references to *playing technique* are quite rare even in the domain of digital musical instruments. Regarding many aspects, the notion *playing technique* is very close to *affordance*. However, while *affordance* focuses on

[149] In Huizinga (1949).

possibilities of interaction, *playing technique* introduces a notion of interaction *practices* that can be useful for the investigation of interaction design beyond sound and music.

Kvifte and Jensen (2007) have investigated the notion of playing techniques in the context of traditional and electronic music. For them '*playing technique is closely linked to the construction of an instrument and to its acoustic qualities; so closely, in fact, that it may often be difficult to speak of construction without implying or referring at the same time to matters of playing techniques.*'[150] What Kvifte and Jensen here call *acoustic qualities* obviously refers to possibilities of producing sound. At the same the time, the notion of playing techniques refers to specific actions that imply the players' capacities. In this sense, playing techniques are potential interactions between a performer and an instrument that allow for the articulation of sound or changes in sound. As Gibson's affordances, playing techniques can not be reduced neither to objective properties of the instrument and the human body nor to subjective action plans that would occur from the perspective of the performer without considering the possibilities of the instrument.

Playing techniques depend on the mechanical properties of the instrument, as well as the performer's physiology and skills. Moreover, they are determined by social and cultural factors, such as the musical practices they belong to. As objects of study, playing techniques are associated with the instruments to which they apply, the composers and performers who invented them, as well as to particular cultural contexts, such as an ethnic group, epoch, style, and repertoire. Musicologists and musicians have, for example, studied the playing techniques of the saxophone (Kientzy 1990) and of digital instruments (Malloch and Wanderley 2007), the violin playing techniques contributed by Nicolo Paganini (Deegan 1941), the playing techniques occurring in particular pieces of Helmut Lachenmann (Hermann and Maciej 2012) and in baroque performance practices (Cyr 1992), as well as the didgeridoo playing techniques practiced by the Yolngu people (Schellberg 1994) in Australia.

The repertoire of playing techniques associated with a given instrument defines in a certain sense its vocabulary.[151] Usually this vocabulary develops over long periods of time. The vocabulary of violin bowing techniques, for example, has evolved over several centuries starting from bowing techniques inherited from it ancestors, the medieval *rebec*, the *renaissance fiddle*, and the *lira da braccio*. The evolution of these techniques is closely related to the evolution of the violin bow. Over the centuries, new playing techniques have been contributed by composers and performers like Corelli, Kreutzer, and Paganini (Boyden 1965; Stowell 1990). This continuos invention of new ways to articulate movement and sound continues in contemporary music. so-called *extended techniques* can be found in the music of Helmut Lachenmann (Her-

[150] In (Kvifte and Jensen 2007).

[151] The term *vocabulary* here stands – metaphorically – for a body of more or less atomic actions that are used to articulate more complex action sequences. Referring to units of articulation in language, playing techniques, in fact, could be arbitrarily compared to syllables, words, and even expressions.

mann and Maciej 2012), as well as many contemporary composers and performers like Mari Kimura (Kimura 1995). Playing techniques are themselves cultural artifacts that are invented and transmitted from generation to generation. Each generation of musicians adds new playing techniques to the existing vocabulary.

Certain playing techniques have been anticipated by the instrument builders that have invented or modified a given instrument. Others emerge from inherent affordances of the instrument. At the end of the 19th century, for example, the archetier François Xavier Tourte collaborated with the violinist Giovanni Battista Viotti on modifications of the violin bow that resulted in the modern bow as it is still used today. Important playing techniques like the modern *staccato* stroke, *spiccato sautillé* directly emerged from this design (Dell'Olio 2009). Even though, extended bowing techniques such as *col legno tratto*, as it appears in Arnold Schoenberg's music, or *subharmonics* (Kimura 1995) have not been foreseen by Tourte and Viotti, they emerged much later without further modification of the instrument itself.

R 29 | Playing techniques define the vocabulary of an instrument and may evolve over large periods of time along with the instrument and musical practices.

Rasamimanana (2008) (see also Rasamimanana et al. 2006; Rasamimanana and Bevilacqua 2008; Rasamimanana et al. 2007, 2009) has studied the movement and sound of conventional violin bowing techniques. In these studies, performances of violin bowing techniques are recorded and analyzed in respect to particular sound and movement qualities. In a first study, Rasamimanana et al. (2006) have recorded the sound and motion of bow strokes performed by trained violinists. The bow strokes have been analyzed according to sound and motion features so that each bow stroke can be represented by a point within a multidimensional space. The dimensions of this space (i.e. the sound and movement qualities) have been chosen in a way that the strokes of particular bowing techniques, *détaché*, *martelé*, *spiccato*, cluster into distinguishable categories. In this representation, each bow stroke of one of the studied playing techniques appears within one of the represented clusters. The distribution of each cluster shows the variability of the sound and motion features for each technique. Bow strokes of ambiguous qualities, in respect to the defined categories (i.e. techniques), appear in-between these clusters.

This study how a particular playing technique is defined by characteristic sound and movement features. Each playing technique allows for certain variations of these features. The clusters representing techniques that allow for continuous transitions between them are merged in the feature space. *Spiccato* violin bow strokes, for example, may vary in their dynamics and duration and tempo. Strokes of stronger dynamics, however, are rather understood as *martelé* and prolonged strokes as *détaché*. Repeated *spiccato* bow strokes allow for a certain range of tempo variations before they become *sautillé* strokes for higher bowing frequencies.[152] While the *détaché*, *martelé*, and *spiccato* bow strokes may exist as atomic actions, other playing techniques like *sautillé*, as well as techniques like *bariolage* or *Kreutzer bowing*, consist in of more complex action sequences (Stowell 1990).

R 30 | Playing techniques are action sequences of characteristic sound and movement features. Each playing technique allows for a certain range of variation of its features and some techniques allow for continuous transitions from one to the other.

While the space of bow strokes studied by Rasamimanana spans over the dimensions of measured sound and movement features, similar spaces could be imagined based on any meaningful set of qualities. Any such space of qualities can be seen as a projection of a hypothetical space spanning over all possible actions that produce sound on a given instrument. This space could be called the space of *sonic affordances* of the given instrument.[153] In abstract terms, a space of *sonic affordances* would be determined by the capacities of the instrument and the capabilities of the player or, more generally, the capabilities of all past and future players of the instrument. Certain regions of this space, representing particular sound and movement qualities, could be excluded due to the constraints of the instrument and the human body, while others would receive particular attention since they correspond to existing playing techniques. The areas of the sonic affordance space that are accessible to a particular player depend on the player's capabilities. The evolution of these areas over larger periods would show the evolution of the player's mastership and virtuosity. Some regions of the space may remain unvisited for aesthetic reasons (e.g. the produced sound is not desirable), others due to ergonomic choices, for example when the same sound can be achieved by another movement.

[152] It is interesting to notice the metaphorical potential of playing techniques that is often reflected by their names. Magnusson (2009) has examined the use of action metaphors in the design of digital musical instruments. His *taxonomy of interaction* lists a wide range of different actions together with their domains of activity, including acoustic instruments (e.g. striking, shaking, and bowing), graphical user interfaces (e.g. moving, selecting, copying), computer games (navigating, shooting, jumping), and machinery (e.g. pressing buttons, moving sliders). In fact, this list could be equally understood as a taxonomy of interaction metaphors or of affordances.

[153] Tanaka et al. (2011) distinguish *physical* and *sonic* affordance. They combine *physical* and *sonic* affordance with *cultural affordance* to define *musical affordance*. While *physical* affordance would correspond to all possible (physical) actions a particular instrument could provide – including hitting a violin on someone's head –, *sonic affordance* is defined by the possibilities to actually make sound. Franinović and Serafin (2013) authors define *sonic affordance* as a sonic indicator of action. However, this definition confuses the concept of affordance as potential action with its perception and recognition. Gibson (1979) explicitly makes this distinction when insisting that affordances exist independently of their perception and recognition by a subject.

In another study, Rasamimanana et al. (2007) investigate the movement strategies players employ to realize a constant evolution of sound qualities of up and down *détaché* bow strokes with increasing tempo (i.e. accelerando). The study shows that in the analyzed performances, the players radically change the movement strategy when reaching a certain frequency. This circumstance may be explained by the fact that the players' movements take into account the resonance frequencies of the bow in conjunction with their hand, arm, and shoulder. The study illustrates how the realization of particular playing techniques takes advantage, not just of the properties of the instrument, but also of the players' physiology. Playing techniques here appear as the interplay between the player, the instrument, and sound. More precisely, playing techniques are negotiated between the capacities of the instrument to produce sound, the performer's motor control capacities, and the desired articulations of sound.

R 31 | Playing techniques are negotiated between the capacities of the instrument to produce sound, the performer's motor control capacities, and the desired articulations of sound.

Notions that are very similar to that of musical playing techniques exist in other artistic disciplines, such as graphic and plastic arts, dance, and circus (e.g. juggling techniques), as well as in sports (e.g. ball playing techniques) and craftsmanship. In all of these fields, action is articulated with the physical properties of a particular artifact (e.g. a brush, a ball, a hammer) or just with the physiology of the body itself (e.g. in dance and certain sports) to achieve a desired effect. The history and evolution of many of these techniques have been documented by practitioners and theoreticians. The development of juggling techniques over many centuries, for example, has been compiled by Ziethen and Serena (2003). Wegener (2005) has investigated *The History of Basketball and Volleyball* and their techniques. Similar to the playing techniques of musical instruments (see for ex. Bouënard et al. 2008; Schoonderwaldt 2009; Bianco 2012), the movements extramusical techniques have been quantified, analyzed, and studied using bioelectric sensors and motion capture.[154] Smith et al. (2009) have for example used electromyogram (EMG) signals for evaluating the process of acquiring motor skills in the learning of basketball dribbling. Lees et al. (2010) have reviewed studies of kicking actions in soccer.

In the same way that it applies to techniques in the performance of extramusical expert tasks, the notion of *playing techniques* can be applied to action sequences which occur in everyday tasks, such as tooth brushing, bicycling, or cooking. In all of these cases, particular techniques do not simply emerge in the interplay of physical constraints, but also represent

[154]The motion studies of Eadweard James Muybridge are often cited as the ancestor of today's practices using digital motion capture devices (see for ex. Rosenbaum 2009).

themselves cultural artifacts. While they are named, learned, and transmitted, they are defined by characteristic features of their movement and of the effect that they produce or the task that they accomplish. As the playing techniques of musical instruments, they evolve with their artifacts and the practices they belong to.

R 32 | Notions similar to that of musical playing technique exist for other expert tasks as well as for everyday activities.

Different aspects in the understanding of playing techniques appear when focusing the attention on each of the involved factors, the user, the artifact user interacts with, and the effect this interaction produces. Even if clearly defined by the conjunction of all three factors, the notion of *affordance* is often associated with artifacts as the possibilities of action that they bear. This includes recurrent concerns in design, whether the techniques applying to a particular artifact are clearly communicated by the artifact or whether users can apply already acquired techniques to it.[155] Also the notion of *function* particularly appears from this perspective (see for ex. Houkes and Vermaas 2010). When focusing on the effect of the interaction or a task to be accomplished, particular techniques are understood and judged through notions like *efficiency* and *ergonomics* (see for ex. Stanton and Young 1999). From the users' point of view, possible keys to the understanding of techniques are *practicability* and *learning*, but also *experience*, *expression*,[156] and further *aesthetic* dimensions (see for ex. Löwgren 2009). Methodologies like *user centered* (Norman 2002) or *experience centered* design (Wright and McCarthy 2010) emphasize this perspective.

R 33 | Different keys for the understanding of playing techniques like *affordance*, *function*, *efficiency*, *practicability*, *learning*, *experience*, and *expression* appear from different perspectives focusing either on the user, the designed artifact or the resulting effect.

A particular perspective that has often guided the examination of techniques and technologies[157] consists in evaluating their efficiency in the accomplishment of a defined effect or task.

[155] These features are often referred to as the instrument's use being *intuitive*.

[156] The international conference on New Interfaces for Expression (NIME) is an example of a community of designers, technologists, and musicians focusing on expressive qualities in the design of digital musical instruments.

[157] The term *technique* here refers to the notion of playing technique as they have been discussed above, while that of *technology* refers to the objects, interfaces, and software modules that supports them in terms of particular affordances.

The evaluation of computer interfaces for the accomplishment of *pointing tasks* is a prominent example of this paradigm (see for ex. MacKenzie and Buxton 1992). Several author have proposed to apply similar methodologies to evaluation of musical instruments. Wanderley and Orio (2002), for example, have proposed to consider *learnability*, *explorability*, *feature controllability*, and *timing controllability* as criteria for the evaluation of digital musical instruments based on musical tasks such as '*simple scales and arpeggios*', '*phrases with different contours*', '*continuous feature modulation*', and '*simple rhythms*'. However, such approaches have been vividly criticized by instrument designers and by theoreticians. Stowell et al. (2009) insist that '*musical interactions have creative and affective aspects*' concluding that '*they cannot be described as tasks for which e.g. completion rates can reliably be measured*'.

Even if existing musical practices may appear as well defined tasks especially in learning, it is difficult to apply this perspective more generally to the evaluation of new instruments and interactive audio applications. The notions of *playing techniques* and *playing scenarios* ultimately underline a dimension of *play*. Gaver (2009) defends a ludic approach to design and categorically rejects task-oriented evaluation. He insists on the difficulties '*to conceive of a task analysis for goofing around, or to think of exploration as a problem to be solved, or to determine usability requirements for systems meant to spark new perceptions*' and proposes several alternatives before concluding that designers '*need to use their personal experiences as sounding boards for the systems they create.*'[158]

Instead of *playing technique*, authors studying the design of digital musical instruments often evoke the notion of *mapping*. Over the past decade, the investigation of different *mapping strategies* and the theorizing on mapping has become an important concern among instrument designers, such as for example in the NIME community (NIME Steering Committee 2013). The focus on mapping in the design of digital musical instruments is justified by the layered architectures that underlie this design approach (see for ex. Miranda and Wanderley 2006). These architectures strictly separate the analysis of incoming motion data streams from the synthesis of outgoing sound streams. *Mapping* here is defined as the central processing layer connecting the output parameters of the input layer with the input parameters of the output layer. However, even if the mapping of parameters is inevitably part of the design of interactive audio applications, such mappings have to consider the affordances and constraints of the overall system (Tanaka 2010; Magnusson 2010).[159] Mappings are an important detail – among many others – in the design of the instruments affordances and it is difficult to imagine this design without the designer anticipating the way the instrument is played. While the design of an instrument can not be reduced to the design of its mappings, it actually makes sense to reduce the design of an instrument to the design of playing techniques. After all, its playing techniques completely

[158] In (Gaver 2009).

[159] Chadabe (2002) has stated that mapping '*is a useful concept when applied to the structure of electronic instruments modeled after traditional acoustic instruments*'.

define the instrument. This includes the playing techniques anticipated by the designer as well as those invented by the users.

> R 34 | The design of interactive audio applications essentially consists in the design of their playing techniques and scenarios.

Apart from the purely technical definition of *mapping*, the term is often used to describe more generally the design of congruencies between the players' actions and the produced sound. Usually, the design of such congruencies requires the adjustment of a variety of components – or *actors* – involved in the processing of a digital instrument including devices, algorithms, representations, and system components (see *2.2.4*). This adjustment ultimately allows the players to control his actions of producing sound and to obtain the desired articulations of sound through the instrument. Given that the control of the instrument depends as much on the possibilities of the as on the players' capacities, the design has to anticipate the way the instrument will actually be played. Here again, mapping appears as a particular point of view on the design of playing techniques. This point of view focuses on creating meaningful congruencies between the players' actions and the articulation of sound.[160]

The relationships between gesture and sound have been comprehensively investigated and theorized in the works of Godøy et al. (see also *2.2.2*). Godøy (2009b) advances the concept of *gestural affordances* in music and sound. This concept preserves a justified ambiguity between gestures which actually produce sound and gestures which accompany sound, as well as between action perceived through sound and sound-related actions performed by a player, listener, or dancer. While Tanaka et al. (2011), limit *gestural affordances* to an instrument's possibilities of producing sound through gestures, Godøy extends this notion to possibilities of associating gestures to sound. Hereby he considers all possible associations including musical performance, dance, and spontaneous movements of listeners. This notion of *gestural affordances* is also related to the concept of *gestural-sonorous objects*. Godøy (2006) develops the idea of *gestural-sonorous objects* as an extension of Schaeffer's *typo-morphology* (Schaeffer 1966). He underlines the gestural character of Schaeffer's *sonorus objects* and connects it to recent studies on sound-related movement, as well as to the principle of *chunking* in the perceptual processing of sensory streams (Miller 1956; Godøy et al. 2010).

Playing techniques can be understood as *gestural-sonorous objects* when focusing on the morphological aspects of sound and motion.[161] This perspective allows for abstracting in-

[160]These congruencies have to be *meaningful* for the players' as well as from the listeners' point of view (Kvifte and Jensenius 2006).

[161]The notion of *morphological aspects of motion* here may refer to trajectories but also to the temporal evolution of other perceived movement features such as speed and effort on different temporal scales.

stances of particular playing techniques, such as the bow strokes in the studies cited above (Rasamimanana 2008), from the physical properties of the involved body parts and artifacts. An abstraction that fits the idea of playing techniques as action sequences that are motivated by the fulfillment of a particular temporal evolution of specific sound features. This abstraction of ideal sound qualities can be found in traditional musical practices. It still strongly resonates in Schaeffer's concept of the *sonorus object* referring '*to every sound phenomenon and event perceived as a whole, a coherent entity, and heard by means of reduced listening which targets it for itself, independently of its origin or its meaning … a sound unit perceived in its material, its particular texture, its own qualities and perceptual dimensions.*'[162] However, the perspective of playing techniques allows for including the physical constraints that strongly determine the relationship between gesture and sound into the investigation of this relationship. Especially for the understanding of coarticulation, the consideration of these constraints is indispensable (Godøy et al. 2009; Bianco et al. 2010).

Bevilacqua et al. have worked on a series of approaches, based on machine learning techniques, that can derive a model of movement-sound relationships from the movements performed by a listener. The first system developed in this series has been the *gesture follower* (Bevilacqua and Müller 2005; Bevilacqua et al. 2007, 2009). This system requires the user to perform a movements while listening to the playback of a pre-recorded sound or by recording a sound (i.e. made with the voice) synchronously to the performed movements.[163] Once one or multiple movement sequences have been associated with sounds in *learning* mode, in *playing* mode, the system continuously matches the learned movement sequences to the incoming motion data and synchronously plays back the associated sounds. The sound is adapted (i.e. through time-streching) to the tempo variations of the performed movements in respect to the learned sequences.[164] The *gesture follower* has been used in various applications to enable users to explore arbitrary movement-sound relationships and to perform with recorded sounds. In a certain sense, the system can be seen as a means of creating playing techniques through listening.

This idea has motivated further research projects on the multimodal analysis and modeling of movement-sound relationships in collaboration with Caramiaux (Caramiaux et al. 2009; Caramiaux 2009; Caramiaux and Schnell 2009; Caramiaux et al. 2010a,b, 2011; Caramiaux 2012) and Françoise (Françoise 2012; Françoise et al. 2013b,a). The approaches developed in these projects seek to overcome the *gesture follower*'s limitation to purely temporal relationships between movements and sound by extending the *temporal mapping* (Bevilacqua

[162] In (Chion 1983).

[163] The performed movements can be captured by any motion capture interface. However, the interface has to deliver parameters that are salient for the movement-sound relationships performed by the user.

[164] Within certain limits, the system can follow tempo changes and jumps in respect to the learned movements sequences. Beyond these limits, the behavior of the system is undefined but in any case stays coherent with the user's movements.

et al. 2011) to further relationships derived from the congruencies between motion and sound features. Inspired by applications that use similar interactive machine learning techniques, Françoise et al. (2013a) describe their approach as *Gesture-Sound Mapping by Demonstration*.[165] Since in in all of these approaches, mappings are learned from actual prototype performances, they implicitly take into account the constraints of the user's motion and the sound processing system.

R 35 | Future applications based on statistical modeling can learn playing techniques through their demonstration by the player.

Many of these research projects exploring the relationship between movement and sound (see for ex. Bevilacqua and Müller 2005; Godøy 2009b) make use of ubiquitous technologies. These technologies are based on camera-based motion capture systems and miniaturized sensors attached to the user's body. Through these technologies, playing techniques emerge directly in the articulation of the players' movements and sound without any apparent mediation by a tangible object. Atau Tanaka, for example, has performed with an instrument based on electroencephalogram (EEG), electromyogram (EMG) and electrooculogram (EOG) that capture bioelectrical signals related to muscle activity (Tanaka 2000). The *BioMuse* interface transforms these signals into data streams controlling digital audio synthesis. Orchestra conducting and dance are starting points and design metaphors for many other applications using motion capture interfaces (see for ex. Marrin Nakra 2000; Salter et al. 2008; Todoroff et al. 2010; Mills III et al. 2010). Even if the notion of instrument usually implies a tangible object, *playing techniques* can equally be defined for these applications.[166] Instead of interacting with a physical object outside their body, the performers play with – and within – the constraints and possibilities of their body in conjunction with constraints and possibilities of the interface and the digital audio processing that it is connected to.

R 36 | The notion of playing techniques equally applies to applications that do not involve any tangible interaction.

Another opportunity that arises from the use of miniaturized motion capture devices lies in the possibility to augment existing objects by digital actors. In the case of *Augmented in-*

[165]The system developed by Françoise et al. (2013a) is based on multimodal hidden Markov models which are also used for acoustic-articulatory inversion and speech-driven character animation.

[166]When considering voice as an instrument, even singing techniques can fit the above definition of playing techniques as the interplay between the performer's motor control, the instrument's capacities to produce sound, and desired articulations of sound.

struments (Bevilacqua et al. 2006) and *hyperinstruments* (Machover 1992), these objects are musical instruments. The enhancement of existing instruments by additional possibilities allows – and even requires – the performers and composers to adapt existing playing techniques and to add new playing techniques to the instruments vocabulary. Using the same technolgies, virtually any object can be transformed into a musical instrument (see for ex. Rasamimanana et al. 2010). In this context, it is very interesting to examine how the *interaction techniques* that are already part of the usage practices of a particular object can be integrated into *playing techniques* of the musical instrument that it becomes (Rasamimanana et al. 2011, 2012). But also in extramusical domains, these technologies have a great potential to integrate digital sound into the interactions with existing everyday objects (see for ex. Rocchesso et al. 2008). These enhancements inevitably – and intentionally – change the way users interact with a given object and allow for clarifying existing or for creating new interaction techniques within the practices it supports. Hereby, the capacity of an object to produce sound ultimately creates references to musical esthetics.

It is interesting to notice that not all parts and aspects of the players' movements in the performance of a particular playing technique necessarily contribute – or obviously seem to contribute – to the produced effect like the articulation of sound or ball movements. This is already the case for many atomic actions such as a bow stroke, ball dribble, or catch, but especially for coarticulatory movements that, for example, connect one violin bow stroke to another or a ball throw to a catch. The movements of a juggler may illustrate this idea insofar as the movements within a particular juggling technique do not constantly influence the movement of the juggling balls. However, the juggler's movements are perfectly adjusted to the dynamics of the balls' (i.e. their inertia, falling, and bouncing). His or her hands may follow and anticipate the ball trajectories even in absence of direct contact.

Several studies have paid attention to ancillary gestures that occur within the playing techniques of musical instruments and revealed their importance in musical performance (see for ex. Wanderley et al. 2005; Rodger et al. 2013).[167] While the players are usually seen as intentionally causing a particular effect, the causal relationship between the players' and the applications actions can be reversed within complex interactions. The players hereby adjust the articulation of their movements to the actions generated by the instrument or application or even accompany these movements. All of these actions are part of the players' engagement with the instrument or application, whether they actually impact on the produced sound or not. On a similar account, Godøy (2009a) emphasizes that the distinction of *sound-producing* gestures and *sound-accompanying* gestures, such as dancing or marching '*may not always be so clear*' arguing that '*musicians may make gestures in performance that are probably not strictly*

[167] Wanderley et al. (2005) show that the ancillary gestures of clarinet players actually cause sound changes. Grond et al. (2010) discuss different technique to sonify these gestures using motion tracking and digital audio processing.

necessary for producing sound, but may be useful for reasons of motor control or physiological comfort, or may have communicative functions towards other musicians or the audience.‘[168]

The idea of interaction as an alternation of influence between the player and the application can be further extended. With certain applications, the player may control only particular aspects of complex sound processes, while other aspects are automatized by the application. This is often the case with applications based on recorded sound materials and/or generative processes. For these applications, playing techniques, instead of producing sound, aim at co-producing sound or at modulating the sound produced by the application. Consequently, rather than ball juggling techniques, they may more resemble a conversation or techniques of navigating a boat in stormy whether.[169] In all of these applications, the player's actions simultaneously or alternately consist in causing sound and reacting to the sound produced by the application within a coherent sequence of actions that constitute a particular playing technique.

R 37 | The playing techniques of audio applications are constituted by action sequences in which the player may simultaneously or alternately cause sound and react to the sound produced by the application.

This definition may apply to interactions on virtually any timescale. On the level of audible frequencies, it may apply to the interaction between the horn player's lips and a resonating tube. On the level of a musician's gestures, it could apply to the interaction between the violin player's arm, hand, bow, and the rest of the instrument (see also Rasamimanana et al. 2007).[170] Aspects of technique that are very similar to playing technique on larger temporal scales are generally considered as elements of musical style and form such as *phrasing* and *ornamentation*, as well as *riffs* and *licks* in jazz and rock music. Aspect of technique on even larger timescales would concern compositional aspects and interactions in which the player, instead of the instrument, is turned towards the dialog with other musicians or the public. At this point, the notion of playing techniques extends into notion of *scenarios* and *strategies*. In these terms, a musical composition, as the anticipation of a musical performance in form of a score, would fit the notion of *scenario*, whereby techniques of improvisation and interpretation are better associated with the notion of *strategy*. While the notion of *scenario* implies a point of view outside a particular anticipated interaction, the notion of *strategy* appears from a point

[168] In (Godøy 2009a).

[169] The metaphor of navigating a boat in stormy whether has been repeatedly evoked by Joel Chadabe in private conversations about different aspects of control in interactive music systems.

[170] In the case of the violin bow it is particularly interesting to notice how the motion aspects of playing techniques integrate vibratory and gestural aspects on a large range of frequencies spanning over the player's audition, haptics, and proprioception.

of view within the interaction. A designer anticipates a particular scenario for a user, but it is the user who applies a particular strategy in his or her engagement with a particular environment. Moreover, the notion of *scenario* implies the anticipation of rather concrete action sequences,[171] while the notion of *strategy* usually refers to a more abstract (i.e. procedural or methodological) anticipation of action.

From this perspective, also playing techniques can be seen as *strategies* or as *scenarios*. The term actually includes both aspects. In the design of interactive audio applications, all three, playing *techniques*, *strategies*, and *scenarios*, can be understood as similar notions referring to the anticipation, formalization, and transmission of complex actions and action possibilities on different levels of temporality. In this sense, all three notions are strongly related to the notion of affordance.[172]

R 38 | Playing techniques, playing strategies, and playing scenarios can be understood as similar notions referring to the anticipation, formalization, and transmission of possibilities of action.

The extended possibilities of generating and supporting action through digital technologies include complex interactional *mechanisms*, *behaviors*, *strategies*, and *scenarios* that are all together objects of design. Hereby, the boundaries between material and social interaction, as well as between human-computer and human-human interaction are fluid. Digital actors may implement behaviors that formerly have been associated with human abilities. Two such applications are for example Pachet's *Continuator* (Pachet 2002, 2004) and the *OMax* system developed by Assayag et al. (Assayag et al. 2006a,b; Assayag and Bloch 2007). Both applications allow for musical interactions in a form of dialog. Imitating the musician's style of playing, they respond to – or *continue* – musical phrases. Both applications model musical style by applying online machine learning techniques to the live input performed by the musician. While the *Continuator* and a first version of *OMax* interact via symbolic input (i.e. MIDI messages), an evolution of *OMax*, first published as *Ofon* (Assayag et al. 2006a), also interacts through sound input.[173] Insofar as the interaction with the *Ofon* system recomposes the musical

[171] Carroll (1995) gives a comprehensive overview of different aspects – and strategies – in *scenario-based* design.

[172] The term *playing techniques* translates in German also to *Spielweisen* (literally *kinds of playing*) and in French to *mode de jeu* (literally *modes of playing*) accentuating their definition as possible ways of playing among other possibilities.

[173] In addition to machine learning techniques, the *Ofon* system (Assayag et al. 2006a), applies music information retrieval methods to the live sound input and recomposes sound segments cut out of the musicians original musical discourse to new musical phrases in response using concatenative synthesis.

phrases played by the interacting performer, the interaction scenario of the application restages the player's solo performance as a musical dialog.

Between the design of digital musical instruments and these complex systems, lies a continuum of different interactions and behaviors. A player may engage with these behaviors through different playing strategies and scenarios. On one end of the scale, players and digital systems are tightly coupled into material interactions. On the other end, this coupling may seem rather loose, but ultimately requires the same permanent and mutual attention of the involved human and non-human actors. Whether these interactions are understood through metaphors of resonance, harmony or counterpoint, and whether they are regarded in terms of control, entrainment, or dialog only depends on the intentions of the observer. Obviously, it makes sense to understand interactive systems that provide interactions inspired by human behaviors through metaphors of social interactions (see for ex. Lewis 2006).

For many of the applications that are presented in chapter 3, the design of playing techniques and scenarios went hand in hand. These playing techniques and scenarios inherit central aspects from the playing techniques and scenarios in musical practices, as well as from other practices of playing. They are articulated with the mechanics and dynamics of objects, spatial topologies, and rule systems, as well as with the players' intentions and capabilities. All of these factors participate in the reenactment of actions, events, and practices that are evoked by the applications. While the anticipation of playing techniques and scenarios is a crucial element of design, players should be enabled to freely follow their own inspirations and invent their own techniques and strategies of playing.

R 39 | The anticipation of playing techniques and scenarios is a crucial element in the design of interactive audio applications.

However, the players should be enabled to freely follow their own inspirations and invent their own techniques and strategies of playing.

Conclusion

This section has explored different aspects of the notion of *reenactment*. It has been motivated by the idea to create a common framework for the understanding of listener and performer interactions with digital applications. In these explorations, interactive audio applications have been regarded in the legacy of both, sound reproduction and musical practices. The introduction provided a definition of *reenactment* in the specific context of interactive audio applications based on recorded sound materials. This definition has been enriched and extended throughout the argumentation of the section.

The first part of the section laid out the phenomenological foundations of the notion of reenactment grounded on the enactive approach to perception and cognition. Some introductory reflections explored how common musical practices particularly display and valorize causal relationships between actions of performers' and instruments and sound. The following argumentation engaged in affirming listening as action showing how listening relates to and extends into further externalized actions performed by the listener. These ideas were applied to interactive audio applications from the point of view of sound reproduction practices and interactive sound installations. In this exploration, the notion of *congruency* has been identified as an important element in the understanding and design of interactions. An interesting property of this notion is that it can be arbitrarily applied to different aspects of action regarding a large range of temporal and spatial dimensions as well as to artifacts and representations.

The focus of argumentation on *action*, instead of *representations* and *artifacts*, introduced in the first part of the section, has been maintained in the second part. This part explored epistemological and semiotic aspects of interactive digital applications. The phenomenological perspective and the focus on action have allowed for investigating interactive technologies beyond commonly maintained distinctions of different categories of interaction and knowledge.

A critical discussion of the distinction of incorporation and inscription practices has introduced the idea that interactive digital applications are media, which support the operation with different aspects of knowledge through nonverbal multimodal interaction. Hereby, it has been proposed to understand interactive digital applications as a form of *essay*. The specificity of interactive digital application compared to other forms of art and entertainment has been defined through three points: the focus of design on how the users interacts, the implication of the users' proprioception, and the users' experience of themselves as actors. Through a discussion of hermeneutic and semiotic aspects, *reenactment* as been defined as a transactional account of the relationships between the designers, the users, the application, and the world.

These ideas have been carried out further, by exploring the articulation between the concepts of *metaphor* and *affordance*. Finally, *playing technique*, along with *playing strategy* and *scenario*, has been identified as a central factor in the design of interactive audio applications. The investigation of these notions and their implications has acknowledged the fluid boundaries between techniques and scenarios, as well as between material and social interactions involving players and interactive systems.

CHAPTER 3

Design in Practice

Introduction

This chapter presents a set of applications that allow for illustrating the ideas and concepts exposed above. It gives the opportunity to complete the abstract considerations of the previous chapters by concrete insights into the craftsmanship of designing interactive audio applications based on recorded sound materials. All of the presented applications have been realized within the past few years. They involve the transformation of digitized sound materials and restage – or *reenact* – these materials within playful scenarios. They all have been created to foster their users engagement in playing (with) sound and music – alone and together with others. All presented scenarios privilege intimate situations of play rather than performances on stage. However, some of the applications have been performed in front of an audience. In some cases, the players have been involved in the design of the applications.

The design of the many applications described below has been inspired by existing audio materials. In this sense, the scenarios of these applications are actually designed as reenactments of the audio materials they are based on (see introduction to 2.2). These audio materials have been analyzed and annotated automatically or by hand. Their annotations include audio descriptors (see for ex. Peeters 2004) that are automatically extracted from the recorded materials, aligned symbolic music representations,[1] as well as further manual segmentations and descriptions. These annotations allow for rendering[2] – or *animating* – the recorded sound materials through transformations in congruence with the players' movements and gestures, while preserving certain of their characteristics and inherent congruencies. The annotation of the note onsets of a music recording, for example, may allow for preserving characteristic properties of these events, while changing tempo, order, and other properties of the events and their sequence. Similarly, the annotation of the energy and other sound characteristics allows the audio system to select sound events that fit certain characteristics of the player's movements. On a smaller temporal scale, the annotation of elementary waveforms of the recording

[1]When centering the analysis of music on recorded performances, aligned symbolic music representations appear as annotations of recorded audio streams by the *onsets* and *durations* of musical events (i.e. *notes*), as well as by pitches and dynamics.

[2]This chapter will consistently use the term *audio rendering* to describe the processing within the interactive audio system including generative models, sound synthesis, and spatialization.

of a harmonic sound, such as produced by the voice or certain musical instruments, permits to transform timing, timbre, and/or pitch while respectively preserving the other sound parameters (Schnell et al. 2000).

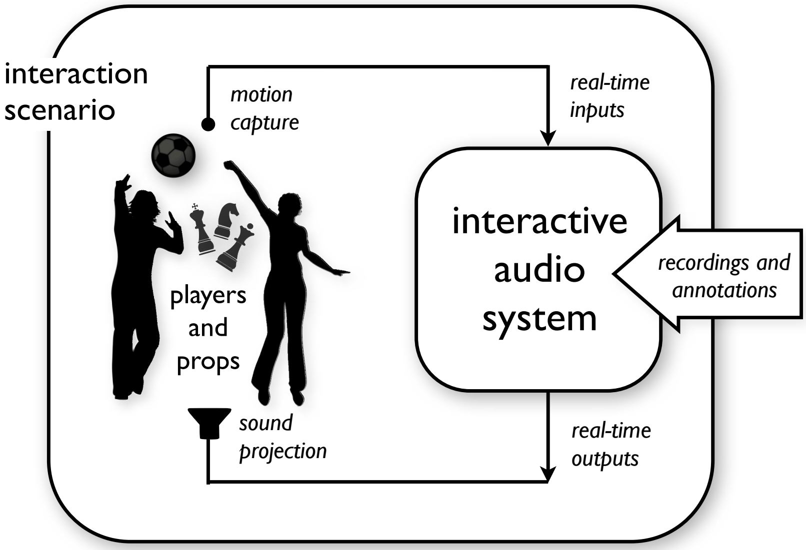


Figure 3.1: Schematic overview of an interaction scenario including the players with optional props, and an interactive audio system processing the recorded and annotated audio materials.

Figure *3.1* shows a schematic overview of an interaction scenario. The scenario essentially includes the players, with optional props, and an interactive audio system. The motion capture of the players' gestures and movements constitutes the input data streams of the interactive audio system. The system processes the annotated audio materials to render the audio output projected into the scene. Most of the applications described in this chapter use the MO wireless motion sensor modules based on inertial measurement units (Fléty et al. 2004; Bevilacqua et al. 2007; Schnell et al. 2011b; Rasamimanana et al. 2011; Bevilacqua et al. 2013). Some applications use similar sensors embedded into mobile phones. Only the *MindBox* project described in *3.2.1* uses different sensors built into a dedicated setup. The projection of the audio output into the environment of the application in all cases uses standard stereo or multichannel systems. The projected sound is spatialized through the assignment of audio channels to individual loudspeakers and intensity-based panning effects.

Figure *3.2* shows an interactive audio system using the variables e, a', and s' as they have been defined in the introduction of *2.2* (see figure *2.2*). The data sets e here represent the recorded scene. Consequently, the data sets e' would refer to the data set that are involved in the reenactment. These data sets include the players' actions captured as a' and the produced sound projected from the audio outputs s', but also the interactive processing based on further

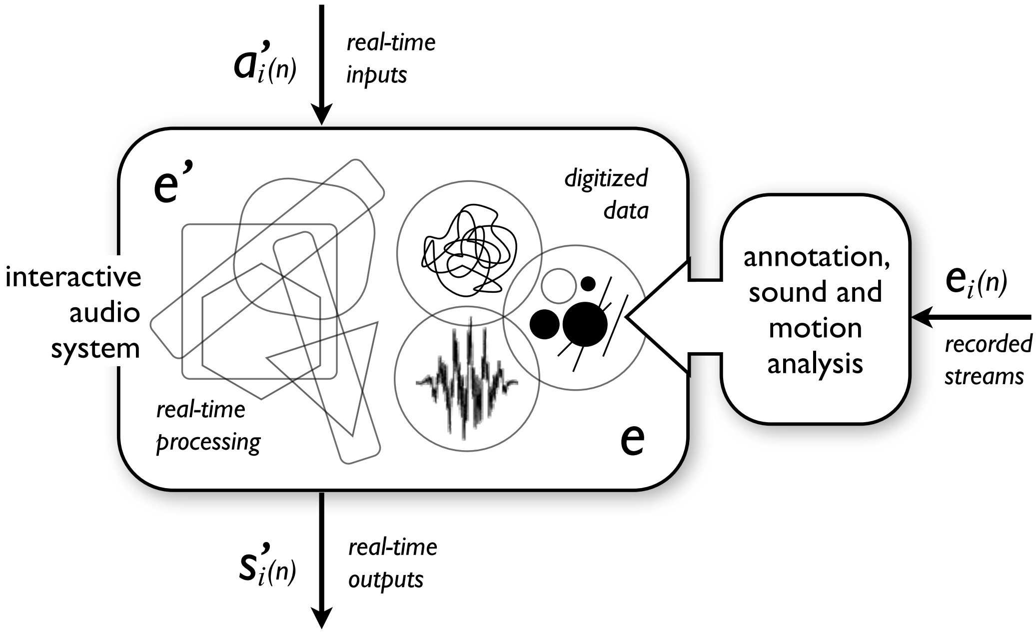


Figure 3.2: Schematic overview of an interactive audio system using the variables defined in figure *2.2*.

representations such as algorithms, models, and additional data sets. While the data sets e are congruent with the recorded scene E, the data sets e' are congruent with its reenactment E'.

However, in practice it is difficult to separate these two data sets. For example, it is not clear to which side should belong the annotations of the recorded data. Their motivation is to assure inherent consistencies of the recorded data sets, while they are transformed in congruency with the players' actions. Similar entanglements between the recorded scene and its reenactment can be found for many parts and aspects of a particular application. For some applications, this ambiguity goes so far that the recorded data is actually recorded and edited with a particular reenactment in mind, anticipating the interaction scenario along with the players' actions and the applied audio processing.

The symmetry between the recorded scene and its reenactment is taken into account in *3.3*. Here, the interactive system X is depicted in the role of mediator between the scene represented by the recorded audio materials and the interaction scenario.[3] What here is called *recorded scene* may be a music performance, but also, for example, an environmental sound texture,[4] or a person speaking. In any case, the interactive audio system has to mediate between the

[3] *Players*, *props*, and *projections* are the principal elements of a scenario. They are either themselves perceived as actors or support the perception of actors.

[4] Sound textures such as produced by natural phenomena, animals, and machines are usually defined as quasi-stationary sounds often have a short-term stochastic or structured behavior (see also Strobl et al. 2006). From the point of view of action, they are generally determined by the polyphony of a great number of actions. Sound textures produced by rain, for example, are determined by the perception of a polyphony of raindrops that are not individually distinguishable as such, but however perceived through the textural character of their accumulation.

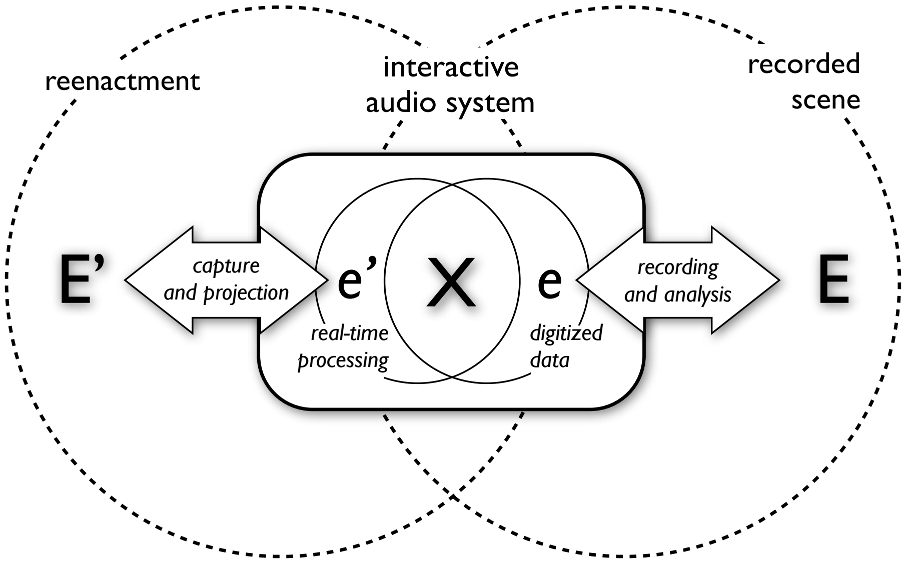


Figure 3.3: Revised version of figure *2.2* illustrating the reenactment of a recorded scene.

preservation of certain – morphological and semantic – aspects[5] of the recorded sound and the players' actions within the interaction scenario.

The mediation between transformation and preservation is an important concern in the design of interactive audio systems. It concerns as much the mediation of low-level congruencies of audio and motion capture signals, as the mediation of spaces, actions, intentions, and practices. As developed in *2.2*, the notion of *reenactment* emphasizes a perspective on this mediation that pays particular attention to the referential and reflective aspects of action. In the case of certain musical applications, this mediation might be understood as the mediation between the competences and intentions of the musicians who perform the recorded pieces and the competences and intentions of the players.[6]

In this sense, applications that propose new interactions for the interpretation of music create novel opportunities to collaborate with musicians on novel recording-based media. These media, instead of representing *one* interpretation of a particular piece of music, represent a space of different possibilities through which the player can navigate. Even in the case that the musicians – and other actors involved in the recordings – stay anonymous, it is important to note to what extent the musical and acoustic qualities of these applications rely on details

In addition, a sound texture like rain generally are modulated by further actions attributed to actors such as wind or to somebody moving an umbrella.

[5]While the *morphological* aspects consist in perceived transitions, states and trajectories of perceived sound features, the notion of *semantic* aspects refers to the listener's recognition of particular actors, intensions, and relationships.

[6]Evidently, common practices of music recording and listening based on classical recording media can also be understood as involving such a mediation of competences and intentions between the performing musician – and further actors involved in the recording – and the listener.

that are contributed by the musicians, instruments, and locations. These details include many aspects of music interpretation, such as articulations and phrasings, as well as acoustic features of instruments, rooms, and recording equipment.

All of the applications described in this chapter have been realized using Max/MSP (Cycling'74 2013) and different audio processing libraries that we developed over the past few years. Almost all applications are based on the *MuBu* library (Schnell et al. 2009) and some use *FTM & Co* (Schnell et al. 2005; Schnell and Schwarz 2005; Bevilacqua et al. 2005), an older framework that, as *MuBu*, provides advanced signal processing functionalities within the Max/MSP programming and processing environment. Most of the functionalities of these libraries have been developed to support the analysis and resynthesis of sound and motion. The audio processing ultimately allows for the animation and reenactment of sound through the interaction of players and automatic processes.

Each of the scenarios presented in this chapter represents on its own an exploratory study. Most of the presented applications have been realized in the framework of interdisciplinary projects. Some have been conceived as demos of interaction concepts and technologies. Others have been developed within collaborations on artistic projects and in the context of music pedagogy. All of these developments have allowed for exploring different concepts and techniques, as well as to confront these concepts and techniques with different groups of users. Most of the presented scenarios have been tried by a considerable amount of players, so that the descriptions of the applications can also report on observations made by designers and users. Many of these observations have informed the design of successive versions of particular scenarios, as well as the design of the scenarios of successive projects. Some of the scenarios have, at least so far, been used by only a few people at the occasion of confidential exchanges. Nevertheless, each of these scenarios has inspired further works currently under development and merit inclusion into this investigation of design in practice. Beyond the exploration of novel interaction scenarios and techniques, all of the presented projects have contributed to the concepts and ideas investigated in chapter 2.

The descriptions of the applications and their scenarios are an occasion to further concretize the understanding of the reenactment as a factor in design. In addition, the terms *reenactment* and *to reenact* are applied the descriptions assuming the connotations developed above. The efficiency of these terms – along with the concepts they refer to – can be judged according to several criteria. First of all, they should allow for highlighting general motivations in the design of the applications' scenarios. Secondly, they should support the understanding of particular actions and interactions that occur within these scenarios. Finally, they can be judged as efficient, when they fluidly integrate into the text and facilitate the description – and understanding – of the presented projects.

The table *3.4* shows a list of the scenarios described below. Even though the scenarios are not presented in chronological order, their descriptions include cross-references that show how some aspects of the design of one scenario has inspired aspects of another. Furthermore, the relationships between different scenarios illustrate how scenarios that are similar from a conceptually or technically point of view can finally create very different interactions from the players' perspective.

Section	**No.**	**Scenario**	**Year**	**Context**	**Users (Approx.)**
3.1.1	1	*Canone alla Sesta*	2009	music pedagogy	music students (< 10)
	2	*Concerto in F-minor*	2010	music pedagogy	music students (< 50)
	3	*Chaconne*	2009	demo, music pedagogy	occasional, music students (< 10)
3.1.2	4	*Rainstick*	2010	demo	public (> 100)
	5	*Shaker,* *MubuFunkScatShare*	2010 2012	demo, performance	
	6	*Surface*	2011	demo	
3.1.3	7	*Party Guitar*	2011	performance, demo	performer, occasional (< 10)
3.2.1	8	*MindBox*	2009	installation	public (> 1000)
3.2.2	9	*Urban Musical Game*	2010	installation	public (> 500)

Figure 3.4: Table of the presented explorative studies mentioning the year of the first showing, the context, and the user group with an approximate number of players who have actually interacted with the designed applications.

Since most of the works presented in this chapter have already been published elsewhere, the discussion of these applications herein is limited to the essential aspects that are relevant in the context of this dissertation. All of the presentations include references to scientific publications, web pages, and audiovisual resources published online.

3.1 Engaging into Sound and Music through Metaphoric Action

Introduction

This section presents applications that allow their players to engage with music through metaphoric actions. In the scenarios of these applications, the players perform musical pieces and sound textures. All applications are based on sound recordings that are rendered through real-time transformations controlled by the players' gestures and movements. The notion of reenactment here is principally related to the way the players' actions metaphorically refer to actions and practices that are evoked by the recordings. This metaphoricity may appear as the cross-reference between musical and extramusical practices or as an exploration of relationships between matter, movement, and sound.

In addition to the metaphorical character of their actions, the players' are actively engaged into performing music. While the applications based on sound textures can be seen as simple digital instruments, the music-based scenarios provide their players with a range of expressive possibilities for (re-)interpreting the musical materials they are based upon.

3.1.1 Playing Bach

This section presents three applications that are based on pieces by Johann Sebastian Bach. The first two applications allow for restaging the interpretation of the Bach pieces within playful scenarios that have been developed in collaboration with teachers and students of a small music and arts school in Paris (Atelier des Feuillantines 2013). Their development was motivated by the idea to create alternative opportunities for the students to engage with musical repertoire. The scenarios have been part of the pedagogical project of the school as complementary elements that complete instrument and music theory classes (Guédy et al. 2011, 2013). The students were involved in the design and performed the scenarios at the school's annual students' concerts.

The third application has been developed for demonstrating the *gesture follower*, a technology performing continuous real-time gesture recognition and alignment (Bevilacqua et al. 2011). As the other scenarios, this application is based on a Bach piece and allows the players to perform the piece through temporal transformations of an audio recording.

The interaction scenarios of the three applications have been presented in a publication which emphasizes the fact that they implement three different paradigms of controlling the temporal transformation of sound recordings (i.e. triggering of successive sound events, continuous synchronization and tempo control, conjunction of discreet and continuous control) (Schnell et al. 2011a).

Figure 3.5: Students performing the 18th Goldberg Variation in the annual concert of the *Atelier des Feuillantines*.

Canone alla Sesta

An application created in collaboration with the *Atelier des Feuillantines* in 2009 restages the 18th Goldberg Variation *Canone alla Sesta* as a chess game. The idea for this scenario had been elaborated with the students in a music theory class. Through analyzing the piece and listening to a recording, the students developed the idea that the tight canon between the soprano and alto voice of the variation could be understood metaphorically in terms of the alternating moves of opponents in a chess game. In addition, the up- and downbeats of each voice have been associated with up and down movements such as the lifting and setting down of the chess pieces during the moves.[7] Figure *3.6* shows the soprano and alto voices of an excerpt of the piece with annotations that mark the alternating voices (figure *3.6* above) and the up and down movements (figure *3.6* below).

Before implementing an interactive application, the students sketched out and tested the scenario by moving arbitrary objects (e.g. rubber gums) on a table while listening to a recording of the piece (see fig. *3.7*). Further sketches involved a chessboard and chess pieces. A first performance at the annual students' concert used a simplified technique to capture the players' moves. In this version, the players had to release and press an FSR pressure sensor with a finger of one hand while simultaneously moving a chess piece with the other (Atelier des Feuillantines 2009). This setup was revised for the final version of the application to use the MO wireless motion sensor modules which have been attached to the players wrists, as well as a piezo microphone fixed on the chessboard (see fig. *3.7* below). In this second version, the players perform the piece directly through moving the chess pieces.

[7] Another metaphor for the up- and downbeats of each voice brought up in the discussions with the students was *breathing*.

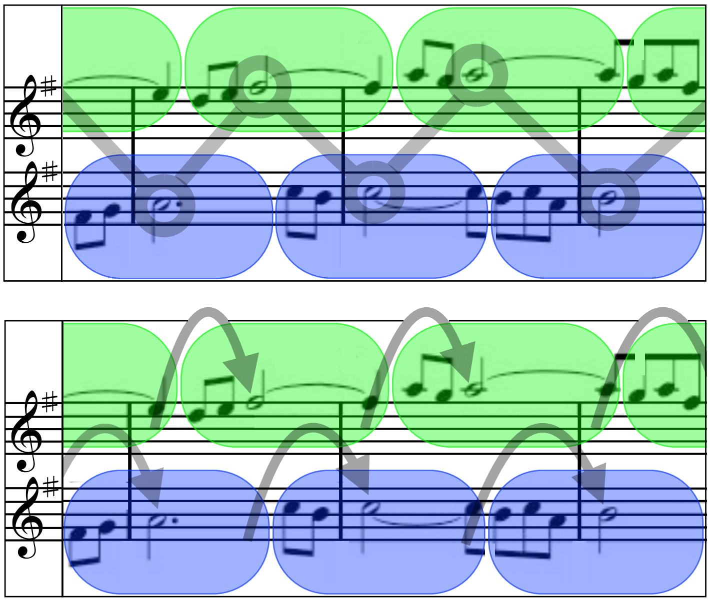


Figure 3.6: Score excerpt of the 18th Goldberg Variation (bars 5 to 8) highlighting the alternating soprano and alto voices (above) as well as the up- and downbeats of each voice (below).

The audio processing of the application is realized by a phase vocoder (Roebel 2003) that renders recorded piano performances in synchronization with the players' moves. Each voice has been recorded separately and annotated with note onset positions. The motion analysis of the players' moves recognizes the lifting and setting down of the chess pieces and follows their trajectory using a real-time alignment to pre-recorded motion templates (Bevilacqua et al. 2011). This information is used to create a temporal mapping between successive moves of chess pieces and successive groups of notes (i.e. phrases) of the recorded performances. The mapping automatically adapts the tempo, as well as the note transitions and durations, of the audio rendering. While the rendering of the soprano and alto voices is respectively controlled by one of the players, the rendering of the accompanying bass voice is derived from the moves of both players.

Even though the players' only influence temporal aspects of the rendering of the recorded performances, the application provides the players with a limited range of possibilities for expressive variations. An important pedagogical element of the project has been that the students' were involved in the design of the reenactment based on their analysis of the piece.

It is interesting to notice the differences between successive variants of the scenario on different stages of the project – the (rubber gum) sketches, the first interactive version of the application used in the students' concert, and the final version of the application. Even though they did not affect the rendering of the piece, the players' actions in the sketches already *enact*

Figure 3.7: Students of the *Atelier des Feuillantines* sketching out the chess scenario with rubber gums on a table (above). Frédéric Bevilacqua and Fabrice Guédy testing a version of the application using the MO modules on the players' wrists and a piezo microphone attached to the chessboard for motion capture (below).

the piece and the recorded performances, but also *reenact* it as a chess game.[8] Other than the sketches, the first version of the application, performed at the students' concert, introduces the possibility of reinterpreting the piece within the limits provided by the application. This act of interpretation – and the consequent difficulty for the players – was obvious for the audience. Most of the audience ignored the players' synchronous actions on the pressure sensor. Consequently, for the audience the reenactment of the piece through the chess game has been complete. However, the players are conscious of controlling the audio rendering with one hand and moving the pieces with another. The final application removes this inconvenience by controlling the audio rendering through the players' moves of the chess pieces.

Nevertheless, the reenactment of the piece as a chess game stays partial. Important aspects of the piece, as well as of the chess game, are not taken into account by the scenario. While the

[8]As many as such distinctions, that of *enactment* and *reenactment* should not be seen as a classification, but as a perspective one chooses for understanding and presenting a particular interaction scenario.

chess game, for example, can develop through very different sequences of moves, the note sequence rendered by the application always stays the same. This illustrates how the reenactment of the piece and its performance creates a particular perspective. From the players' point of view this perspective is constituted and constrained by the possibilities of interaction provided by a particular application. From perspective of an outside observer, such as the audience, further projections are possible.[9]

Concerto in F-minor – Largo

A second application, created in collaboration with the *Atelier des Feuillantines* in 2010, is based on the 2nd movement of the *Concerto in F-minor*[10] by Johann Sebastian Bach. In this piece, the solo harpsichord or piano is accompanied by an orchestra that plays pizzicato chords in regular 8th-notes throughout almost the entire movement. In the scenario of this application, the pizzicato chords of the orchestra accompaniment are reenacted by players throwing and catching a ball. Each catch corresponds to a pizzicato. Set up around the pianist, the players perform the accompaniment chord by chord. Hereby, the ball is thrown in a regular trajectories from one player to another around and over the pianist.

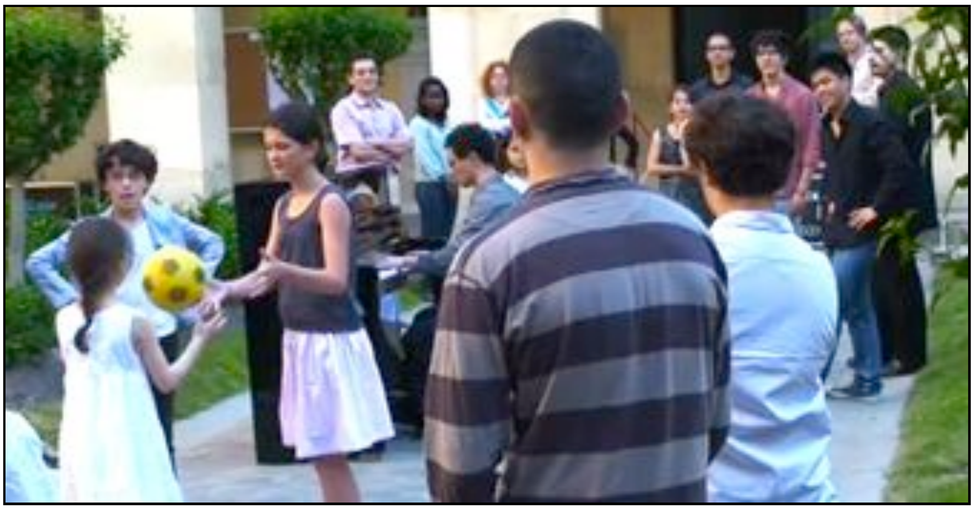


Figure 3.8: Students performing the 2nd movement of Bach's *Concerto in F-minor* with their teacher who plays the solo piano (in the center).

The audio rendering of this scenario is based on the recording of an orchestra playing the pizzicato accompaniment. Each chord onset of the recording has been annotated so that the audio system can render one segments one by one without additional transformations. The ball

[9]In this sense, the design of interactive applications and similar technologies for stage performances diverges considerably.

[10]The 2nd movement itself is in Ab-major.

is equipped with a wireless motion sensor module.[11] The analysis of the motion data allows for reliably distinguishing the catching of the ball from other movements and impacts. Since a pizzicato chord is rendered only when the ball actually is thrown and caught – or rebounds from a surface –, the players quickly associate the pizzicato sounds to the action of throwing the ball into the air.

Also in this scenario, the players' influence on the interpretation is limited on temporal aspects.[12] An interesting feature of this scenario is that the players have to concentrate together on the performance of the accompaniment in interaction with the pianist. Even if only two players are involved in the performance of each chord (i.e. one who throws and another who catches the ball), the realization of a regular tempo, as well as agogic variations demands the anticipation of the balls trajectory, and thus an adjustment of each player's actions to the actions of the whole group.

A variant of the scenario that has been realized for a public performance at the annual students' concert of the school (Atelier des Feuillantines 2010). In this extension, the students perform the piece together with their teacher who plays the solo piano. The children are disposed in multiple circles of 4 or 5 players around the piano. Each of the groups performs a section of the piece and passes the ball to the next group at the end of the section. This way, in addition to the reenactment of the orchestra by the ball players, and the pizzicatos by the ball catches, the musical form of the piece is reenacted by the groups of players passing the ball among them.[13]

Chaconne

As mentioned above, this scenario was developed as a demo of the *gesture follower* (Bevilacqua et al. 2011). It is based on a recording of the first eight bars of Johan Sebastian Bach's *Chaconne* for solo violin performed by Jascha Heifetz. In the first part of the scenario, the player performs a continuous sequence of movements while listening to the recording of the performance. In the second part, the player performs the piece through the same sequence of movements varying in tempo. The player's movements are captured by a wireless motion capture module the player holds in his hands or attaches to the wrist.

During the training phase of the gesture follower in the first part of the scenario, the motion capture data, representing the sequence of movements performed by the player, is associated with the audio recording as a template. During the performance, in the second part of the

[11] A first implementation of the scenario used a *Wiimote* controller. Further implementations used a MO wireless motion sensor module.

[12] An extended version of the application also rendered the energy of the balls impact when it is caught as a modulation of the pizzicatos dynamics (i.e. sound intensity and timbre). However, this version made the performance much more difficult and was discontinued after some initial tests.

[13] It also makes sense to claim that the topology of the players' positions and the trajectories of the ball *reembody* formal aspects of the piece.

scenario, this template is used to align the audio recording to the variations of the movement sequence in real-time (Bevilacqua et al. 2011). Hereby, the audio rendering of the application may use a phase vocoder or granular synthesis to apply the tempo variations of the movement sequence to the audio recording.[14]

The essential elements of this scenario can been applied to any audio recording. However, in the context of this study, it is interesting to consider the possibilities provided by the gesture follower for reenacting a particular piece and interpretation by different movement. A variant of this scenario that, apart from demos, has been used in the context of music pedagogy and artistic performances, consists in training the gesture follower with a sequence of conducting gestures.[15] Once the gesture follower is trained this way, the player can perform the piece through conducting gestures. Hereby, the temporal variations of the conducting gestures result in temporal variations of the audio rendering.

The sequence of movements performed during the training phase may be seen as a gestural representation that *enacts* the piece and its interpretation. This representation is a translation of how the player understands the recorded piece *in terms of* a sequence of movements. In this sense, the training of the application is an act of creating a *metaphor*.[16] Moreover, the gestural metaphor can be seen as an *analysis* of the recording by the player. This analysis may concern formal aspects, as for example rhythm and a partitioning of the piece into different segments, as well as expressive aspects, such as articulations and phrasings. In the performance phase of the scenario, the application inverts the relationship between the recorded interpretation and the associated sequence of movements. The gestural representation of the recording, which the player created in the training phase, determines the player's possibilities during the performance.

In spite of the strong constraints that apply to this scenario due to the limited representation of the geometry[17] and temporality[18] of the player's movements, playing with the application is a rewarding experience for the player. With a minimum of explanations and experience, novices

[14]When the player's sequence of movements during the performance differs too much from the recorded template, the audio rendering also may jump back and forth in the recording or freeze at a certain position.

[15]For training the gesture follower one has to conduct once through the whole recording of the piece.

[16]It is interesting to consider in this context the notions of *metonymy* and *synecdoche*. The audio rendering, in a certain sense, completes the movements of the player that only partially represents the recoded piece and interpretation.

[17]The player's movements are represented within the application as a set of motion parameters (i.e. six axis of acceleration and rotation velocity) that do not represent all aspects of the player's motion. In addition, the application does not take into account similarities between movements that are evident for the player. While for example for the player a circular movement with the hand in one direction only differs in one aspect from a circular movement in the other direction (i.e. the direction), the application is not able to represent this transformation.

[18]As mentioned above, the inversion of the relationship between the recorded performance and the sequence of movements associated by the player is based on an unidirectional temporal alignment that essentially allows for varying the playback speed of the recorded performance through the variation of the performed sequence of movements. This also means that the player has to remember the sequence of movements that he carried out in the training phase to be able to produce a pertinent variation of it during the performance.

usually learn quickly to work with the application's constraints and create different reenactments of the recorded performance. Each time the application is trained with a new sequence of movements, the performance movements may radically change, as well as the transformations resulting in variations of the movements. Thus, the characteristics of the movements that the player uses for training the system are as important as their synchronization with the recorded interpretation of the piece. The challenge – and pleasure – in playing with the application consists as much in training it with a sequence of movements that enact certain characteristics of the recorded performance, as it lies in reenacting the recorded performance through variations of these movements. The player deliberately chooses a perspective that may be similar to the interactions of conductor or further engage with musical details like an instrument player.

3.1.2 Retilting, Reshaking, Rescratching

This section summarizes three scenarios that have been developed between 2010 and 2011 as demos of the MO wireless motion sensor modules (Interlude Project Consortium 2011; Schnell et al. 2011b; Bevilacqua et al. 2013). While the scenarios presented in the last section are based on music recordings, these scenarios use recordings of sound textures.

All three scenarios restage the behavior of sound textures within interactions with arbitrary everyday objects. The associations between action, object, and sound explore metaphorical relationships between these elements and play with everyday experiences. The applications can be regarded as interactive animations. Similar to visual animations, these animations are composed of elements that refer to former experiences of the player, while introducing fictional elements extending and contradicting these experiences.

A Rainstick

This scenario reenacts sound textures as an interaction that evokes a rainstick or similar objects. The player in this scenario generates sound textures by tilting an arbitrary object. Tilting the object to one side, produces a sound evoking a granular or liquid substance, or a solid mass running inside the object. When keeping the object tilted, the sound runs out after a short time as if the substance has accumulated at one end of the object. The object has to be tilted in the opposite direction to produce sound again.

The scenario was developed out of the sound and interaction design created for a dance performance. In this performance, one of the dancers manipulated a box that produced sounds when it was moved. The music for this performance was created by the composer Pierre Jodlowski, who also contributed the sound design for the box. Some years later, in 2010, this sound design has been integrated into the interactive audiovisual installation *Gainstick* (Leslie et al. 2010) before it further evolved into a demo application for the MO modules. The original materials, recorded and edited by Pierre Jodlowski for this scenario, consists of a set of sound files, one for each sound texture. Each sound file is composed of 40 segments of a duration of one second in which the same sound texture occurs in different intensities in increasing order. This principle has been applied to sound textures evoking very different materials such as coins, stones, wood, voices, and electric guitar sounds. Through granular synthesis, sound textures of a large range of intensities can be convincingly rendered from these recorded materials.

Miniaturized wireless motion sensor modules like the MO can be embedded into practically any object to measure the object's inclination through an accelerometer. The application uses this measurement for controlling the calculation of a simple numerical model that consists in a mass which slides on a rail. The rail is limited on both ends and its inclination can be controlled by a parameter. The mass' sliding velocity is used to control the intensity of the rendered sound texture. To reinforce the impression of displacement, the sound is panned between two audio channels.

In addition to the original sound textures designed by Jodlowski, the application includes an extended mode that allows for recording arbitrary sound textures. The recorded sound textures are automatically segmented and analyzed according to an intensity measure. In this mode, virtually any sound texture and object can be integrated into the scenario. Hereby, elongated objects may better correspond to the idea of a rainstick than for example spherical objects. Certain sound textures are better for creating a rainstick-like experience than others. However, within this scenario, it is particularly interesting to explore the reenactment of sound textures and objects that do not correspond to the expectations of a rainstick.

The traditional rainstick creates a homogeneous sound texture of intense rain when tilted strongly. Carefully tilting a rainstick, generates more delicate textures whereby single sound events of percolating grains may become distinguishable. With certain sound textures a very similar behavior can be obtained. With other sound textures, the scenario may still correspond to the experience of tilting an object, for example when the sound texture evokes trickling water or rolling solid objects. For further sound textures, such as the voice and guitar sounds designed by Jodlowski, the relationship between the tilting action and the produced sound texture do not correspond to any experience in everyday life. In the extended mode of the application, the players can record sound textures that they produce, for example, with their voice. Since the recorded sounds are analyzed on the fly and instantly available for playing, the original experience of how the sound has been produced is immediately confronted to its reenactment.

The question of what actually is reenacted in this scenario can be answered from different perspectives. As the musical scenarios, this scenario can be simply regarded as reenacting the recorded sound textures. However, similar to the way that the recorded music performances refer to a musical piece and particular musical practices, also the recorded sound textures evoke further experiences. Consequently, their reenactment as *the sound caused by something moving* *inside an object*, creates a perspective on these experiences in which they are redefined in relationship to the elements that are present in the actual interaction.

It has been particularly interesting to notice how the experience of the titling interaction changes with different objects and sound textures. We observed, how the players adapt their interactions to different sound textures. Even if the form and mass of the physical object stay the same, the experience of the interaction considerably changes with different sound textures. Similarly, the interaction with one the same sound texture is very different depending on the chosen object. Using the application ourselves, we noticed how the constant relationship between inclination and acceleration imposed by the model create the impression that lighter sounding substances and objects are moving against more friction, while heavier sounds give an impression of less friction. This can be explained by the fact that the lag between the inclination of the rail and the movement of the mass can be perceived, either as inertia, or as friction. Heavier objects have more inertia, and since they move as easily as lighter objects, they are perceived as having less friction.

This illustrates how the application ultimately plays with the multimodal perception of physical parameters such as mass, friction, and velocity, as well as different states of matter (e.g. solid, liquid, and granular). This perception emerges from the way a sound texture reacts to the players' manipulation of the object. In the interaction with sound textures that cannot be easily related to the hidden movement of a substance inside the object, such as voices, this perception of physical parameters is purely metaphoric.

A Shaker

Similarly to the rainstick scenario, this scenario reenacts sound textures as the interaction with an arbitrary object. Instead of tilting the object, the player shakes it to produce percussive sound events as if playing with a shaker.

The first version of this scenario used manually annotated sound materials that had been produced for the *MindBox* installation (see *3.2.1*). The annotations mark precisely each percussive event of the recording of a performance that principally consists of beatboxing and body percussion. In addition to the segmentation, each percussive event is described by a value corresponding to the perception of its intensity. The audio rendering plays the percussive events on a regular beat grid.[19] At each beat, a new percussive event is rendered that corresponds to the intensity of the player's shaking movement. The shaking intensity is derived from an accelerometer embedded into the object that the player shakes. This can be the wireless sensor module itself or another object to which the sensor module is attached.

As the *MindBox* installation, the shaker scenario creates a semantic contradiction between the human origin of the sound events and the perfect regularity of the rhythm generated by

[19]The application generates events in a period of 150 ms corresponding to 16th-notes in a tempo of 100 BPM.

the application. By playing with the application, the players, in a certain sense, enter into this contradiction and explore it through their actions. As with the rainstick scenario, this scenario has been extended to be able to play with arbitrary recorded sounds. With this extension, the players' can shake practically any sound and object and create their own associations and contradictions.

A notable aspect of this application is that the players do not actually cause each sound event nor the tempo of the rendered rhythm. They only determine the intensity of the rhythmic sound texture by the intensity of their shaking. Nevertheless, when the object is held still, no sound is produced, and the sound energy is reactively modulated by the player's action. Other than with a usual shaker, it is rather difficult for the player to produce single percussive events with this application. Players who try to play this way, easily become frustrated. However, this frustration does not occur when the players focus their interaction on a continuously modulated shaking movement and leave the generation of rhythm to the application. We have observed that most of the players spontaneously adjust the tempo of their shaking to the rhythm of the sound rendered by the application. While for some players it seems to be easy to share the control over the produced sound with the application, other players have more difficulty to engage with application in the proposed way.[20]

(*MubuFunkScatShare*)

Recently, the shaker scenario was further extended to an application that allows multiple players to play music together. A first variant of this scenario has been presented as the *MubuFunkScatShare* performance by the performer Atau Tanaka (TEDx Panthéon Sorbonne 2012). A second version has been published and demonstrated on stage at the CHI conference in Paris (Tanaka et al. 2013).

During the *MubuFunkScatShare* performance, Tanaka records himself playing guitar riffs and scat singing before inviting members of the public to join him on stage and to play together shaking smart phones. The smart phones control the rendering of the rhythmic sound textures as described above. Since four smart phones are connected to the same interactive audio system, the rhythmic rendering of different recordings controlled by four players can easily be synchronized.

The scenario of *MubuFunkScatShare* can be presented as a novel form of mediation between a performing musician and an audience in which members of the audience can participate in the performance. The players, at the same time, reenact the musicians' performance and perform with the musician (see figure *3.9*).

[20] A study of how and under which circumstances the players achieve this synchronization and appreciate playing with the application is pending.

Figure 3.9: Atau Tanaka performing *MubuFunkScatShare* at the TEDx Panthéon Sorbonne with Natalianne Boucher and Baptiste Caramiaux.

A Surface

In this scenario, recorded sound textures are reenacted as manual actions performed on a solid surface such as *scratching*, *scrubbing*, *tapping*, and *caressing*. The gestures and movements which the player performs on the surface are translated into an articulation of recorded sound textures.

Other than most of the scenarios described this section, the player's movements are not captured by inertial sensors, but by one or multiple piezo microphones attached to the surface. The interactive audio system underlying this scenario extracts in real-time a set of audio descriptors from the sounds that are captured by the microphone. These parameters are used to control the rendering of sound textures that thus correspond in many details to the player's actions.

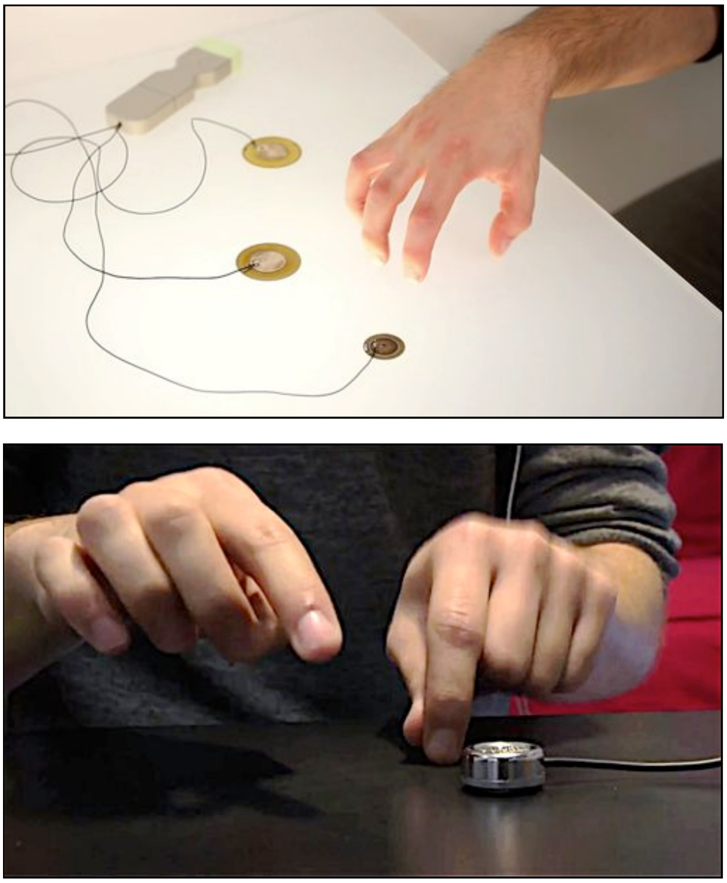


Figure 3.10: Julien Bloit performing sound textures with the MO modules extended by piezo microphones (above). Bruno Zamborlin performing an early prototype of the *Mogees* project (below).

The current version of the application[21] is based on an audio mosaicing technique (Casey 2003; Lazier and Cook 2003).[22] As for the rainstick scenario, the audio renderer resynthesizes the recorded sound textures through granular synthesis. For the rainstick, the evolution of the intensity of the rendered sound texture was controlled by a simple model of a mass moving inside the rainstick. In this application, the evolution of intensity and timbre of the sound captured by the piezo microphone determines the evolution of the generated sound textures[23] (Schnell 2011).

[21] A first version of the scenario implemented by Nicolas Rasamimanana and Julien Bloit, classified the sounds produced by different movements through a standard classifier based on Gaussian mixture models (GMM). In a training phase of this scenario, the classifier was trained with frames of audio descriptors corresponding to different movements that could be recognized during the performance. The recognized classes were used to change between the presets of a sound texture synthesizer.

[22] This version became an early prototype of the *Mogees* project (Zamborlin 2011) that, even though it soon focused on other technologies, was named *Mosaicing Gestural Surface* after this prototype.

[23] The real-time audio analysis of the incoming audio extracts MFC coefficients in intervals of 5 to 10 ms. The coefficients control a descriptor-driven granular synthesis engine rendering grains of the recorded sound textures with about the same period.

This procedure can be seen as a multidimensional mapping between the timbre space of the movements on the surface and the timbre space of the recorded sound textures. In the training phase of the application, the players can choose the range of movements that they want to use to reenact the chosen sound textures.[24] This way, the players may, for example, reenact sound textures of rich variations in intensity and timbre through careful movements with few variations.

Other than in the rainstick scenario where the players' movements influence the sound via a physical model, in this scenario the players' gestures and movement have an immediate impact on the generated sound. In addition, the player, as well as an outstanding observer, can relate the produced sound textures to the sound that the players' movements would cause on the surface. In this sense, the produced sound textures appear as an amplification and *enstrangement* of the players' actions.

3.1.3 The Party Guitar

The *Party Guitar* has been developed in 2010 as prototype for the free play mode of a dance and karaoke game in collaboration with a company (VoxLer 2013). The scenario is based on the original recording of the song *Party in the U.S.A.* performed by the singer Miley Cyrus (Pokerdance et al. 2009). Within this scenario, the player reenacts a guitar riff, which repeats literally over almost the whole duration of the song, through movements in the air that evoke the movements of playing a guitar.[25]

The guitar sounds performed by the player are derived from the guitar riff that occurs solo in the introduction of the song. In some aspects, the application is similar to the application underlying the shaker scenario described above.[26] As with the shaker application, the guitar recording is segmented into percussive events. The events rendered on a regular beat grid of 16th-notes in response to the intensity of the player's air-guitar strokes.[27] The beat grid of the audio rendering is synchronized with the playback of a recording of the song from which the

[24]The mapping between the timbre spaces of the movements on the surface and the recorded sound textures is achieved by normalizing the extracted MFCC audio descriptors. While the normalization of the recorded sound textures can be calculated after extracting the descriptors when they are loaded into the application, the normalization of the input sounds appears as a training phase of the application in which the user has to perform movements on the surface that are representative of the movements performed during playing.

[25]The initial idea behind this project was to sonify the movements of players moving to the song that can be overlaid to the recording. Inspired by the given audio recording, this idea has been replaced by the *Party Guitar* scenario.

[26]Developed about a year earlier, the *Party Guitar* scenario has contributed to the development of the *MubuFunkScatShare* performance described above.

[27]The players either hold a wireless sensor module in their hand or attach a module to their wrist. Instead of the acceleration of the players' movements, this application derives the intensity of the strokes from the rotation velocity measures by the module's gyroscopes at each beat. This way, rather than strumming a chord at a given position, the player has to strum a beat at a given time.

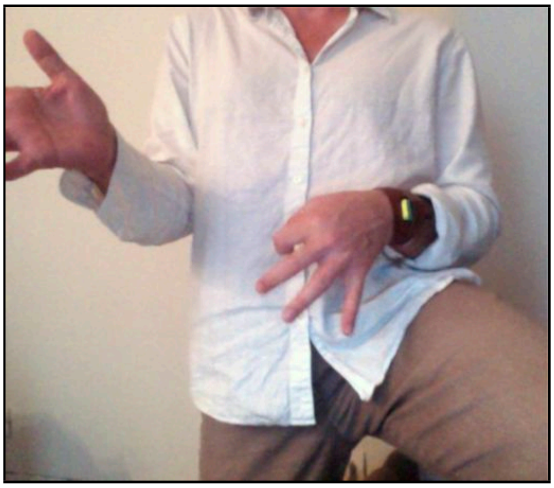


Figure 3.11: Player performing the *Party Guitar* with a wireless motion capture module attached to the wrist of his right hand.

guitar riff has been removed.[28] Other than with the shaker, the intensity of the player's strokes in this application is categorized into two levels, soft and strong. In the performance, these levels correspond to the damped short and sustaining notes or chords of the riff. Below a certain intensity threshold, a stroke is ignored and no new segment is rendered. In addition to these categories of *intensity* and *sustain*, each segment of the guitar riff is described by a harmonic category. This category corresponds to a set of chords which the segment can co-occur with. The audio rendering of the application selects on each beat a segment that corresponds to the intensity of the player's stroke and the current chord of the song playback.[29]

The interaction of the player in this scenario can be understood as *assisted* playing. Whatever the player does will fit the rhythm and harmonic progression of the synchronized playback. Nevertheless, the player has a considerable amount of freedom to perform different riffs and can produce performances of very different characteristics and qualities.

Even though this scenario does not refer to extramusical practices such as the scenarios of chess and ball games presented above, it clearly refers to a musical practice other than guitar performance which is *air-guitar* performance. In fact, the application allows the player to reenact a guitar performance as an air-guitar performance. But other than in air-guitar performance, the *Party Guitar* players not only *enact* the guitar performance of a particular piece through their gestures, but actually *reenact* it through their performance.[30]

[28]Since the guitar riff repeats almost literally, its removal could be achieved by subtracting the solo guitar of the first four bars of the song's recording from the rest of the recording where it occurs with only a few modifications.

[29]To make the performance of musically consistent guitar riffs even easier for the player, some further constraints are applied to the rendering of the sound segments.

[30]Here again, the distinction between *enacting* and *reenacting* is rather fragile. However, it may be justified by the idea that the air-guitar performer does *nothing but* imitating the actions of a guitar player, while the player of this scenario actually *re*-interprets a piece of music.

Conclusion

This section has presented seven scenarios in which players reenact recorded music and sound textures. Each of the scenarios has been briefly illustrated and discussed including technical details. The metaphoricity of the scenarios has been discussed showing how they allow the players to engage with recorded music performances and sound textures in terms of actions that refer to further actions and practices beyond them.

Two scenarios based on Bach pieces have illustrated this metaphoricity as the reenactment of these pieces *in terms of* a chess and ball game. In the *Party Guitar* scenario, the players reenact a guitar riff *in terms of* an air-guitar performance. Other scenarios have illustrated how the production of sound textures is reenacted as the performance of percussion instruments, such as a shaker and a rainstick, as well as through gestures and movements performed on a solid surface. Moreover, it has been shown how the movements and topologies implied in these scenarios enact and embody aspects of music composition and interpretation that are not made explicit in usual performances. These movement and topologies allow for creating additional metaphorical relationships between the recorded music and its reenactment by the players. The description of a scenario based on an excerpt of a Bach's *Chaconne*, presented the *gesture follower* technology that allows players to create their own gestural metaphors for the reenactment of the recorded violin performance.

Following the discussion of different aspects of reenactment in 2.2, the question of what these scenarios actually reenact has been answered in two ways. As developed in the introductions to 2.2 and this chapter, the notion of *reenactment* first of all has been applied to the act of restaging the scenes and actions, represented by recorded materials, through a particular interaction scenario. The other perspective emerged when considering that the sound recordings not only represent a particular recorded musical performance or sound texture, but further experiences and understandings of action and sound. In this sense, the discussion of the scenarios based on sound textures, has emphasized how the reenactment of these sounds explores the players' experience of movement and state of matter through sound.

The examples presented in this section have concretized the idea of reenactment as a reflection on interaction through interaction and, ultimately, the reflection on existing practices by relating them to other practices – musical or extramusical, existing or novel. The descriptions of the scenarios have underlined how the scenarios explore the boundaries between performing and listening to sound and music. It has presented the music-based scenarios, developed in very different contexts, such as music pedagogy, music performance, and computer games, as novel ways to engage with music and to mediate between performers and listeners.

3.2 Playing Games, Playing Music

Introduction

This section presents two interactive installation projects developed in 2009 and 2010. Both projects restage existing games as an occasion to engage in music performance. While the audiovisual installation *MindBox* is based on a slot machine, in *Urban Musical Game* the players play music through a ball game. In both projects, playing music and playing games merge into a single activity of playing. In their interaction scenarios, equally applies to playing music and playing games. *MindBox*, additionally explores the connections between playing recordings and gambling.

These projects share a common predecessor. The concert and installation *Players – Twilight Zone* has been developed in 1997 for the *Steirischer Herbst* contemporary arts festival in Graz in collaboration with Peter Böhm, Ulli Fussenegger, and Corinne Schweizer. In this project, music performance was staged together with chess and pinball-playing.[31] The concert involved two chess players who played with a time control of 30 minutes for each player. During the concert, an abstract graphical representation was derived from the positions and moves of the chess pieces in the game. This graphical representation guided the improvisations of three musicians playing double bass, guitar, and clarinet. Each time a chess piece was moved on the chessboard, one of the musicians improvised a musical section of a determined duration,[32] interpreting the display according to a specific system of symbols and rules. When a chess piece was eliminated from the board, all of the recorded sections that concerned the piece were replayed through a pinball game. The three pinball machines played in the concert were extended by a MIDI interfaces which connected the machines' control systems to an interactive audio system.[33] This system rendered the recorded improvisations through transformations by granular synthesis according to the actions occurring in the pinball games such as the player's controls (i.e. coin slot, flipper buttons, and spring launcher), events triggered on the trajectories of the balls, and the scoring.[34]

[31] The installation opened the day after the concert and allowed the audience to play chess and pinball to reenact the musicians' improvisations recorded during the concerts.

[32] The duration of an improvised section was determined by the time elapsed since the last move and typically between a few seconds and a few minutes.

[33] The *Bally Williams* pinball machines used in the project had 64 sensors – essentially switches and light barriers – monitoring the states and actions within the machine. These values were translated into 64 note events by the MIDI extensions. A MIDI interfaces directly connected to a machine's 8 kHz CPU multiplexing eight sensors on each of eight sensor inputs. Consequently, each of the 64 streams had a temporal resolution (sample period) of 1 ms.

[34] The interactive audio system was implemented in Max/FTS running on an *IRCAM Signal Processing Workstation* (ISPW). All other processing and display was implemented in Max running on Mac OS 9.

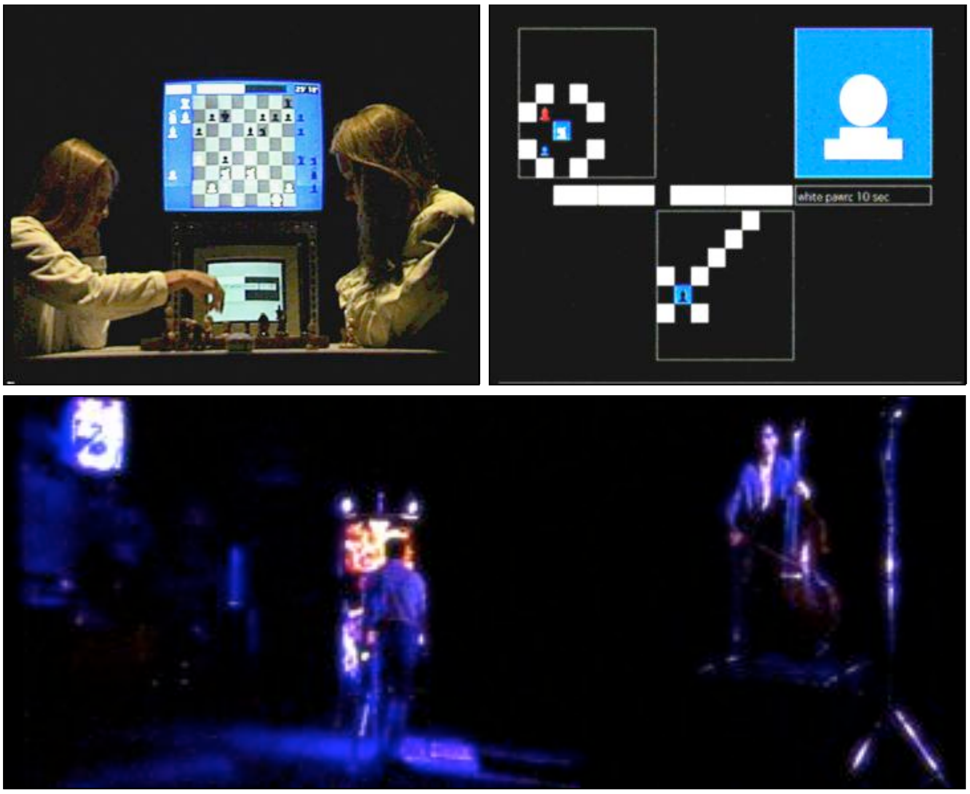


Figure 3.12: Betacam video images of the *Players – Twightlight Zone* concert. The two chess players (above left). The display guiding the musicians' improvisation derived from the chess game showing two running sections and one – of a white pawn – in preparation (above right). Peter Böhm performing with a pinball machine and Uli Fussenegger improvising on the double bass (below).

A remarkable aspect of the project is how the rules and dynamics of games, music, and sound are completely entangled. The overall form of the concert is determined by the evolution of the chess game limited to 60 minutes (i.e. 30 minutes for each player). The musicians follow – actually *reenact* – the chess game through its abstraction by the graphics display and the rule system guiding their improvisation. The improvisations are further reenacted through the pinball games.[35] While the interaction of the musicians improvisation with the chess match focuses on aspects of rules, strategy, and long-term evolution throughout the concert, the interaction with the pinball machines highlights aspects of articulation, technique, and short-term action. This entanglement of music- and game-playing that strongly determines all elements of the *Players – Twightlight Zone* performance, can also be found – even though with less complexity – in the projects presented in the sections below.

[35]The chess players as well as the pinball players in the *Players – Twilight Zone* concert were explicitly instructed to play without paying attention to sound and music. On the contrary, the musicians has all freedom to adjust their improvisations to the global musical context within the given rule system.

3.2.1 MindBox

The interactive audiovisual installation *MindBox* (Graupner et al. 2011; Graupner 2010, 2011) was created in 2009 in collaboration with Christian Graupner and Roberto Zappalà. Since its development, the installation has been shown at numerous festivals and exhibitions all over the world. It was awarded with the second prize at the 2011 third annual *Margaret Guthman Musical Instrument Competition* organized by the Georgia Institute of Technology.

The installation is based on a modified slot machine that is completed by a triptych video screen and a three channel audio system. In the principal scenes of the installation,[36] the players can reenact solo performances of dance, beatboxing, and body percussion, by operating the lever and buttons of the slot machine interface (see figure *3.13*).

Figure 3.13: *MindBox* scenography and interface design by Christian Graupner. The slot machine completed with a triptych video screen displaying three images of the performer (left). The slot machine interface with nine buttons and a lever (right).

Without the players' interaction, the materials are rendered automatically in shifting forward-backwards loops. [37] In this mode, the players – actually still spectators and listeners – can watch and listen to the juxtaposition of three different instances of the performer executing repeating patterns of movement and sound that continuously evolve. The audio rendering hereby assures that percussive events of the performances are always aligned to a constant beat

[36]This description of the installation focuses on the three principal scenes of the installation, that are based on the same interactions with the slot machine interface, but different recorded performances. The performances differ in camera perspective and style.

[37]Since the sequence played forward is always longer than the reversed sequence, the rendering advances slowly through the whole duration of a recorded performance. The rendering of a performance recording of about six minutes linear playback duration takes in this mode about one hour and a half.

grid.[38] The unreversed rendering of audio segments in a constant beat contributes to the impression of continuously performed movement patterns, even the video sequences are actually looped and rendered reversely about half of the time. Non-percussive audio segments are rendered through granular synthesis synchronously with the video images running forward and backward.[39]

During the automatic rendering, the players can intervene at any moment through different interactions using the control elements of the slot machine interface (see figure *3.13*, right). The controls allow the players to recompose – and *reenact* – the beatboxing and body percussion performance of the three displayed performers creating complementary patters of movement and sound. The three buttons on the left of the interface allow for freezing the movement and sound of each of the displayed performers for the time a button is pressed.[40] Pushing the button on the right, causes the three displayed performers to resynchronize to perfectly identical movement and sound. The buttons in the middle provide fast forward and backward in two different speeds like common media players.

Figure 3.14: Players engaging with the *MindBox* at different showings of the installation (photos by Christian Graupner and Jürgen Lösel).

A more performative interaction with the installation is provided by the lever. Whenever a player moves the lever, the automatic rendering mode is deactivated. Instead the players

[38]The percussive events are played on a 16th-notes beat grid corresponding to a tempo of 100 BPM. The time correction of the percussive events in respect to the video, necessary to generate the constant beat grid, generally stays below the level of perceptibility.

[39]The audio rendering is based on semi-automatic annotations (i.e. annotation that associate automatic segmentation and extraction of descriptors with manual corrections and complementary descriptions) of all percussive events occurring in the sound of the recordings. These annotations principally consist of a sound segments with a precise reference position for each percussive event. The reference position allows for rendering a percussive sound event precisely on the beat grid independently of it attack time. The annotations of non-percussive segments are automatically derived from those of the percussive events. The non-percussive segments are rendered granular synthesis that preserves their textural characteristics.

[40]In the sound, freezing just applies to non-percussive sound segments.

can animate the three displayed performers through the movements of the lever. Hereby, the generated movement patterns seamless continue exactly with the movement and sound at the positions where the automatic rendering mode was left. When moving the lever regularly back and forth the generated patterns are similar to those produced in automatic mode. However, by modulating their movements, players can explore different movement patterns, rhythmic structures, and sound textures. As with the automatic mode, the audio rendering generates percussive events in a constant beat synchronized to the visualized movement sequences. Using the freeze buttons, the players can reduce the animation to one or two displayed performers. When the player releases the lever, the performances seamlessly continue in automatic mode as described above.

This functioning allows the players for fluidly alternating between watching and listening to the displayed performers and performing themselves. We have observed how multiple players collaborate on exploring the possibilities of the installation whereby each player focused on one functionality (i.e. freezing buttons, fast forward and backward buttons, and manipulating the lever).[41] Some players, as well as groups of players, engaged with the installation for periods of more than an hour.

Similar to the rainstick scenario presented above in *3.1.2*, the *MindBox* scenario implies a generative model that contributes to the reenactment of the recorded performances. In the rainstick scenario, the motion model (i.e. of a mass sliding on a rail) essentially translates the movement of an object that the players' hold in their hands into sound textures behaving like sounds generated by a mass moving inside the object. Apart from the players' actions, the only forces appearing in this scenario are due to gravity and inertia. Other than these movements, the motion model that animates the displayed *MindBox* performers extends the recorded movements and sound textures into a sustained ostinato – often actually a canon – of slowly evolving patterns. Even though the audiovisual rendering in *MindBox* produces fluid movement and sound patterns, avoiding any discontinuities, the perfectly regular beat and repetition of sequences create the perception of mechanically generated action. On the other side, the visuals and sound transmit a very lively impression of a human performing with his body that continuously contradicts this perception. The tension between these elements is coherent with the tension between the general visual impression of the slot machine and the human performer.

The players may at any moment interrupt the automatic reenactment of the recorded performances as a *man-machine* and introduce themselves as actors into the game. Directly animating the displayed performers through their own actions, the players are able to perform their own reenactments of the recorded performances. Even if the interaction in this mode essentially consists in scrubbing forward and backward through a short sequence of a recorded perfor-

[41]The scenario of multiple players collaboratively interacting with the installation had not been particularly anticipated in the *MindBox* design. After having observed the public interacting with the installation at it first showing, we have adapted a few details of the interaction design to better support collaborative playing.

mance, the players can create very different reenactments of the same sequence of movement and sound and exploring different gestures and patterns. In these reenactments the displayed performers may appear as avatars that extend the players' actions into skilled performances, but also as rather ridiculous jumping-jacks constrained to follow any of the players' movements.

3.2.2 Urban Musical Game

Urban Musical Game (NoDesign 2011b,a; Rasamimanana et al. 2012) is an interactive installation that has been developed for the *Futur en Seine* international festival for technology and innovation in the Paris area (Cap Digital 2013). The installation was presented outdoors in a public space[42] for 10 days during the festival in spring 2011. Early prototypes were presented to a larger public in early 2011 during a residency.[43]

Figure 3.15: Players engaged in different game scenarios of the *Urban Musical Game* installation: volleyball (above left), basketball (above right), the bomb game (below left), music-playing (below right).

[42]The installation was presented on the place in front of the IRCAM next to the Pompidou Center in Paris (WGS 84 coordinates: 48.85973, 2.35160).

[43]These prototypes have been developed during a residency of three weeks at the *Centquatre 104* culture center in Paris. The public of the culture center was invited at the end of each week to try the current state of the developed scenarios and system components.

The scenarios of the installation bring together ball-playing with music-playing. While in some of the scenarios the players manipulate balls for making music, other scenarios focus on a ball game whereby the ball motion and the evolution of the game are related to the articulation and evolution of sound and music. Figure *3.15* shows players engaging in different playing scenarios.

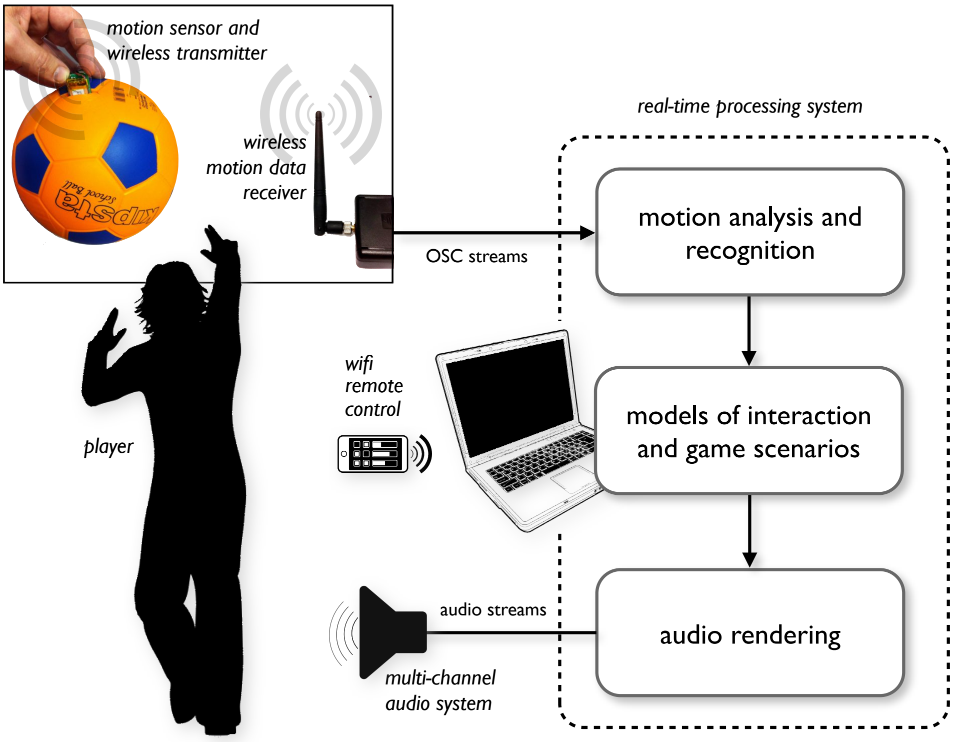


Figure 3.16: Schematic overview of the *Urban Musical Game* interactive system.

Technically the installation is based on foam balls with embedded wireless sensor modules. Figure *3.16* shows a schematic overview over the system. The motion capture data allows for recognizing different events and motion states, as well as for estimating continuous motion and posture parameters such as intensities, angles, and durations. Controlled by the motion capture data, the system renders recorded audio materials. The audio materials used in all scenarios are sound effects and loop samples that have been designed for the installation by the composer Andrea Cera. The loop samples are tracks of 8 or 16 bars performed with acoustic and electronic instruments in compatible keys and three different tempos of 66.66, 100, and 133.33 beats per minute. They are composed to around 20 to 30 arrangements of 6 complementary tracks each that superpose to complex loops of a variety of different popular music genres. By combining tracks of different genres, keys, and tempos, further variations can be created. Similar as in the scenarios described above, the rendering of these loops uses beat synchronous concatenative synthesis[44] and granular synthesis to allow for transforming the tempo and temporality of the materials in response to the players' actions. The sound projected into the installation space was spatialized through intensity-based panning.

The final installation at the *Futur en Seine* festival occupied an outdoor space of 9 x 9 meters. At each corner of the square was installed a loudspeaker system, as well as a wireless receiver which transmitted the data from the motion capture modules in the balls to the real-time processing system. An overview of this setup in given in figure *3.17*.

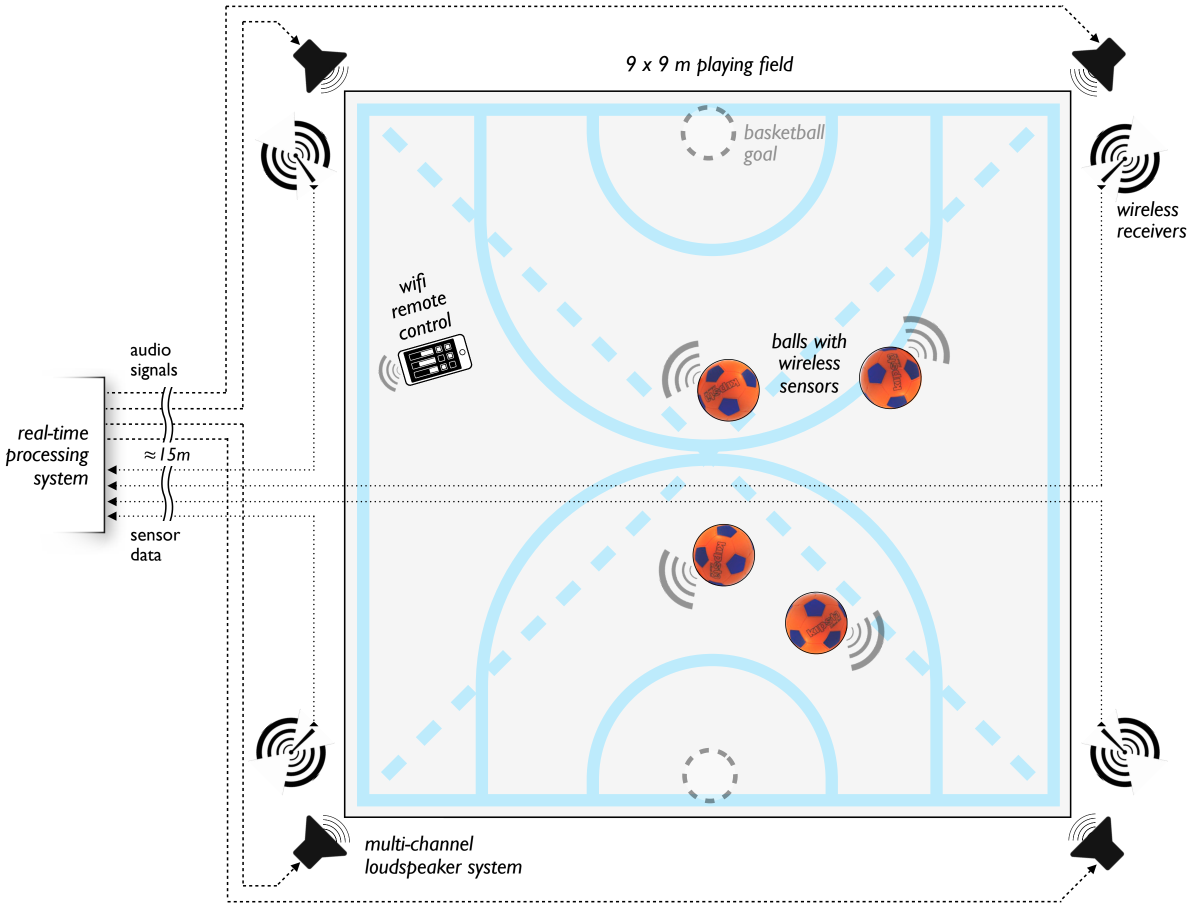


Figure 3.17: Schematic overview of the *Urban Musical Game* technical setup.

The installation was conducted by one or two tutors who invited the public to different games and accompanied the players. A remote control[45] allowed the tutors to change and parametrize the different playing scenarios from the playing field.[46]

[44]Other than in the scenarios presented above, the beat synchronous audio rendering for *Urban Musical Game* also allows for preserving groove (i.e. deviations of note onsets from a regular beat grid).

[45]The remote control was implemented using the application *TouchOSC* running on a smart phone communicating with the system via WLAN.

[46]The computer running the real-time processing system as well as the audio amplifiers were placed indoors at about 15 m distance from the installation. The cables connecting the indoor and outdoor system components included audio, Ethernet, and electric power supply.

The scenarios that were selected for presentation at the festival were *The Band*, where multiple players manipulate balls to perform music together, as well as a second, a simplified basketball game with sound effects and a musical accompaniment.[47]

The Band

In this scenario of the installation, the players are primarily engaging in music-playing. The balls here become tangible interfaces of digital musical instruments that are each played by one or multiple players. Generally, multiple players are invited to play together with three or four balls. Each of the balls, played by one or multiple players, is related to the rendering of one or multiple loop tracks (i.e. a solo instrument or ensemble). Even though the players can vary the rhythmic structure of the loop tracks through the ball movements, the rendering of all performed tracks is synchronized to a common and steady musical tempo.

Figure 3.18: Still image from a movie showing some of the designers (Nicolas Rasamimanana, Julien Bloit, and Norbert Schnell) performing different loop tracks through different playing techniques in the band scenario of *Urban Musical Game*.

While in the ball scenario presented above in *3.1.1* only catching the ball had an effect on sound rendering, for this scenario, a set of playing techniques has been designed (see figures *3.18* and *3.19*). These playing techniques have determined the implementation of movement analysis and audio rendering techniques for the application. A remarkable difference of playing music with a ball as compared to traditional musical instruments lies in the autonomy of the

[47]The other two selected scenarios were a musical volleyball game (without net) and The Bomb, a game, inspired by an existing ball game, where the players throw and catch the ball until – after a determined but unknown time – a player is eliminated when catching the ball with the the sound of an explosion.

manipulated object. Other than a musical instrument, the balls' movements are not constantly driven by the player, but also evolve without the player directly influencing its movement. This is the case, for example, when the ball is thrown, but also when it is rolled or spun on a surface. The movements of the ball here can also be understood as – and used for – sustaining the players' movements.

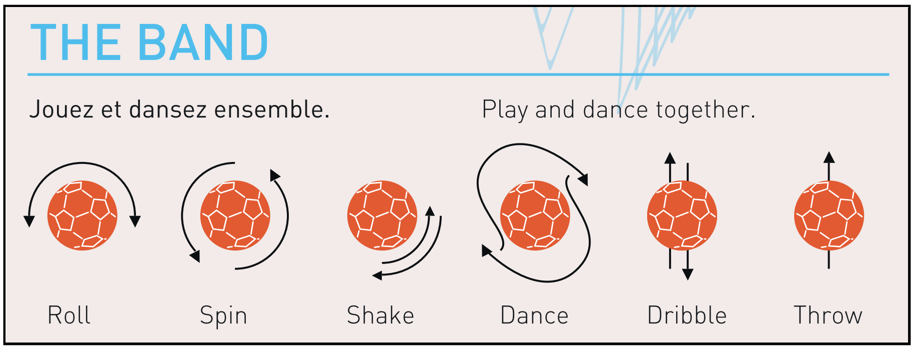


Figure 3.19: Detail of a poster inviting the public to participate in the *Urban Musical Game* installation that briefly introduces the different scenarios and suggests some playing techniques for the band scenario.

The wireless motion capture modules embedded into each ball continuously measure the forces applied to the ball and its rotation velocity in three dimensions.[48] From these data streams, the motion analysis of the real-time processing system extracts events (e.g. impacts, throws, catches), states (e.g. still, flying, shaken, spinning on surface), and continuous parameters that, that describe movement qualities which are specific for each state (e.g. intensities, velocities, angular positions).

The challenge of the sound and interaction design, especially for this scenario, consisted in creating congruencies between the evolution of states and movements of the ball with the evolution of sound and music that allow the players to engage in music performance. As in many scenarios of this kind, an important congruency concerns the correspondence between the intensity of the ball movements (i.e. of shaking, spin, dribbling, etc.) and the intensity of sound. The implementation of this detail in the audio rendering of this scenario represents an interesting example of how fine details in the design of congruencies between movement and sound influence on the way a player engages with sound. While in several scenarios described above, the relationship between movement and sound intensity has been achieved by selecting sound segments of which the perceived intensity corresponds to the intensity of movement, the audio rendering of *The Band* used a low-pass filter to produce sounds of different intensities.

[48]The modules integrate three-dimensional accelerometers and gyroscopes and provide six values of acceleration and rotation velocity with a period of 5 or 10 ms.

We have considered two techniques for applying a low-pass filter varying with the movement energy of the ball to percussive sound segments of the recorded loops. In one technique, the segments are concatenated before applying a low-pass filter to the resulting sound stream. In this case, the player perceives the continuous effect of the filter reacting to the ball movements and, consequently, engages in performing with this effect. Instead, the technique that we preferred, applied to each percussive sound segment a static low-pass filter corresponding to the movement energy at the beginning of the segment, before concatenating the filtered segments. This technique fairly approximates the behavior of percussion instruments, such as drums and plucked string instruments, in respect to the players' actions.[49] Even though the sound segments are automatically sequenced according to the pre-recorded loop samples, this technique allows the players to engage in the performance of percussive events rather than of a continuous filter.

From a certain point of view, this scenario may appear similar to the scenario presented above where a group of players reenacts the orchestral accompaniment of a Bach concerto (see *3.1.1*). Although, the performance with the balls in the *The Band* of *Urban Musical Game*, allows for influencing much more details of the rendered sounds (i.e. intensity, rhythmic patterns, pitch, sound effects), The players can explore very different reenactments of the sample loops through a variety of movements and interactions between multiple players. Moreover, they are encouraged to invent their own playing techniques and scenarios.

A Musical Basketball Game

In this scenario, the players engage into a basketball-like game. The game opposes two teams of 3 to 6 players on a 9 by 9 meters playing field as show in figure *3.17*. Instead of the usual backboards and baskets suspended above the field, the goals are two cardboard barrels of about one meter height which are placed at the center of two opposite edges of the field (see figure *3.17* and *3.20*). The game is accompanied and commented by a musical environment that reflects the dynamics and evolution of the game. The ball movements are sonified by sound effects.

Different sound effects are associated with various ball actions such as falling, catching, dribbling, and spinning.[50] The musical accompaniment is based on the same loop samples as the music-playing scenarios described above. At the beginning of the game, the music starts with a sparse bass loop and at each goal new musical elements are added to the arrangement.[51]

[49] Many of the recorded loop samples are recordings of acoustic instruments that are rendered in different intensities through the described technique.

[50] Most of the sound effects use granular synthesis to render sound textures based on sound recordings.

[51] Goals are automatically detected by wireless motion sensors inside the goals. For reliably distinguishing goals from other movements captured by the sensors, the system compares the movements captured by the sensor inside the goals with the movements of the ball.

Figure 3.20: Still images of a movie (NoDesign 2011a) showing players engaging in the basketball scenario of *Urban Musical Game*.

An arrangement of a particular musical genre is associated with each team (e.g. *electro* against *hard rock*). After each goal, the music continues with the arrangement of the team that scored the goal. In addition, the rhythm of the rendered sample loops reacts to the density and fluidity of the game.[52] In response to the variation of these parameters, the rendered rhythmic patterns alternate between different variants of the musical arrangements, as well as between three beat patterns of 16th, 8th-triplet, and 8th beats of the ongoing steady tempo.

Even if the players in this scenario principally concentrate on the basketball game, the influence of the sound effects and musical environment on the players actions and the dynamics of the game is considerable. By translating the dynamics and evolution of the game into sound and music, the sound effects and the musical environment become an extension of the game, but also a commentary. They enhance the players' excitement and entrain them in a rhythmic flow.[53] This entrainment is very different than the entrainment by a musical accompaniment that would not react to the evolution of the game.

In the music-playing scenario of *The Band* described above, the recorded sample loop tracks are reenacted as instrument performances through the interactions with the balls. Their reenactment in the basketball scenario creates a very different perspective on the same materials. Instead of musical instruments and their playing techniques, the scenario refers to different *genres* and *grooves* that evoke different emotional states (i.e. *moods*) and varying levels of excitement. These moods and levels of excitement are modulated in congruence with the basketball game. The design of the loop samples has paid careful attention to these aspect of music that become the principal object of reenactment in this scenario.

[52]To extract approximate measurements of the density and fluidity of the game, the system extracts measures of the overall activity as well as the density and regularity of impacts from the ball's motion capture data over periods of a few seconds.

[53]Many players interviewed after the game confirmed this impression. The interviews were recorded directly after the games during the installation at the *Futur en Seine* festival.

Conclusion

This section has presented three scenarios in which playing games and music-playing converge to a common activity. Since the scenarios of *MindBox* and *Urban Musical Game* rely on similar ideas and technologies as the scenarios of the last section, the descriptions have focused on aspects that distinguish these scenarios from those described before. While the scenarios described in *3.1* focused on the idea of metaphorical action, this section has insisted on the entanglement between music and games whereby *playing* itself becomes a metaphor.

The introduction to this chapter presented the concert and installation *Players – Twilight Zone* as a predecessor of the projects presented in the chapter. The description of the concert, which involved professional game and music players, particularly emphasized the strong conceptual entanglement between game-playing (i.e. chess and pin ball) and music-playing. This entanglement included aspects of rules, strategy, and technique. Similar entanglements have been found in the projects that are described in this section.

Similar to the scenarios of Bach pieces reenacted in terms of ball and chess games described in *3.1.1*, the metaphoricity of the scenarios presented in this section is created through the players' actions, as well as through the involved objects such as the *MindBox* slot machine and the balls in *Urban Musical Game*. However, in the former scenarios, the act of playing games remained largely symbolic. The chess players did not really engage in a chess game and the ball players, instead of actually pursuing a ball game, focused on interpreting a Bach piece.

This strongly metaphoric character is also present in the scenario of *MindBox*. Even though the *MindBox* players, start the game by inserting a coin into the slot machine, the aim of the game is not to win money – or loose – as is usually the case with gambling.[54] Nevertheless, the players engage with a game that inherits important elements from the interaction with a slot machine. Apart from the control elements (i.e. the lever and buttons), *MindBox* inherits from slot machine gambling the idea of constantly creating new coincidences in the composition of three display elements in perpetual movement.[55] From this perspective, the scenario achieves an exemplary synthesis of game- and music-playing.

In *Urban Musical Game*, the balanced relationship between game-playing and music-playing is achieved by proposing several complementary scenarios. While in *The Band* the players primarily engage in music-playing through ball games, the basketball scenario primarily consists in a defined game that is accompanied by a reactive environment of music and sound effects. Both scenarios build on joined competences in music and ball-playing whereby the players focus more on one or the other.

[54]Further scenes of the installation that have not been described above actually create further connections of *gambling* with *speculating* on global financial markets.

[55]A promotional text presenting the installation described the conceptual affiliation of *MindBox* to slot machine playing employing some irony by stating that the installation '*restages gambling as the player's addictive engagement in the performance of dance and music.*'

In the *MindBox* installation, but also in the music-playing scenario of *Urban Musical Game*, the players are invited to explore and improvise with objects and musical materials. Even though the installations are designed with a certain set of playing techniques in mind, we have observed how players invent their own techniques, strategies, and rules to play alone and with others.

CHAPTER 4

Conclusions and Perspectives

The previous two chapters investigated two key concepts, *animation* and *reenactment*, that inform the design of interactive audio applications based on recorded sounds. The investigation of *animation* allowed for laying out a comprehensive cultural background that drew on both current and historical influences from philosophy, science, and technology. It concluded by presenting interaction design as a form of animation in the sense of *creating and exploring action*.

However, the principal contribution of this dissertation lies in the comprehensive exploration of an extended concept of *reenactment*. The idea of reenactment first of all connects sound and interaction design to the concepts of enaction and embodied cognition grounded on phenomenology. The central proposal of this dissertation has been developed on these foundations through the investigation of phenomenological, epistemological, and semiotic aspects of sound reproduction and interaction design, but also by introducing secondary concepts such as *congruency*, *metaphor*, and *play*. This proposal presents interactive digital applications as a form of interactive digital *essays*, based on nonverbal multimodal interactions, that explore their users' experience of themselves as actors. Supporting this idea, reenactment has been defined as an action-based extension of the notion of *reproduction* in the sense of reenacting action representations *through* and *within* the users' interactions.

The ideas developed in chapter 2 were further consolidated in chapter 3, which presented nine interactive audio applications developed within the past few years. The projects are separated into two sections that emphasize respectively *metaphorical action* and the *exploration of play* in music, games, and sports. However, metaphor and play are essential factors of the sound and interaction design of all the application scenarios which have been presented.

Conceptual Advances

The in-depth investigation of concepts in this dissertation largely exceeds the framework of sound and interaction design based on recorded sound. Beyond the key concepts of animation and reenactment, this dissertation has explored further notions that have a great potential for future investigation such as *action*, *congruency*, *metaphor*, and *playing technique*. The following paragraphs summarize the most important findings and the perspectives they may open up.

Re-reenactment

The concept of reenactment as it is elaborated here extends and concretizes existing notions of *re-enactment*. By *re-enactment* several authors (see for ex. Godøy 2006; Fadiga et al. 2009) have referred to the simulation of action by our nervous system that is an essential aspect of our perception and cognition. While this simulation to a large extent shares cognitive resources with the performances and imagination of action, re-enactment essentially describes an internal process in which we still and silently re-enact the actions that we perceive and cogitate on.

Instead of being limited to the still and silent simulation of action, the notion of reenactment proposed in this dissertation also embraces externalized action. The example of listening shows how our perception of sound is not only connected to the change of physiological and emotional states like our heart rate, breathing, and mood, but also shows how we support our listening via bodily movements to better localize sound sources, change perspective, and further explore our sonic environment. In addition, listening may entrain further bodily actions like tapping, rocking, and dancing. Perception and action here merge into a complex cognitive process that includes arousal and affect, variation of perspective, exploration, and entrainment.

The extended idea of reenactment supposes that within this cognitive process, all of our actions and perceptions are connected to an entire network of actions that we have experienced before. The notion of reenactment precisely allows for highlighting a particular connection within this network. In this context, actions may refer to – or *reenact* – concrete *actions* that we have experienced before, but also to more abstract notions of action, such as *events*, *techniques* or *practices*. For the design of interactive audio applications based on recorded sounds, reenactment specifically describes the relationship between the actions of the player or listener interacting with the designed application and the actions that are evoked by the recorded sound materials (see *2.2*). This idea has been the principal object of investigation of this dissertation. However, the concept applies also in other contexts.

When applying this idea of reenactment to a musical performer, instead of *performing* or *interpreting* a particular piece – which would refer to a composition or score –, his or her actions could be seen as *reenacting* particular aspects of action associated with the piece. More precisely, the musician's actions could reenact actions that are encoded in the composition (e.g. breathing, falling) or the act of composing itself. Moreover, the reenactment might refer to particular musical or social practices associated with the piece and its historical epoch such, as ceremony or dance. This example suggests how the concept of reenactment might be generalized. Instead of referring to recorded actions, it refers to actions that are implied in a musical composition or actions emanating from its context. Further examples are given in chapter *3*.

Action-Action Relationships

The idea of reenactment – and with it this whole dissertation – relies on the idea of *action*. But *what is* action and how is it articulated by the notion of reenactment?

The relationship between the actions of a violinist and resulting sound are commonly referred to as *action-sound* – or *movement-sound* – relationships (see for ex. Jensenius 2007). However, this seems to suggest that the sound produced is not part of the violinist's action. This is not satisfying, since sound is actually what makes the violinist's action an action – that of making sound.[1] On the other hand, it makes sense to consider what the violinist is doing with the violin independently of the sound it produces, or to consider the sound emitted by the violin independently of its player. This separation becomes even more evident in regard to interactive digital applications that allow for loose relationships between what the user does and the generation of sound, attributed to the application.[2]

This dilemma disappears when accepting the idea that an action is always an interaction[3] and that sound is a form of interaction itself (see 2.2.2). This leads to the idea that complex interactions can be arbitrarily decomposed into any number of elementary interactions. Finally, it is this act of decomposing interactions from a particular point of view that defines the implied actors. When referring to *action-sound* relationships, one implicitly decomposes the violinist's actions of making sound into two elements: the interaction between violinist and violin, on the one hand, and the violin's action of producing sound on the other.[4] In fact, when watching and listening to a violin performance, both the violinist's interaction *with* the violin – the piece of wood with four strings caressed with the hairs of a horse – and the violinist's actions *through* the violin are present. The composition and decomposition of interactions into interactions can be seen as a basic *action-action* relationship. Actions here are related to each other by composition.

Finally, reenactment describes another such action-action relationship. Insofar as *reenacting* can be seen as *referring* to action through action (see 2.2.5), it should be emphasized that this action is *ongoing* and *experienced*. This notion of ongoing and experienced action is essential to studying and reasoning on action. Representations of action such as recordings and motion graphs, but also models, algorithms, and programs only become ongoing actions – and understandable as such – through their playback, execution, or interpretation.[5] From this

[1]The problem is not solved by referring to a *movement-sound* relationship, since it degrades the interaction between the violinist and the violin to a movement, which does not seem right either.

[2]However, it should be insisted that, especially in the design of digital music instruments, it is important to consider the action of making sound as a whole.

[3]Even though, action is always interaction, we generally refer to *action* when it is not necessary to define who or what exactly is interacted with.

[4]The violin's action of producing sound is an interaction with the listener.

[5]Abstract notions of actions like *gesture*, *technique*, *practice*, and *event* have to be used with caution in this context, since they can refer to action in manifold ways. *Gesture*, for example, can describe an ongoing action.

perspective, action is acknowledged as an intersubjective experience that can be shared and communicated among actors through reenactments.[6] In this sense, action always consists in an *action-action relationship*. A relationship of experienced action with action that is evoked within this experience and, finally, *constitutes* this experience.

Interaction Techniques

Interaction design is concerned with the anticipation of ongoing and experienced action. After all, the principal occupation of the interaction designer is to imagine and anticipate the users' interactions with an application. It is in this sense that *playing techniques* and *playing scenarios* have been proposed as possible formalizations of potential action and action possibilities, similar to the notion of affordance (see in *2.2.7*). As affordance, the notion of playing technique contributes to a transactional approach to the investigation of interaction (see *2.2.5*). Instead of trying to define the involved actors and their interactions, it aims at observing, understanding, and anticipating how interaction emerges in a given environment.

In the virtually unlimited space of action possibilities, the notion of *playing technique* allows for further qualifying and distinguishing specific clusters of potential action. These clusters can then be put into relationship to specific qualities of the involved actors, their capabilities, and their cultural context. Similar to the notion of scenario, the notion of playing techniques – or more generally *interaction* techniques – fills the gap between *affordance* and *experience* (i.e. as in *experience centered* design). It enables the anticipation and study of *how* users actually interact with designed applications and *what* makes them interact in a certain way.

Metaphor

The idea of reenactment and action-action relationships is closely related to the notion of *metaphor*. Studies on metaphor have shown that metaphor is a fundamental element in making sense of our interactions with the world beyond language. Although in language, metaphor is usually defined as the understanding of one *thing* in terms of another, it applies equally to action. In addition, metaphors in language often refer to action (see *2.2.6*).

In the design of computer interfaces, metaphor was for a long time misunderstood and frowned upon as the graphical simulation of everyday and work environments. But the comprehension of metaphor has fundamentally evolved in the context of tangible, touch, and motion-based interfaces. In this context, instead of objects and environments, metaphor directly applies

However, the term is often used to refer to symbolic abstractions of action that tend to loose their connection to actual actions. While this abstraction may be intended when regarding gestures in communication, these symbolic abstractions are less useful in the study of interactions in other contexts.

[6] Leman (2008) advances the idea of *second person* descriptions that '*reflect involvement ... in a context of intersubjective communication.*'

to ongoing and experienced action. The *pinch* gesture for interacting with graphical objects through touch-based interfaces, for example, is, beyond the linguistic metaphor of the word describing the interaction, a metaphorical action referring to *expanding* and *compressing*.

The descriptions of interactive sound installations in *2.2.4* and playful pedagogical applications in *3.1* have particularly insisted on the *metaphoricity* of the listeners' and players' interactions. Through the interactions provided by these applications, the musical phrases of a canon, for example, are understood in terms of the moves in a chess game, and the regular pizzicatos of an orchestra accompaniment in terms of the throws and catches of a ball game. With the applications described in *3.2*, *playing* itself, finally becomes a metaphor that allows for exploring manifold relationships between music, games, and audiovisual reproductions.

The notion of metaphor is relatively unexplored in the realm of nonverbal actions that are not related to communication. However, it seems to be an important key to the understanding of interactions and ought to be further investigated in the context of interaction design along with other semiotic relationships that specifically inform the design of digital applications. In particular, the study of sound-related action and action-related sound opens a wide playground for studies that may also invite aesthetic and artistic contributions.

Interactive Digital Applications as Enactive Essays

The reenactment of action through recording and playback has become an omnipresent ingredient in our lives. The relatively recent development of interactive digital technologies enables us to create novel reenactments that are not limited to our vision and audition, but invite further parts of our cognition into this process.

Digital technologies allow for freely operating with bodily action in the same way that text allows for operating with words and, finally, with knowledge and thought. In this sense, digital applications have been presented as a form of *essay*. These essays invite their users into explorations and reflections through nonverbal multimodal interactions in which the users experience themselves as actors (see *2.2.5*). This *enactive* reflection is based on digital reenactments that are further based on digital representations of action (i.e. models, algorithms, programs). Interaction design here becomes the act of reflecting on both *on* and *through* action. This reflection operates with enactive knowledge and addresses, beyond vision and audition, the users' motor system and proprioception.

Given the explosion of interactive artworks, digital games, mobile applications, and digitally augmented objects over the past two decades, the revolution of enactive reflection based on digital interactions has already been underway for some time. However, the study of the semiotics, hermeneutics, and epistemology of interaction design as well as a clear consciousness of the implications of interactive digital authorship is still at its beginning. The concepts developed in this dissertation will hopefully contribute to theoretical and practical advances.

Agency and Congruency

The question '*who is actually acting here?*' is an important aspect of our perception and cognition. This concerns not only the identification of actors in our environment, but also the understanding of ourselves as actors through the observation of our own actions. It is with this understanding that *agency*[7] and the contribution of action to self-recognition have been investigated in neuroscience and psychology over the past decades.[8]

The exploration of action and agency is an important aspect of many interactive digital applications and a chief concern of interaction design more generally (see *2.1.6*). The discussion of interactions with sound reproduction technologies in *2.2.3* examined the notion of *congruency* as an important factor in the perception of our own actions and the attribution of perceived actions to other actors. From this perspective, interaction design has been defined as the act of creating congruencies between a multitude of actions and actors implied in the design of interactive applications (see *2.2.3*). Beyond the innumerable elements that compose an application from the designer's point of view, the actors that actually *appear* within the interaction scenario may still be numerous. Apart from the users interacting with the application,[9] they include physical objects that are manipulated by the users (e.g. balls) and digital actors appearing through sound, visuals, or motorized mechanical (i.e. robotic) elements.

Even though the argumentation of this dissertation has focused on reenactment, the questions of *how*, and *under which circumstances*, congruencies contribute to the perception of agency, certainly merits further exploration. From the perspective of action-action relationships, these questions can be reformulated as the questions under which circumstances two actions are perceived independently and when they are integrated into a single one (see above). These investigations in sound and interaction design can begin from the already considerable corpus of existing studies in neuroscience. Moreover, it seems that the knowhow acquired in the design of interactive applications, and especially audio applications, can contribute to this research through interdisciplinary projects that, again, might also include practice-based research on aesthetic and artistic questions.

[7]The argumentation of this dissertation has avoided operating with the concept of *agency* insofar as it could reintroduce an idea of self-acting actors (see *2.1.6*). However, here it is reintroduced in its correct sense.

[8]Some of these studies have concerned the investigation of neurological and psychological diseases like schizophrenia and Huntington's disease (see for ex. Repp and Knoblich 2007).

[9]In the phrasing '*users interacting with the application*', the application appears as an actor. However, from the users point of view, the application may disappear behind other material and digital actors the users engages into interaction with.

Design Perspectives

Further perspectives emerging from the investigations of this dissertation and the practice it relates to, directly concern the development of future design projects. Three axes of development are summarized in the following.

Interactive Music Recordings

The applications described in *3.1.1* (*Playing Bach*) and *3.1.3* (*The Party Guitar*) support interaction scenarios in which the players reinterpret recorded music. The *MubuFunkScatShare* performance presented in *3.1.2* can also be counted among these scenarios. An import aspect of the pedagogical projects presented in *3.1.1* is that the music students are involved in the design of these applications. Nevertheless, they can also been seen as novel ways of engaging with music that are situated somewhere half way between listening to music and performing music. Even if the audio recordings used in these projects are anonymous, these kinds of applications have a great potential for collaborating with musicians – performers, but also composers – on *interactive music recordings*.

In fact, the production of such interactive music applications strongly resembles the production of commercially distributed music recordings from many points of view. This concerns especially the techniques and competences employed in the different phases of production including recording, editing and mixing, and post-production. Only the interaction design adds a new component and new competences to the production process. As a result, instead of just a single interpretation that the listener can repeatedly play back, the music is delivered as an interactive media which the listener can *play with*.

In popular music, distribution formats and web-based applications that allowed listeners to remix songs have already existed since the 1990s (e.g. the *RMF* file format released by the *Headspace* company or the *MXP4* format). More recently, many interactive music games have been released on game and mobile platforms (e.g. Björk's *Biophilia* mobile phone application and the game *Rez* released by *Sega*). For classical music, interactive museum installations like in the *Haus der Musik* in Vienna, for example, invite visitors to make the experience of conducting an orchestra by controlling the speed of audiovisual recordings through conducting-like movements.

Within the past few years, motion capture and audio processing technologies as well as mobile and web platforms have reached a level of maturity that affords the development of a new generation of such applications. The fact that these applications are no longer screen-based, but controlled through tangible interfaces or free body movements – within collaborative interaction scenarios – introduces new elements in this context. Another advancement lies in the combination of recordings with musical models that allow for musically meaningful and consistent transformations of the recorded materials (see for ex. Fabiani 2011).

The prototypes that we have developed over the past few years and that are partially presented in chapter *3* show how recordings can be the basis of stimulating playing scenarios even when using rather simple motion capture and audio rendering techniques. The applications can be understood from the point of view of *assisted* music playing, where part of the musical action is generated by the application, while the players focus on particular aspects of musical interaction (e.g. control of tempo and/or dynamics). Other applications – or the same applications from a different perspective –, support the navigation or exploration of musical materials that represent different parts, interpretations, and versions of a given piece.

A key element in the design of these interaction scenarios lies in the bodily engagement of the players in music playing. The idea of reenactment and animation provides a conceptual ground for the design of such interactions.

Adaptive Reenactments or Learning Instruments

Some of our research and development projects have been driven by the idea of creating interactive audio applications which *adapt* to the player's movements in relationship to sound. The *Chaconne* scenario described in *3.1.1*, based on a technique known as *gesture follower*, is a relatively early realization of this idea (see *2.2.7*).[10]

The overall scenario of these applications has not changed since the beginnings of our explorations several years ago. As a first step, the player performs movements while listening to a sound recording. From the congruencies between the recorded motion and sound, the application derives a model of the implied relationships between movement and sound. The model ultimately allows the player to perform with sound through variations of the originally recorded movements. This process can be seen as the application learning specific playing techniques intended by the player, especially when additional tangible objects are involved, but also with free bodily movements.

The actions that the player associates with the recorded sound materials can be understood as gestural reenactments of these sounds. They are '*actions that could have produced*' the recorded sounds or '*actions through which the player would like to generate*' these sounds. They implicitly inform the application of the player's intentions and skills of making sound as well as the constrains of the player's physiology, involved tangible objects, and the motion capture interface. Congruencies between particular characteristics of movement and particular characteristics of sound thus inform the application that the player would like to control those characteristics of sound through those characteristics of movement. Incongruencies indicate

[10] In fact, it is the development of the *gesture follower* that has lead us to this kind of applications and finally inspired the idea of instruments that can learn relationships between movements and sound recordings.

that the application has to complete certain aspects of sound that are not controlled by the player. The player would expect that this completion is not only coherent with the aspects of sound that he or she controls, but also with the ongoing temporal evolution of sound.

The challenges in the technical implementation of this idea lie in defining what exactly is meant by each part of the above description.

- How to find the congruencies between movement and sound?
- What are the characteristics of movement and sound that are involved in this process?
- How does the application complete the characteristics of sound that are not controlled by the player?
- What does it mean that this completion is coherent with the ongoing temporal evolution of sound?

All of these questions are strongly tied to the projection, analysis, and synthesis of action and sound in respect to perceptual characteristics. They are, finally, essential questions that concern the sound and interaction design with recorded sound materials more generally, beyond the idea of learning instruments.

Research and Therapy

A third perspective that should be mentioned in this context concerns research related to the understanding and therapy of psychological, neuronal, and physiological diseases and disorders. Especially music therapy, additionally may be concerned with social disorders and social integration. In this context, sound and music may be used for the sonification of movement, enhancing motor control and proprioception, as a means of expression, and as references to the patients' environment.

The potential which arises from the work with recorded sound materials and through the perspective of reenactment, concerns the introduction of – explicit or implicit – references to sound-related interactions. As in other applications, these references may concern phenomena (i.e. nature, mechanisms, utterings), events, and practices such as music, dance, sports, and games. Through recordings of the patients' actions, these references may point to the patients themselves. Existing recordings might introduce references to their natural, social, and cultural environment. Beyond the notion of agency already mentioned above, objects worthy of exploration in this context include motor control and sensory substitution.

Bibliography

Ament, V. T. (2012). *The Foley Grail: The Art of Performing Sound for Film, Games, and Animation*. Taylor & Francis.

Aristotle and Smith, J. A. (2006). *On the Soul*. Digireads.com.

Armstrong, N. (2006). *An Enactive Approach to Digital Musical Instrument Design*. PhD thesis, Princeton University.

Assayag, G. and Bloch, G. (2007). Navigating the Oracle: A Heuristic Approach. In *Proceedings of the International Computer Music Conference*, pages 405–412, Copenhagen, Denmark.

Assayag, G., Bloch, G., and Chemillier, M. (2006a). OMax-Ofon. In *Proceedings of the 3rd International Conference on Sound and Music Computing*, Marseille, France.

Assayag, G., Bloch, G., Chemillier, M., Cont, A., and Dubnov, S. (2006b). Omax Brothers: A Dynamic Topology of Agents for Improvization Learning. In *Workshop on Audio and Music Computing for Multimedia, ACM Multimedia 2006*, Santa Barbara, USA.

Assayag, G. and Dubnov, S. (2004). Using Factor Oracles for machine Improvisation. *Soft Computing*, 8-9:604–610.

Atelier des Feuillantines (2009). Students performing the 18th Goldberg Variation by J. S. Bach. Documentation Movie. http://vimeo.com/7310159.

Atelier des Feuillantines (2010). Students performing the 2nd movement of the Concerto in f-major by J. S. Bach accompanying thier teather through a ball game. Documentation Movie. http://vimeo.com/12948099.

Atelier des Feuillantines (2013). Presentation of the activities of the *Atelier des Feuillantines*. Website. http://www.feuillantines.com/.

Baird, D. (2004). *Thing Knowledge: A Philosophy of Scientific Instruments*. University of California Press.

Ballas, J. A. (1993). Common Factors in the Identification of an Assortment of Brief Everyday Sounds. *Journal of Experimental Psychology: Human Perception and Performance*, 19(6).

Barad, K. (1998). Getting Real: Technoscientific Practices and the Materialization of Reality. *Differences: A Journal of Feminist Cultural Studies*, 10(2).

Barthes, R. (1982). *L'obvie et l'obtus : Essais critiques III*. Seuil.

Battier, M. (2007). What the GRM Brought to Music: From Musique Concrète to Acousmatic Music. *Organized Sound*, 12(3):189–202.

Begault, D. R. (1994). *3-D Sound for Virtual Reality and Multimedia*. Academic Press Professional, Inc., San Diego, CA, USA.

Bergson, H. (1908). *Essai sur les données immédiates de la conscience*. F. Alcan.

Bergson, H. and Pogson, F. L. (1910). *Time and Free Will, An Essay on the Immediate Data of Consciousness*. S. Sonnenschein & Co, New York, The Macmillan Company.

Berkhout, A. J., de Vries, D., and Vogel, P. (1993). Acoustic Control by Wave Field Synthesis. *Journal of the Acoustical Society of America*, 93(5):2764–2778.

Bertelson, P. and Radeau, M. (1981). Cross-Modal Bias and Perceptual Fusion with Auditory-Visual Spatial Discordance. *Perception & Psychophysics*, 29(6):578–584.

Berwick, C. (2006). Forty Harmonious Voices Drown Out Your Woes: Cardiff at MoMA. *Bloomberg.com*.

Bevilacqua, F., Guédy, F., Fléty, E., Leroy, N., and Schnell, N. (2007). Wireless Sensor Interface and Gesture-Follower for Music Pedagogy. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, New York, USA.

Bevilacqua, F. and Müller, R. (2005). A Gesture Follower for Performing Arts. In *Proceedings of the 6th International Gesture Workshop*, Ile de Berder, France.

Bevilacqua, F., Müller, R., and Schnell, N. (2005). MnM: A Max/MSP Mapping Toolbox. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 85–88, Vancouver, Canada.

Bevilacqua, F., Rasamimanana, N., Fléty, E., Lemouton, S., and Baschet, F. (2006). The Augmented Violin Project: Research, Composition and Performance Report. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 402–406, Paris, France.

Bevilacqua, F., Schnell, N., Rasamimanana, N., Bloit, J., Flety, E., Caramiaux, B., Françoise, J., and Boyer, E. (2013). De-Mo: Designing Action-Sound Relationships with the MO Interfaces. In *CHI '13 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '13, pages 2907–2910, New York, USA. ACM.

Bevilacqua, F., Schnell, N., Rasamimanana, N., Zamborlin, B., and Guédy, F. (2011). Online Gesture Analysis and Control of Audio Processing. In Solis, J. and Ng, K., editors, *Musical Robots and Interactive Multimodal Systems*, volume 74 of *Springer Tracts in Advanced Robotics*, pages 127–142. Springer Berlin Heidelberg.

Bevilacqua, F., Zamborlin, B., Sypniewski, A., Schnell, N., Guédy, F., and Rasamimanana, N. (2009). Continuous Realtime Gesture Following and Recognition. In *Gesture in Embodied Communication and Human-Computer Interaction: Lecture Notes in Computer Science (LNCS)*. Springer Verlag.

Bianco, T. (2012). *Gestural Control in Trumpet Performance: Experiments and Modeling*. PhD thesis, Université Pierre et Marie Curie (UPMC).

Bianco, T., Freour, V., Rasamimanana, N., Bevilaqua, F., and Caussé, R. (2010). On Gestural Variation and Coarticulation Effects in Sound Control. In *Proceedings of the 8th International Conference on Gesture in Embodied Communication and Human-Computer Interaction*, GW'09, pages 134–145, Berlin, Heidelberg. Springer-Verlag.

Black, M. (1954). Metaphor. *Proceedings of the Aristotelian Society, New Series*, 55:273–294.

Black, M. (1979). More on Metaphor. *Metaphor and Thought*.

Blackwell, A. (2006). The Reification of Metaphor as a Design Tool. *ACM Transactions on Computer-Human Interaction*, 13(4):490–530.

Blauert, J. (1997). *Spatial Hearing - Revised Edition: The Psychophysics of Human Sound Localization*. Mit Press.

Blume, E. (2008). *Bernhard Leitner – Berlin, Hamburger Bahnhof*. Number 11 in Museum für Gegenwart. Nationalgalerie, Staatliche Museen zu Berlin, DuMont, Köln.

Bodden, M. and Igsleder, H. (2002). Active Sound Design: Vacuum Cleaner. *Proceedings of Forum Acusticum, Special Issue of Revista de Acustica*, 33.

Bolia, R. S., D'Angelo, W. R., and McKinley, R. L. (1999). Aurally Aided Visual Search in Three-Dimensional Space. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 41(4):664–669.

Bolivar, V., Cohen, A. J., and Fentress, J. C. (1994). Semantic and Formal Congruency in Music and Motion Pictures: Effects on the Interpretation of Visual Action. *Psychomusicology*, 13(1):28–59.

Bouënard, A., Gibet, S., and Wanderley, M. (2008). Enhancing the Visualization of Percussion Gestures by Virtual Character Animation. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 38–43, Genova, Italy.

Boyden, D. D. (1965). *The History of Violin Playing, from Its Origins to 1761 and Its Relationship to the Violin and Violin Music*. Oxford University Press.

Bregman, A. S. (1994). *Auditory Scene Analysis: The Perceptual Organization of Sound*. Mit Press.

Bruner, J. S. (1966). On Cognitive Growth. In Bruner, J. S., Olver, R. R., and et. al., G. P. M., editors, *Studies in Cognitive Growth: A Collaboration at the Center of Cognitive Studies*, pages 30–67. Wiley, New York.

Burke, K. (1969). *A Grammar of Motives*. University of California Press.

Buxton, B. and Buxton, W. (2007). *Sketching User Experiences: Getting the Design Right and the Right Design*. Elsevier/Morgan Kaufmann.

Cadoz, C., Luciani, A., Florens, J.-L., and Castagné, N. (2003). ACROE - ICA: Artistic Creation and Computer Interactive Multisensory Simulation Force Feedback Gesture Transducers. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 235–246, Montreal, Quebec, Canada.

Calvert, G. A., Spence, C., and Stein, B. E. (2004). *The Handbook of Multisensory Processes*. Mit Press.

Camurri, A. and Volpe, G. (2004). *Gesture-Based Communication in Human-Computer Interaction: 5th International Gesture Workshop, Selected Revised Papers*. Springer, Genova.

Cap Digital (2013). Presentation of the *Futur en Seine* festival. Website. `http://www.futur-en-seine.fr/`.

Caramiaux, B. (2009). Similarités Morphologiques entre Geste et Son et leur Détection Temporelle. In *JJCAAS (Journées des Jeunes Chercheurs en Acoustique, Auditon et Signal audio)*, Marseille, France.

Caramiaux, B. (2012). *Studies on the Gesture – Sound Relationship in Musical Performance*. PhD thesis, Université Paris VI, Ircam – Centre Pompidou.

Caramiaux, B., Bevilacqua, F., and Schnell, N. (2009). Towards a Gesture-Sound Cross-Modal Analysis. In *Gesture in Embodied Communication and Human-Computer Interaction, Lecture Notes in Computer Science (LNCS)*. Springer Verlag.

Caramiaux, B., Bevilacqua, F., and Schnell, N. (2010a). Analysing Gesture and Sound Similarities with a HMM-based Divergence Measure. In *Proceedings of the 7th International Conference on Sound and Music Computing*, Barcelona, Spain.

Caramiaux, B., Bevilacqua, F., and Schnell, N. (2010b). Study on Gesture-Sound Similarity. In *3rd Music and Gesture Conference*, Montreal, Canada.

Caramiaux, B. and Schnell, N. (2009). Towards an Analysis Tool for Gesture/Sound Mapping. In *Proceedings of the 8th International Gesture Workshop*, Bielefeld, Germany.

Caramiaux, B., Susini, P., Bianco, T., Bevilacqua, F., Houix, O., Schnell, N., and Misdariis, N. (2011). Gestural Embodiment of Environmental Sounds: An Experimental Study. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 144–148, Oslo, Norway.

Carroll, J. M. (1995). *Scenario-based Design: Envisioning Work and Technology in System Development*. Wiley.

Casey, M. (2003). Sound Replacement, Beat Unmixing and Audio Mosacing: Creative Uses of MPEG-7 Audio. In *Proceedings of the 6th International Conference on Digital Audio Effects (Tutorial Session)*, London, UK.

Celermajer, D. (2006). Seeing the Light and Hearing the Call: The Aesthetics of Knowledge and Thought. *Journal of Literature and Aesthetics*, 16(2).

Chadabe, J. (2002). The Limitations of Mapping as a Structural Descriptive in Electronic Instruments. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, NIME '02, pages 1–5, Singapore. National University of Singapore.

Chen, J. L., Penhune, V. B., and Zatorre, R. J. (2008). Listening to Musical Rhythms Recruits Motor Regions of the Brain. *Cerebral Cortex*, 18:2844–2854.

Chion, M. (1983). *Guide des objets sonores: Pierre Schaeffer et la recherche musicale*. Buchet/Chastel, English translation by John Dack and Christine North.

Cholodenko, A. (1993). *The Illusion of Life: Essays on Animation*. Indiana University Press.

Cholodenko, A. (2007). *The Illusion of Life II: More Essays on Animation*. Power Publications.

Cienki, A. and Müller, C. (2008). *Metaphor and Gesture*. John Benjamins Publishing Company.

Clandinin, D. J. (1985). Personal Practical Knowledge: A Study of Teachers' Classroom Images. *Curriculum Inquiry*, 15(4):361–385.

Clarke, E. F. (2005). *Ways of Listening: An Ecological Approach to the Perception of Musical Meaning*. Oxford University Press.

Clayton, M., Sager, R., and Will, U. (2004). In Time with the Music: The Concept of Entrainment and its Significance for Ethnomusicology. *ESEM CounterPoint*, 1.

Collins, N. (2010). *Introduction to Computer Music*. John Wiley & Sons.

Connerton, P. (1989). *How Societies Remember*. Cambridge University Press.

Cook, N. (1998). *Analysing Musical Multimedia*. Clarendon Press.

Cooke, D. (1959). *The Language of Music*. Oxford University Press on Demand.

Cooper, A., Reimann, R., and Cronin, D. (2012). *About Face 3: The Essentials of Interaction Design*. Wiley.

Coppola, P. (1944). *Dix-sept ans de musique à Paris, 1922-1939*. Slatkine.

Cycling'74 (2013). Max/MSP Product Presentation. http://cycling74.com/products/max/.

Cyr, M. (1992). *Performing Baroque Music*. Scholar Press.

Damasio, A. R. (1994). *Descartes' Error: Emotion, Reason and the Human Brain*. G. P. Putnam.

Davide Rocchesso, F. F. (2003). *The Sounding Object*. Mondo Estremo, Italy.

de Bruijn, W. (2004). *Application of Wave Field Synthesis in Videoconferencing*. PhD thesis, TU Delft.

Deegan, M. A. (1941). *Contributions to the Technique of Violin Playing Made by Nicolo Paganini*. PhD thesis, University of Rochester.

Dell'Olio, P. (2009). *Violin Bow Construction and its Influence on Bowing Technique in the Eighteenth and Nineteenth Centuries*. PhD thesis, Florida State University.

Demers, J. and Coombe, R. (2006). *Steal This Music: How Intellectual Property Law Affects Musical Creativity*. University of Georgia Press.

Dewey, J. and Bentley, A. F. (1949). *Knowing and the Known*. Beacon Press.

Dickreiter, M., Dittel, V., Hoeg, W., and Wöhr, M. (2008). *Handbuch der Tonstudiotechnik*. De Gruyter.

Dohrn-van Rossum, G. (1996). *History of the Hour: Clocks and Modern Temporal Orders*. University of Chicago Press.

Donin, N. (2004). Le travail de la répétition. Deux dispositifs d'écoute et deux époques de la reproductibilité musicale, du premier au second après-guerre. *Circuit. Musiques contemporaines*, 14-1:53–85.

Dourish, P. (2004). *Where the Action Is: The Foundations of Embodied Interaction*. MIT Press.

Dubnov, S. and Assayag, G. (2002). Universal Prediction Applied to Stylistic Music Generation. *Mathematics and Music. A Diderot Mathematical Forum*.

Eckel, G. (2001). The Vision of the LISTEN Project. In *Proceedings of the 7th International Conference on Virtual Systems and Multimedia (VSMM'01)*, VSMM '01, pages 393–396, Washington, DC, USA. IEEE Computer Society.

Eco, U. (1986). *Semiotics and the Philosophy of Language*. Indiana University Press.

Eitan, Z. and Granot, R. (2006). How Music Moves. *Music Perception*, 23(3):221–248.

Essl, G. and O'Modhrain, S. (2006). An Enactive Approach to the Design of New Tangible Musical Instruments. *Organized Sound*, 11(3):285–296.

Fabiani, M. (2011). *Interactive Computer-Aided Expressive Music Performance – Analysis, Control, Modification, and Synthesis Methods*. PhD thesis, KTH School of Computer Science, Stockholm, Sweden.

Fadiga, L., Craighero, L., Buccino, G., and Rizzolatti, G. (2002). Speech Listening Specifically Modulates the Excitability of Tongue Muscles: a TMS study. *European Journal of Neuroscience*, 15:399–402.

Fadiga, L., Laila, C., and D'Ausilio, A. (2009). Broca's Area in Language, Action, and Music. *Annals of the New York Academy of Sciences*, 1169:448—58.

Farnell, A. J. (2007a). Marching Onwards: Procedural Synthetic Footsteps for Video Games and Animation. In *Proceedings of the Pure Data Convention*, Montreal, Canada.

Farnell, A. J. (2007b). Synthetic Game audio with PureData. In *Proceedings of the Audio Mostly Conference*, Ilmenau, Germany.

Fastl, H. and Zwicker, E. (2007). *Psychoacoustics: Facts and Models*. Springer.

Fingerhut, M. (2004). Music Information Retrieval, or How to Search for (and Maybe Find) Music and Do Away with Incipits. In *IAML-IASA 2004 Congress*, Oslo, Norvège.

Flach, R., Knoblich, G., and Prinz, W. (2004). Recognizing One's Own Clapping: The Role of Temporal Cues in Self-Recognition. *Psychological Research, 69*, 69:147–156.

Fléty, E., Leroy, N., Ravarini, J.-C., and Bevilacqua, F. (2004). Versatile Sensor Acquisition System Utilizing Network Technology. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, Hamamatsu, Japan.

Fog, C. L. and Pedersen, T. H. (1999). Tools for Product Optimisation. In *Proceedings of the Conference Human Centered Processes*, Brest, France.

Françoise, J. (2012). A Hierarchical Approach for the Design of Gesture-to-Sound Mappings. In *Proceedings of the 9th International Conference on Sound and Music Computing*, Copenhagen, Denmark.

Françoise, J., Lallemand, I., Artières, T., Bevilacqua, F., Schnell, N., and Schwarz, D. (2013a). Perspectives pour l'apprentissage interactif du couplage geste-son. In *Journées d'Informatique Musicale*, Paris, France.

Françoise, J., Schnell, N., and Bevilacqua, F. (2013b). Gesture–Sound Mapping by Demonstration in Interactive Music Systems. In *Proceedings of the 21st ACM International Conference on Multimedia*, Barcelona, Spain.

Franinović, K. and Serafin, S. (2013). *Sonic Interaction Design*. Mit Press.

Fraser, J. T. (1990). *Of Time, Passion, and Knowledge: Reflections on the Strategy of Existence*. Princeton University Press.

Friberg, A., Bresin, R., and Sundberg, J. (2006). Overview of the KTH Rule System for Musical Performance. *Advances in Cognitive Psychology*, 2(2):145–161.

Gallup, G. G. (1970). Chimpanzees: Self Recognition. *Science*, 167:86–87.

Gann, K. (2006). *The Music of Conlon Nancarrow*. Cambridge University Press.

Gaver, W. (1989). The SonicFinder: An Interface That Uses Auditory Icons. *Human-Computer Interaction*, 4(1):67–94.

Gaver, W. W. (1986). Auditory Icons: Using Sound in Computer Interfaces. *Human-Computer Interaction*, 2(2):167–177.

Gaver, W. W. (1991). Technology Affordances. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '91, pages 79–84, New York, NY, USA. ACM.

Gaver, W. W. (2009). Designing for Homo Ludens, Still. In Binder, T., Löwgren, J., and Malmborg, L., editors, *(Re)searching the Digital Bauhaus*, pages 163–178. Springer London.

Gelatt, R. (1955). *The Fabulous Phonograph: From Tin Foil to High Fidelity*. Lippincott.

Gervain, J., Macagno, F., Cogoi, S., Peña, M., and Mehler, J. (2008). The neonate brain detects speech structure. *Proceedings of the National Academy of Sciences of the United States of America*, 105(37):14222–14227.

Gerzon, M. (1985). Ambisonics in Multichannel Broadcasting and Video. *Journal of the Audio Engineering Society*, 33(11):859–871.

Ghirardelli, T. G. and Scharine, A. A. (2009). Auditory-Visual Interactions. In E. Rash, C., editor, *Helmet-mounted Displays: Sensation, Perception, and Cognition Issues*. U.S. Army Aeromedical Research Laboratory.

Gibet, S., Courty, N., and Kamp, J.-F. (2006). *Gesture in Human-Computer Interaction and Simulation: 6th International Gesture Workshop, Revised Selected Papers*. Springer, Berder Island, France.

Gibson, J. J. (1966). *The Senses Considered as Perceptual Systems*. Houghton Mifflin, Boston.

Gibson, J. J. (1977). The Theory of Affordances. *Perceiving, Acting, and Knowing: Toward an Ecological Psychology*.

Gibson, J. J. (1979). *The Ecological Approach To Visual Perception*. Houghton Mifflin.

Giordano, B. and McAdams, S. (2006). Material Identification of Real Impact Sounds: Effects of Size Variation in Steel, Glass, Wood, and Plexiglass Plates. *The Journal of the Acoustical Society of America*, 119(2):1171–1181.

Godøy, R. I. (2006). Gestural-Sonorous Objects: Embodied Extensions of Schaeffer's Conceptual Apparatus. *Organised Sound*, 11(2):149–157.

Godøy, R. I. (2009a). Geometry and Effort in Gestural Renderings of Musical Sound. In Sales Dias, M., Gibet, S., Wanderley, M. M., and Bastos, R., editors, *Gesture-Based Human-Computer Interaction and Simulation*, pages 205–215, Berlin, Heidelberg. Springer-Verlag.

Godøy, R. I. (2009b). Gestural Affordances of Musical Sound. In Godoy, R. I. and Leman, M., editors, *Musical Gestures: Sound, Movement, and Meaning*. Routledge, New York, USA.

Godøy, R. I., Haga, E., and Jensenius, A. R. (2005). Playing "Air Instruments": Mimicry of Sound-Producing Gestures by Novices and Experts. In Gibet, S., Courty, N., and Kamp, J.-F., editors, *Gesture in Human-Computer Interaction and Simulation: 6th International Gesture Workshop*, volume 3881/2006, pages 256–267, Berder Island, France. Springer-Verlag.

Godøy, R. I., Haga, E., and Jensenius, A. R. (2006). Exploring Music-Related Gestures by Sound-Tracing – A Preliminary Study. In *2nd International Symposium on Gesture Interfaces for Multimedia Systems (GIMS2006)*, Leeds, UK.

Godøy, R. I., Jensenius, A. R., and Nymoen, K. (2009). Coarticulation of sound and movement in music. In *Proceedings of ESCOM 2009: 7th Triennial Conference of the European Society for the Cognitive Sciences of Music*, Jyväskylä, Finland.

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

Godøy, R. I. and Leman, M., editors (2010). *Musical Gestures: Sound, Movement, and Meaning*. Routledge, New York.

Gondan, M., Niederhaus, B., Rösler, F., and Röder, B. (2005). Multisensory Processing in the Redundant-Target Effect: A Behavioral and Event-Related Potential Study. *Perception & Psychophysics*, 67(4):713–726.

Gould, G. (1966). The Prospects of Recording. *High Fidelity*, 16(4):46–63.

Gould, G. (2004). The Prospects of Recording. In Cox, C. and Warner, D., editors, *Audio Culture: Readings in Modern Music*, Continuum Books. Bloomsbury.

Graupner, C. (2010). MindBox. Demonstration Movie. http://vimeo.com/15879203.

Graupner, C. (2011). MindBox Trailer 2. Promotional Movie. http://vimeo.com/22128501.

Graupner, C., Zappalà, R., Schnell, N., and Peters, N. (2011). MindBox. In Behrendt, F. and Lossius, T., editors, *Sonic Interaction Design – Exhibition Catalogue (Norwegian Museum of Science, Technology, and Medicine)*, pages 28–31. Bergen Center for Electronic Arts (BEK).

Greenfield, P. (1991). Language, Tools and Brain: The Ontogeny and Phylogeny of Hierarchically Organized sequential behavior. *Behavioral and Brain Sciences*, 14:531–551.

Grond, F., Hermann, T., Verfaille, V., and Wanderley, M. M. (2010). Methods for Effective Sonification of Clarinetists' Ancillary Gestures. In *Proceedings of the 8th International Conference on Gesture in Embodied Communication and Human-Computer Interaction*, GW'09, pages 171–181, Berlin, Heidelberg. Springer-Verlag.

Gruman, G. and Hattersley, M. (2009). *Mac OS X Snow Leopard Bible*. Wiley.

Guédy, F., Bevilacqua, F., Fober, D., and Robert, L. (2013). Etudier et jouer une oeuvre musicale en manipulant des objets. In *Actes des Journées d'Informatique Musicale JIM2013*, Paris, France.

Guédy, F., Bevilacqua, F., Rasamimanana, N., Schnell, N., Fléty, E., Delabie, C., and Rigolot, F. (2011). L'expérimentation musicale et pédagogique sur le geste de l'Atelier des Feuillantines avec l'équipe Interaction Musicale Temps Réel de l'Ircam. Online Article *l'éducation musicale*. http://www.leducation-musicale.com/ircam.pdf.

Gusev, X. (2009). Akustische Technologien und Klangkunst: Anwendungen, Intentionen, Wirkungen. Master's thesis, Technische Universität Berlin, Fachgebiet Audiokommunikation.

Haga, E. (2008). *Correspondences between Music and Body Movement*. PhD thesis, Faculty of Humanities University of Oslo.

Hansen, K. F. (2010). *The Acoustics and Performance of DJ Scratching. Analysis and modeling*. PhD thesis, Kungl Tekniska Högskolan.

Haspels, J. J. (1987). *Automatic Musical Instruments: Their Mechanics and Their Music, 1580 - 1820*. Nirota, Muziekdruk C.V.

Haueisen, J. and Knösche, T. R. (2001). Involuntary Motor Activity in Pianists Evoked by Music Perception. *Journal of Cognitive Neuroscience*, 13(6):786–792.

Hayles, N. K. (2008). *How We Became Posthuman: Virtual Bodies in Cybernetics, Literature, and Informatics*. University of Chicago Press.

Heidegger, M. (1967). *Sein und Zeit*. Max Niemeyer Verlag Tübingen, 11th edition.

Heidegger, M. and Stambaugh, J. (1996). *Sein und Zeit*. State University of New York Press.

Henle, P. (1981). Metaphor. *Philosophical Perspectives on Metaphor*.

Hermann, M. and Maciej, W., editors (2012). *Erweiterte Spieltechniken in der Musik von Helmut Lachenmann / Extended Techniques in the Music of Helmut Lachenmann*. Breitkopf & Härtel.

Hoeg, W., Steffen, F., and Steinke, G. (1983). Delta Stereophony: A sound system with true direction and distance perception for large multipurpose halls. *Journal of the Audio Engineering Society*, 31:500–511.

Houkes, W. and Vermaas, P. E. (2010). *Technical Functions: On the Use and Design of Artefacts*. Springer Netherlands.

Howard, I. P. and Templeton, W. B. (1966). *Human spatial orientation*. Wiley.

Huizinga, J. (1949). *Homo Ludens*. Routledge & Kegan Paul Limited.

Hume, D. (1757). *Four Dissertations: I. The Natural History of Religion.* A. Millar.

Huron, D. B. (2006). *Sweet Anticipation: Music And the Psychology of Expectation*. Mit Press.

Husserl, E. (1913). *Ideen zu einer reinen Phänomenologie und phänomenologischen Philosophie – Allgemeine Einführung in die reine Phänomenologie*. M. Niemeyer.

Ihde, D. (1990). *Technology and the Lifeworld: From Garden to Earth*. Indiana University Press.

Interlude Project Consortium (2011). MO – Musical Objects. Demonstration Movie. http://www.youtube.com/watch?v=Uhps_U2E9OM.

Isaacson, W. (2007). *Einstein: His Life and Universe*. Simon & Schuster.

Jehan, T. (2005). *Creating Music by Listening*. PhD thesis, Massachusetts Institute of Technology, MA, USA.

Jensenius, A. R. (2007). *Action–Sound: Developing Methods and Tools to Study Music-Related Body Movement*. PhD thesis, University of Oslo.

Jensenius, A. R. (2012). Motion-sound Interaction Using Sonification based on Motiongrams. In *Proceedings of the 2012 International Conference on Advances in Computer-Human Interactions*, pages 170–175, Valencia, Spain.

Johnson, M. (1981). *Philosophical Perspectives on Metaphor*. University of Minnesota Press.

Johnson, M. (1989). Embodied Knowledge. *Curriculum Inquiry*, 19(4):361–377.

Jordà, S. (2005). *Digital Lutherie – Crafting Musical Computers for New Musics' Performance and Improvisation*. PhD thesis, Universitat Pompeu Fabra, Barcelona, Spain.

Juntunen, M.-L. (2004). *Embodiment in Dalcroze Eurhythmics*. PhD thesis, University of Oulu, Finland.

Kepler, J., von Dyck, W., and Caspar, M. (1937). *Johannes Kepler Gesammelte Werke*. C.H. Beck'sche Verlagsbuchhandlung, München.

Kerins, M. (2011). *Beyond Dolby (Stereo): Cinema in the Digital Sound Age*. Indiana University Press.

Kersten, S. and Purwins, P. (2010). Sound Texture Synthesis with Hidden Markov Tree Models in the Wavelet Domain. In *Proceedings of the 7th International Conference on Sound and Music Computing*, Barcelona, Spain.

Kersten, S. and Purwins, P. (2012). Fire Texture Sound Re-Synthesis Using Sparse Decomposition and Noise Modelling. In *Proceedings of the 15th International Conference on Digital Audio Effects*, York, UK.

Kientzy, D. (1990). *Saxologie : du potentiel acoustico-expressif des 7 saxophones*. PhD thesis, Université de Paris VIII.

Kilgore, R., Chignell, M., and Smith, P. (2003). Spatialized Audioconferencing: What are the Benefits? In *Proceedings of the 2003 Conference of the Centre for Advanced Studies on Collaborative Research*, CASCON '03, pages 135–144, Toronto, Canada. IBM Press.

Kimura, M. (1995). "Subharmonics:" An extended technique for the violin. *The Journal of the Acoustical Society of America*, 97(5):3270–3270.

Kittler, F. A. (1986). *Grammophon, Film, Typewriter*. Brinkmann & Bose.

Knoblich, G. and Flach, R. (2003). Action Identity: Evidence from Self-Recognition, Prediction, and Coordination. *Consciousness and Cognition*, 12(4):620–632.

Koetsier, T. (1997). *On the Prehistory of Programmable Machines: Musical Automata, Looms, Calculators*. Vrije Universiteit.

Krumhansl, C. L. (1997). An Exploratory Study of Musical Emotions and Psychophysiology. *Canadian Journal of Experimental Psychology – Revue canadienne de psychologie experimentale*, 51(4):336–353.

Kuo, S. M. and Morgan, D. R. (1999). Active Noise Control: A Tutorial Review. *Proceedings of the IEEE*, 87(6):943–973.

Kvifte, T. and Jensen, S. (2007). *Instruments and the Electronic Age*. Lulu.com.

Kvifte, T. and Jensenius, A. R. (2006). Towards a Coherent Terminology and Model of Instrument Description and Design. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, NIME '06, pages 220–225, Paris, France. IRCAM - Centre Pompidou.

Lahav, A., Boulanger, A., Schlaug, G., and Saltzman, E. (2005). The Power of Listening: Auditory-Motor Interactions in Musical Training. *Annals of the New York Academy of Sciences*, 1060(1):189–194.

Lahav, A., Saltzman, E., and Schlaug, G. (2007). Action Representation of Sound: Audiomotor Recognition Network While Listening to Newly Acquired Actions. *The Journal of Neuroscience*, 27(2):308–314.

Lakoff, G. (1987). The Death of Dead Metaphor. *Metaphor and Symbolic Activity*, pages 143—147.

Lakoff, G. and Johnson, M. (1980). *Metaphors We Live By*. University of Chicago Press.

Lakoff, G. and Turner, M. (1989). *More Than Cool Reason: A Field Guide to Poetic Metaphor*. University of Chicago Press.

Latour, B. (1987). *Science in Action: How to Follow Scientists and Engineers Through Society*. Harward University Press.

Latour, B. (1996). On Actor-Network Theory - A Few Clarifications. *Soziale Welt*, 4(47):369–381.

Latour, B. (2005). *Reassembling the Social: An Introduction to Actor-Network-Theory*. Oxford University Press, USA.

Lazier, A. and Cook, P. (2003). MoSievius: Feature Driven Interactive Audio Mosaicing. In *Proceedings of the 6th International Conference on Digital Audio Effects*, London, UK.

Lees, A., Asai, T., Andersen, T. B., Nunome, H., and Sterzing, T. (2010). The Biomechanics of Kicking in Soccer: A Review. *Journal of Sports Science and Medicine*, 28(8):805–17.

Leman, M. (2008). *Embodied Music Cognition and Mediation Technology*. MIT Press.

Lesiuk, T. (2005). The effect of music listening on work performance. *Psychology of Music*, 33(2):173–191.

Leslie, G., Zamborlin, B., Jodlowski, P., and Schnell, N. (2010). Grainstick: A Collaborative, Interactive Sound Installation. In *Proceedings of the International Computer Music Conference*, New York, USA.

Lessig, L. (2008). *Remix: Making Art and Commerce Thrive in the Hybrid Economy*. Penguin Press.

Lewis, G. (2006). Living with Creative Machines: A Composer Reflects. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 414–415, Paris, France. IRCAM - Centre Pompidou.

Lewis, G. E. (2000). Too Many Notes: Computers, Complexity and Culture in Voyager. *Leonardo Music Journal*, 10(1):33–39.

Liberman, A. (1985). The Motor Theory of Speech Perception Revised. *Cognition*, 21(1):1–36.

Lindemann, E. (2007). Music Synthesis with Reconstructive Phrase Modeling. *IEEE Signal Processing Magazine*, 24(2):80–91.

Löwgren, J. (2006). Articulating the Use Qualities of Digital Designs. *Aesthetic Computing*, pages 383–403.

Löwgren, J. (2009). Towards an Articulation of Interaction Aesthetics. *New Review of Hypermedia and Multimedia*, 15(2).

Ludden, G. D. S., Kudrowitz, B., Schifferstein, H. N. J., and Hekkert, P. (2012). Surprise and Humor in Product Design. Designing Sensory Metaphors in Multiple Modalities. *Humor*, 25(3):285 – 309.

Machover, T. (1992). Hyperinstruments: A Progress Report, 1987-1991. Technical report, Massachusetts Institute of Technology, Media Laboratory.

MacKenzie, I. S. and Buxton, W. (1992). Extending Fitts' Law to Two-Dimensional Tasks. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '92, pages 219–226, New York, USA. ACM.

Maggiore, A. and Meneghetti, U. (2011). Antique Applications of Cam Mechanisms. In *13th World Congress in Mechanism and Machine Science*, Guanajuato, México.

Magnusson, T. (2009). *Epistemic Tools – The Phenomenology of Digital Musical Instruments*. PhD thesis, University of Sussex.

Magnusson, T. (2010). Designing Constraints: Composing and Performing with Digital Musical Systems. *Computer Music Journal*, 34(4):62–73.

Malloch, J. and Wanderley, M. M. (2007). The T-Stick: from musical interface to musical instrument. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, NIME '07, pages 66–70, New York, USA. ACM.

Marrin Nakra, T. (2000). *Inside the "Conductor's Jacket": Analysis, Interpretation and Musical Synthesis of Expressive Gesture*. PhD thesis, Massachusetts Institute of Technology.

Maturana, H. R. and Varela, F. J. (1980). *Autopoiesis and Cognition: The Realization of the Living*. D. Reidel Publishing Company.

McDermott, J. H. and Simoncelli, E. P. (2011). Sound Texture Perception via Statistics of the Auditory Periphery: Evidence from Sound Synthesis. *Neuron*, 71(5):926–940.

McGurck, H. and MacDonald, J. W. (1976). Hearing Lips and Seeing Voices. *Nature*, 264(246–248).

McNeill, D. (2008). *Gesture and Thought*. University of Chicago Press.

Merleau-Ponty, M. (1945). *Phénomènologie de la perception*. Gallimard, Paris.

Merleau-Ponty, M. (1962). *Phenomenology of Perception*. Routledge & Kegan Paul.

Merleau-Ponty, M. and Deprun, J. (1997). *L'union de l'âme et du corps chez Malebranche, Biran et Bergson: notes prises au cours de Maurice Merleau-Ponty à l'École Normale Supérieure (1947-1948)*. Librairie philosophique, J. Vrin.

Meyer, J. (2004). *Akustik und musikalische Aufführungspraxis: Leitfaden für Akustiker, Tonmeister, Musiker, Instrumentenbauer und Architekten*. Verlag Erwin Bochinsky.

Miller, G. A. (1956). The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information. *Psychological Review*, 63:81—97.

Milliman, R. E. (1982). Using Background Music to Affect the Behavior of Supermarket Shoppers. *Journal of Marketing*, 46(3):86–91.

Mills III, J. A., Astrinaki, M., Chessini Bose, R., Rebousière, L., and Todoroff, T. (2010). Augmented Conductor. In Dutoit, T. and Macq, B., editors, *QPSR of the numediart Research Program*, volume 3, pages 71–76. numediart Research Program on Digital Art Technologies.

Minard, R. (1993). *Sound Environments / Klangwelten*. Akademie der Künste, Berlin.

Miranda, E. R. and Wanderley, M. M. (2006). *New Digital Musical Instruments: Control And Interaction Beyond the Keyboard*. A-R Editions.

Molnar-Szakacs and Overy (2006). Music and Mirror Neurons: From Motion to 'E'motion. *Social Cognitive and Affective Neuroscience*, 1(3):235.

Morein-Zamir, S., Soto-Faraco, S., and Kingstone, A. (2003). Auditory Capture of Vision: Examining Temporal Ventriloquism. *Cognitive Brain Research*, 17(1):154–163.

Morton, P. (2010). *A Historical Introduction to the Philosophy of Mind (2nd Edition: Readings with Commentary)*. Broadview Press.

Möser, M. (2009). *Technische Akustik*. Springer.

Müller, C. (2009). *Metaphors Dead and Alive, Sleeping and Waking: A Dynamic View*. University of Chicago Press.

Müller, M. (2007). *Information Retrieval for Music and Motion*. Springer Verlag.

Naveda, L. (2011). *Gesture in Samba: A Cross-Modal Analysis of Dance and Music from the Afro-Brazilian Culture*. PhD thesis, Ghent University.

Naveda, L. A. and Leman, M. (2008). Sonification of Samba dance using periodic pattern analysis. In Barbosa, A., editor, *Proceedings of the 4th International Conference on Digital Arts (ARTECH 2008)*, pages 16–26, Porto, Portugal. Portuguese Católica University.

Neubarth, K., Bergeron, M., and Conklin, D. (2011). Associations between Musicology and Music Information Retrieval. In *Proceedings of the 12th International Society for Music Information Retrieval Conference*, pages 429–434, Miami, USA.

Nielsen, M. and Dissanayake, C. (2004). Pretend Play, Mirror Self-Recognition and Imitation: A Longitudinal Investigation Through the Second Year. *Infant Behavior and Development*, 3:342–365.

Nierhaus, G. (2009). *Algorithmic Composition: Paradigms of Automated Music Generation*. Mathematics and Statistics. Springer.

NIME Steering Committee (2013). International Conference on New Interfaces for Musical Expression. http://www.nime.org/.

NoDesign (2011a). Documentation of the *Urban Musical Game* basketball game. Documentation Movie. https://vimeo.com/26413625.

NoDesign (2011b). Presentation of an *Urban Musical Game* prototype at the Centquatre (104) in Paris. Documentation Movie. https://vimeo.com/22120867.

Noë, A. (2005). *Action in Perception*. MIT Press.

Norman, D. A. (1999). Affordance, Conventions, and Design. *interactions*, 6(3):38–43.

Norman, D. A. (2002). *The Design of Everyday Things*. Basic Books.

Nymoen, K., Caramiaux, B., Kozak, M., and Torresen, J. (2011). Analyzing Sound Tracings: A Multimodal Approach to Music Information Retrieval. In *Proceedings of the 1st International ACM Workshop on Music Information Retrieval with User-Centered and Multimodal Strategies*, MIRUM '11, pages 39–44, New York, USA. ACM.

Nymoen, K., Godøy, R. I., Torresen, J., and Jensenius, A. R. (2012). A Statistical Approach to Analyzing Sound Tracings. In Ystad, S., Aramaki, M., Kronland-Martinet, R., Jensen, K., Mohanty, S., Nymoen, K., Torresen, J., Godøy, R., and Jensenius, A., editors, *Speech, Sound and Music Processing: Embracing Research in India*, volume 7172 of *Lecture Notes in Computer Science*, pages 120–145. Springer Berlin Heidelberg.

Ord-Hume, A. W. J. G. (1973). *Clockwork music: An Illustrated History of Mechanical Musical Instruments from the Musical Box to the Pianola, from Automaton Lady Virginal Players to Orchestrion*. Crown Publishers.

Pachet, F. (2002). Multimedia at Work - Playing with Virtual Musicians: The Continuator in Practice. *MultiMedia, IEEE*, 9(3):77–82.

Pachet, F. (2004). Beyond the Cybernetic Jam Fantasy: The Continuator. *Computer Graphics and Applications, IEEE*, 24(1):31–35.

Pachet, F., Roy, P., Moreira, J., and d'Inverno, M. (2013). Reflexive Loopers for Solo Musical Improvization. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '13*, pages 2205–2208, Paris, France.

Parviainen, J. (2002). Bodily Knowledge: Epistemological Reflections on Dance. *Dance Research Journal*, 34(1):11–26.

Patel, A. (2006). Musical Rhythm, Linguistic Rhythm, and Human Evolution. *Music Perception*, 24(1):99–104.

Patel, A., Iversen, J., Bregman, M., and Schulz, I. (2009a). Experimental Evidence for Synchronization to a Musical Beat in a Nonhuman Animal. *Current Biology*, 19(10):827–830.

Patel, A. D. (2010). *Music, Language, and the Brain*. Oxford University Press, USA.

Patel, A. D., Hallam, S., Cross, I., and Thaut, M. (2009b). Music and the Brain: Three Links to Language. In Hallam, S., Cross, I., and Thaut, M., editors, *The Oxford Handbook of Music Psychology*, pages 208–216. Oxford University Press.

Peeters, G. (2001). *Modèles et modélisation du signal sonore adaptes à ses caractéristiques locales*. PhD thesis, Université Paris VI.

Peeters, G. (2004). A Large Set of Audio Features for Sound Description (Similarity and Classification) in the CUIDADO Project. Technical report, IRCAM – Centre Pompidou.

Peirce, C. S. and Moore, E. C. (1998). *Charles S. Peirce: The Essential Writings*. Prometheus Books.

Petts, L. and Andrews, F. (1983). *The Story of Nipper and the His Master's Voice Picture Painted by Francis Barraud*. Talking machine review.

Pieraccini, R. (2012). *The Voice in the Machine: Building Computers that Understand Speech*. MIT Press.

Pokerdance, Ipodnano05, et al. (2009). Party in the U.S.A. Wikipedia Article. http://en.wikipedia.org/wiki/Party_in_the_U.S.A.

Polansky, R. M. (2007). *Aristotle's De Anima*. Cambridge University Press.

Polanyi, M. (1966). *The Tacit Dimension*. Peter Smith.

Prinz, W. (1997). Perception and Action Planning. *European Journal of Cognitive Psychology*, 9(2):129–154.

Queneau, R. (1963). *Exercices de style*. Gallimard, Paris, France.

Rasamimanana, N. (2008). *Geste instrumental du violoniste en situation de jeu : analyse et modélisation*. PhD thesis, Université Pierre et Marie Curie.

Rasamimanana, N., Bernardin, D., Wanderley, M. M., and Bevilacqua, F. (2007). String Bowing Gestures at Varying Bow Stroke Frequencies: A Case Study. In *Advances in Gesture-Based Human-Computer Interaction and Simulation*. Springer Verlag.

Rasamimanana, N. and Bevilacqua, F. (2008). Effort-based Analysis of Bowing Movements: Evidence of Anticipation Effects. *Journal of New Music Research*, 37-4:339–351.

Rasamimanana, N., Bevilacqua, F., Bloit, J., Schnell, N., Fléty, E., Cera, A., Petrevski, U., and Frechin, J.-L. (2012). The Urban Musical Game: Using Sport Balls as Musical Interfaces. In *CHI '12 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '12, pages 1027–1030, New York, USA. ACM.

Rasamimanana, N., Bevilacqua, F., Schnell, N., Guédy, F., Flety, E., Maestracci, C., Zamborlin, B., Frechin, J.-L., and Petrevski, U. (2011). Modular Musical Objects Towards Embodied Control Of Digital Music. In *5th International Conference on Tangible, Embedded, and Embodied Interaction*, Funchal, Portugal.

Rasamimanana, N., Bloit, J., and Schnell, N. (2010). Mo Kitchen. Demonstration Movie. https://www.youtube.com/watch?v=v7_cHlsQaGw.

Rasamimanana, N., Fléty, E., and Bevilacqua, F. (2006). Gesture Analysis of Violin Bow Strokes. In Sylvie Gibet, Nicolas Courty, J.-F. K., editor, *Lecture Notes in Computer Science, Vol. 3881: Gesture in Human-Computer Interaction and Simulation: 6th International*

Gesture Workshop, GW 2005, Berder Island, France, May 18-20, 2005, Revised Selected Papers, pages 145–155. Springer Berlin / Heidelberg.

Rasamimanana, N., Kaiser, F., and Bevilacqua, F. (2009). Perspectives on Gesture-Sound Relationships Informed from Acoustic Instrument Studies. *Organised Sound*, 14-2:208–216.

Reighley, K. B. (2000). *Looking for the Perfect Beat: The Art and Culture of the Dj*. Pocket Books.

Repp, B. H. and Knoblich, G. (2007). Toward a Psychophysics of Agency: Detecting Gain and Loss of Control over Auditory Action Effects. *Journal of Experimental Psychology: Human Perception and Performance*, 33(2):469–482.

Riskin, J. (2003). Eighteenth-Century Wetware. *Representations*, 83(1):97–125.

Riva, G., Davide, F., and IJsselsteijn, W. W. A. (2003). *Being There: Concepts, Effects and Measurements of User Presence in Synthetic Environments*. IOS Press, Ohmsha.

Roads, C. (1996). *The Computer Music Tutorial*. Mit Press.

Rocchesso, D. (2011). *Explorations in Sonic Interaction Design*. Logos Verlag Berlin.

Rocchesso, D., Serafin, S., Behrendt, F., Bernardini, N., Bresin, R., Eckel, G., Franinovic, K., Hermann, T., Pauletto, S., Susini, P., and Visell, Y. (2008). Sonic Interaction Design: Sound, Information and Experience. In *CHI '08 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '08, pages 3969–3972, New York, USA. ACM.

Rodger, M. W., O'Modhrain, S., and Craig, C. M. (2013). Temporal Guidance of Musicians' Performance Movement is an Acquired Skill. *Experimental Brain Research*, 226(2):221–230.

Roebel, A. (2003). A New Approach to Transient Processing in the Phase Vocoder. In *Proceedings of the 6th International Conference on Digital Audio Effects*, pages 344–349, London, United Kingdom.

Roebel, A. (2010). Shape-Invariant Speech Transformation with the Phase Vocoder. In *InterSpeech*, Makuhari, Japan.

Rosenbaum, D. A. (2009). *Human Motor Control*. Elsevier Science.

Sachs, J. (2005). Aristotle: Motion and its Place in Nature. http://www.iep.utm.edu/aris-mot/.

Salter, C. L., Baalman, M. A. J., and Moody-Grigsby, D. (2008). Between Mapping, Sonification and Composition: Responsive Audio Environments in Live Performance. In Kronland-Martinet, R., Ystad, S., and Jensen, K., editors, *Computer Music Modeling and Retrieval. Sense of Sounds*, volume 4969 of *Lecture Notes in Computer Science*, pages 246–262. Springer Berlin Heidelberg.

Schaeffer, P. (1966). *Traité des objets musicaux : essai interdisciplines*. Éditions du Seuil.

Schafer, R. M. (1977). *The Tuning of the World*. Knopf.

Schafer, R. M. (1994). *The Soundscape: Our Sonic Environment and the Tuning of the World*. Destiny Books.

Schellberg, D. (1994). *Didgeridoo: Ritual Original and Playing Techniques*. Red Wheel/Weiser.

Schnell, N. (2011). Real-Time Audio Mosaicing. Webpage. http://imtr.ircam.fr/imtr/Real-Time_Audio_Mosaicing.

Schnell, N., Bevilacqua, F., Guédy, F., and Rasamimanana, N. (2011a). Playing and Replaying – Sound, Gesture and Music Analysis and Re-Synthesis for the Interactive Control and Re-Embodiment of Recorded Music. In von Loesch, H. and Weinzierl, S., editors, *Klang und Begriff: Gemessene Interpretation - Computergestützte Aufführungsanalyse im Kreuzverhör der Disziplinen*. Schott Verlag, Mainz, Germany.

Schnell, N., Bevilacqua, F., Rasamimanana, N., Blois, J., Guédy, F., and Fléty, E. (2011b). Playing the "MO" – Gestural Control and Re-Embodiment of Recorded Sound and Music. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 535–536, Oslo, Norway.

Schnell, N., Borghesi, R., Schwarz, D., Bevilacqua, F., and Müller, R. (2005). FTM – Complex Data Structures for Max. In *Proceedings of the International Computer Music Conference*, Barcelona, Spain.

Schnell, N., Peeters, G., Lemouton, S., Manoury, P., and Rodet, X. (2000). Synthesizing a Choir in Real-Time Using Pitch Synchronous Overlap Add (PSOLA). In *Proceedings of the International Computer Music Conference*, Berlin, Allemagne.

Schnell, N., Röbel, A., Schwarz, D., Peeters, G., and Borghesi, R. (2009). MuBu & Friends - Assembling Tools for Content Based Real-Time Interactive Audio Processing in Max/MSP. In *Proceedings of the International Computer Music Conference*, Montreal, Canada.

Schnell, N. and Schwarz, D. (2005). Gabor, Multi-Representation Real-Time Analysis/Synthesis. In *Proceedings of the 8th International Conference on Digital Audio Effects*, pages 122–126, Madrid, Spain.

Schoonderwaldt, E. (2009). *Mechanics and Acoustics of Violin Bowing - Freedom, Constraints and Control in Performance*. PhD thesis, KTH Computer Science and Communication.

Sherrington, C. S. (1907). On the Proprioceptive System, Especially in Its Reflex Aspect. *Brain*, 29(4):467–482.

Small, C. (1998). *Musicking: The Meanings of Performing and Listening*. University Press of New England.

Smith, M., Salvendy, G., Abe, S., Nozawa, T., and Kondo, T. (2009). *Lecture Notes in Computer Science*, volume 5617, pages 459–465. Springer Berlin Heidelberg.

Stacey, J. and Suchman, L. (2012). Animation and Automation –The Liveliness and Labours of Bodies and Machines. *Body & Society*, 18(1):1–46.

Stanton, N. A. and Young, M. (1999). *Guide to Methodology in Ergonomics: Designing for Human Use*. Taylor & Francis.

Sterne, J. (2003). *The Audible Past: Cultural Origins of Sound Reproduction*. Duke University Press.

Sterne, J. (2006). The Death and Life of Digital Audio. *Interdisciplinary Science Reviews*, 31(4):338–348.

Stowell, D., Robertson, A., Bryan-Kinns, N., and Plumbley, M. D. (2009). Evaluation of Live Human-Computer Music-Making: Quantitative and Qualitative Approaches. *International Journal of Human-Computer Studies*, 67(11):960–975.

Stowell, R. (1990). *Violin Technique and Performance Practice in the Late Eighteenth and Early Nineteenth Centuries*. Cambridge University Press.

Strack, F., Martin, L. L., and Stepper, S. (1988). Inhibiting and Facilitating Conditions of the Human Smile: A Nonobtrusive Test of the facial Feedback Hypothesis. *Journal of Personality and Social Psychology*, 54(5):768–777.

Strobl, G., Eckel, G., and Rocchesso, D. (2006). Sound Texture Modeling: A Survey. In *Proceedings of the 3rd International Conference on Sound and Music Computing*, pages 61–65, Marseille, France.

Suchman, L. A. (2007). *Human-Machine Reconfigurations: Plans and Situated Actions*. Cambridge University Press.

Supper, M. (2001). A Few Remarks on Algorithmic Composition. *Computer Music Journal*, 25(1):48–53.

Szeliski, R. (2010). *Computer Vision: Algorithms and Applications*. Springer London.

Tanaka, A. (2000). Musical Performance Practice on Sensor-Based Instruments. In Wanderley, M. M. and Battier, M., editors, *Trends in Gestural Control of Music*. (CD-ROM), IRCAM - Centre Pompidou, Paris.

Tanaka, A. (2010). Mapping out Instruments, Affordances, and Mobiles. In of Technology Sydney, U., editor, *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 88–93, Sydney, Australia.

Tanaka, A., Altavilla, A., and Spowage, N. (2011). Gestural Musical Affordances. In *Proceedings of the 8th International Conference on Sound and Music Computing*, Padova, Italy.

Tanaka, A. and Bongers, B. (2001). Global String: A Musical Instrument for Hybrid Space. In *Proceedings of Cast01: Living in Mixed Realities*, pages 177–181, Sankt Augustin, Germany.

Tanaka, A., Caramiaux, B., and Schnell, N. (2013). MubuFunkScatShare: Gestural Energy and Shared Interactive Music. In *CHI '13 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '13, pages 2999–3002, New York, USA. ACM.

TEDx Panthéon Sorbonne (2012). Atau Tanaka performing MubuFunkScatShare at the TEDx Panthéon Sorbonne. Movie. http://www.youtube.com/watch?v=6G2H-LUHrtw.

Todoroff, T., Ben-Madhkour, R., Binon, D., Bose, R., and Paesmans, V. (2010). FireTraSe: Stereoscopic Camera Tracking and Wireless Wearable Sensors System for Interactive Dance Performances - Application to "Fire Experiences and Projections". In Dutoit, T. and Macq, B., editors, *QPSR of the numediart research program*, volume 3, pages 9–24. numediart Research Program on Digital Art Technologies.

Tomayko-Peters, A. (2006). Maestro Frankenstein. http://arvidtp.net/sw/maestro_frankenstein.php.

Usher, A. P. (1988). *A History of Mechanical Inventions*. Dover.

van der Heide, E. (1998). A World Beyond the Loudspeaker. http://www.evdh.net/awbl/.

Varela, F. J., Thompson, E. T., and Rosch, E. (1992). *The Embodied Mind: Cognitive Science and Human Experience*. MIT Press.

Vercoe, B. (1984). The Synthetic Performer in the Context of Live Performance. In *Proceedings of the International Computer Music Conference*, pages 199–200, Paris, France.

Vergez, C. and Tisserand, P. (2006). The BRASS Project, from Physical Models to Virtual Musical Instruments: Playability Issues. In Kronland-Martinet, R., Voinier, T., and Ystad, S., editors, *Computer Music Modeling and Retrieval*, volume 3902 of *Lecture Notes in Computer Science*, pages 24–33. Springer Berlin Heidelberg.

Voegelin, E. (2000). Order and History. In Sandoz, E., editor, *The Collected Works of Eric Voegelin*, volume 18 (5). University of Missouri Press.

von Loesch, H. and Weinzierl, S., editors (2011). *Klang und Begriff: Gemessene Interpretation - Computergestützte Aufführungsanalyse im Kreuzverhör der Disziplinen*. Schott Verlag, Mainz.

VoxLer (2013). Presentation of the activities of *VoxLer*. Website. http://www.voxler.fr/.

Walton, K. L. (1978). Fearing Fictions. *The Journal of Philosophy*, 75(1):5–27.

Wanderley, M. M. and Orio, N. (2002). Evaluation of Input Devices for Musical Expression: Borrowing Tools from HCI. *Computer Music Journal*, 26(3):62–76.

Wanderley, M. M., Vines, B. W., Middleton, N., McKay, C., and Hatch, W. (2005). The Musical Significance of Clarinetists' Ancillary Gestures: An Exploration of the Field. *Journal of New Music Research*, 34(1).

Warner, T. (2003). *Pop Music: Technology and Creativity : Trevor Horn and the Digital Revolution*. Ashgate.

Warusfel, O. and Eckel, G. (2004). LISTEN – Augmenting Everyday Environments Through Interactive Soundscapes. In *Virtual Reality for Public Consumption, IEEE Virtual Reality 2004 Workshop*, Chicago, USA.

Wegener, F. (2005). *The History of Basketball and Volleyball*. GRIN Verlag.

Weinzierl, S. (2008). *Handbuch der Audiotechnik*. Springer.

Weinzierl, S. and Franke, C. (2002). 'Lotte, ein Schwindel!' – Geschichte und Praxis des Musikschnitts am Beispiel von Beethovens 9. Symphonie. In *Proceedings of the VDT International Audio Convention (22. Tonmeistertagung)*, Hannover, Germany.

Weiss, A. S. (2002). *Breathless: Sound Recording, Disembodiment, and the Transformation of Lyrical Nostalgia*. Wesleyan University Press.

Welch, W. L. and Burt, L. B. S. (1994). *From Tinfoil to Stereo: The Acoustic Years of the Recording Industry, 1877-1929*. University Press of Florida.

Wessel, D. (2006). An Enactive Approach to Computer Music Performance. In Orlarey, Y., editor, *Le Feedback dans la Creation Musical*. Studio Gramme, Lyon, France.

White, M. (1996). *The High Fidelity Turntable System and the Creation of Hip Hop Music: An Organological Inquiry*. University of Washington.

Wiener, N. (1961). *Cybernetics or Control and Communication in the Animal and the Machine*. M.I.T. Press.

Wills, D. (1995). *Prosthesis*. Stanford University Press.

Wilson, S. M., Saygin, A. P., Sereno, M. I., and Iacoboni, M. (2004). Listening to Speech Activates Motor Areas Involved in Speech Production. *Nat Neurosci*, 7(7):701–702.

Wishart, T. (1996). *On Sonic Art*. Harwood Academic Publishers.

Wittgenstein, L. (1965). *The Blue and Brown Books*. Harper Collins.

Wright, P. C. and McCarthy, J. (2010). *Experience-Centered Design: Designers, Users, and Communities in Dialogue*. Morgan & Claypool Publishers.

Zamborlin, B. (2011). Mogees. Webpage. http://www.brunozamborlin.com/mogees/.

Zatorre, R. J., Chen, J. L., and Penhune, V. B. (2007). When the Brain Plays Music: Auditory-Motor Interactions in Music Perception and Production. *Nature Reviews Neuroscience*, 8(7):547–558.

Zbikowski, L. (2008). Metaphor and Music. In Gibbs Jr., R. W., editor, *The Cambridge Handbook of Metaphor and Thought*, pages 502—524. Cambridge University Press.

Zhu, J. and Thagard, P. (2002). Emotion and Action. *Philosophical Psychology*, 15(1):19–36.

Ziethen, K.-H. and Serena, A. (2003). *Virtuosos of Juggling: From the Ming Dynasty to Cirque du Soieil*. Renegade Juggling.

List of Figures and Tables