Department of Computer Science

Creating and evaluating embodied interactive experiences: case studies of full-body, sonic and tactile enaction.

Roberto Pugliese

Aalto University

DOCTORAL
DISSERTATIONS

Aalto University publication series
DOCTORAL DISSERTATIONS 205/2015

Creating and evaluating embodied interactive experiences: case studies of full-body, sonic and tactile enaction.

Roberto Pugliese

A doctoral dissertation completed for the degree of Doctor of Science (Technology) to be defended, with the permission of the Aalto University School of Science, at a public examination held at the lecture hall TU1 of the school on 11 December 2015 at 12.

Aalto University
School of Science
Department of Computer Science

Supervising professor
Prof. Tapio Takala

Thesis advisor
Prof. Tapio Takala

Preliminary examiners
Prof. Roope Raisamo,
University of Tampere,
Finland

Assist. Prof. Cumhur Erkut,
Aalborg University,
Denmark

Opponents
Assoc. Prof. Zsófia Ruttkay
Moholy-Nagy University of Art and Design
Hungary

Aalto University publication series
DOCTORAL DISSERTATIONS 205/2015

ISBN 978-952-60-6553-3 (printed)
ISBN 978-952-60-6554-0 (pdf)
ISSN-L 1799-4934
ISSN 1799-4934 (printed)
ISSN 1799-4942 (pdf)
http://urn.fi/URN:ISBN:978-952-60-6554-0

Unigrafia Oy
Helsinki 2015

Finland

Aalto University

Aalto University, P.O. Box 11000, FI-00076 Aalto www.aalto.fi

Abstract

Author
Roberto Pugliese

Name of the doctoral dissertation
Creating and evaluating embodied interactive experiences: case studies of full-body, sonic and tactile enaction.

Publisher School of Science

Unit Department of Computer Science

Series Aalto University publication series DOCTORAL DISSERTATIONS 205/2015

Field of research Interactive Digital Media

Manuscript submitted 17 August 2015 **Date of the defence** 11 December 2015

Permission to publish granted (date) 14 October 2015 **Language** English

☐ **Monograph** ☒ **Article dissertation (summary + original articles)**

Abstract

This thesis contributes to the field of embodied and multimodal interaction by presenting the development of different original interactive systems. Using a constructive approach, a variety of real-time user interaction situations were designed and tested, two cases of human-virtual character bodily interaction, two interactive sonifications of trampoline jumping, collaborative interaction in mobile music performance and tangible and tactile interaction with virtual sounds.

While diverse in terms of application, all the explored interaction techniques belong to the context of augmentation and are grounded in the theory of embodiment and strategies for natural human-computer interaction (HCI). The cases have been contextualized within the umbrella of enaction, a paradigm of cognitive science that addresses the user as an embodied agent situated in an environment and coupled to it through sensorimotor activity. This activity of sensing and action is studied through different modalities: auditory, tactile and visual and combinations of these. The designed applications aim at a natural interaction with the system, being full-body, tangible and spatially aware. Particularly sonic interaction has been explored in the context of music creation, sports and auditory display. These technology-mediated scenarios are evaluated in order to understand what the adopted interaction techniques can bring to the user experience, how they modify impressions and enjoyment. The publications also discuss the enabling technologies used for the development, including motion tracking and programmed hardware for the tactile-sonic interaction and sonic and tangible interaction.

Results show that combining full-body interaction with auditory augmentation and sonic interaction can modify the perception, observed behavior and emotion during the experience. Using spatial interaction together with tangible interaction or tactile feedback provides for a multimodal experience of exploring a mixed reality environment where audio can be accessed and manipulated with natural interaction. Embodied and spatial interaction brings playfulness to a mobile music improvisation, shifting the focus of the experience from music-making towards movement-based gaming. Finally, two novel implementations of full-body interaction based on the enactive paradigm are presented. In these designed scenarios of enaction the participant is motion tracked and a virtual character rendered as a stick figure is displayed in front of her on a screen. Results from the user studies show how the involvement of the body is crucial in understanding the behavior of a virtual character or a digital representation of the self in a gaming scenario.

Keywords Embodied interaction, enaction, multimodal, sonic interaction, full-body, audio augmented reality, avatars

ISBN (printed) 978-952-60-6553-3 **ISBN (pdf)** 978-952-60-6554-0

ISSN-L 1799-4934 **ISSN (printed)** 1799-4934 **ISSN (pdf)** 1799-4942

Location of publisher Helsinki **Location of printing** Helsinki **Year** 2015

Pages 160 **urn** http://urn.fi/URN:ISBN:978-952-60-6554-0

Preface

Writing this section gives me satisfaction and relief. It arrives after five years of work, exciting at times, frustrating others. Luckily, I had the opportunity to work together with people that eased the process. I want to thank warmly Klaus and Archontis who co-authored much of my research. Sharing problems and solutions with the whole research group, Meeri, Tuukka, Jussi H., Jussi T., Jari, has helped me a lot and made the daily work much more enjoyable.

I am grateful to my supervisor Tapio "Tassu" Takala for his advice and encouragement; I particularly appreciated his flexibility and support for my artistic career which I was able to build on in parallel with the doctoral research. Although not co-authors in this dissertation, I want to mention two people I am very fond of, Mauri and Pia. Meeting and working with them was among the highlights of these years.

Even if it is natural for the unpleasant memories to blur, I clearly remember times of discomfort and research dizziness when finding the right path felt beyond my reach. In those moments Katarina brought stability and affection to the plate and motivating me to push forward. I owe her so much. Finally, as in the best tradition of fairy tale endings, the conclusion of this thesis coincided with an episode of major significance in my life, my first son Leo was born. For both serendipity as well as causality this thesis is dedicated to him. May the world he brings forth be a beautiful one.

Helsinki, November 11, 2015,

Roberto Pugliese

Contents

List of Publications

This thesis consists of an overview and of the following publications which are referred to in the text by their Roman numerals.

I Roberto Pugliese, Klaus Lehtonen. A framework for motion based bodily enaction with virtual characters. In *11th International Conference on Intelligent Virtual Agents (IVA 2011)*, Reykjavik, Iceland, Lecture Notes in Computer Science, Volume 6895, pp. 162-168, September 2011.

II Roberto Pugliese, Archontis Politis, Tapio Takala. Spatial rendering of audio-tactile feedback for exploration and object interaction in virtual environments. In *Proceedings of the 9th Sound and Music Computing Conference*, Copenhagen, Denmark, pp. 241-248, July 2012.

III Roberto Pugliese, Koray Tahiroğlu, Callum Goddard, James Nesfield. A qualitative evaluation of augmented human-human interaction in mobile group improvisation. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, University of Michigan, Ann Arbor, May 2012.

IV Luca Turchet, Roberto Pugliese, Tapio Takala. Physically based sound synthesis and control of jumping sounds on an elastic trampoline. In *Proceedings of ISon 2013, 4th Interactive Sonification Workshop*, Fraunhofer IIS, Erlangen, Germany, pp. 87-94, December 2013.

V Roberto Pugliese, Archontis Politis, Tapio Takala. ATSI: augmented and tangible sonic interaction. In *Proceedings of the Ninth Interna-*

tional Conference on Tangible, Embedded, and Embodied Interaction, Stanford, California, USA, ACM, pp. 97-104, January 2015.

VI Roberto Pugliese, Tapio Takala. Sonic trampoline: how audio feedback impacts the user's experience of jumping. *MultiMedia, IEEE*, Volume 22, no. 1, pp. 74-79, January - March 2015.

VII Roberto Pugliese, Klaus Förger, Tapio Takala. Game experience when controlling a weak avatar in full-body enaction. In *15th International Conference on Intelligent Virtual Agents (IVA 2015)*, Delft, Netherlands, Lecture Notes in Computer Science, Volume 9238, pp. 418-431, August 2015.

Author's Contribution

Publication I: "A framework for motion based bodily enaction with virtual characters"

The paper describes a novel framework and implementation of full-body interaction with a virtual character using the concept of enaction in bodily interaction. The system allows authoring rules for the behavior of a virtual character. Motion features are extracted from the participants actions and used to continuously generate appropriate full-body reactions in the virtual character presented in front of them. Through a user study based on subjective evaluation and interview we demonstrated the potential of continuous motion-based interaction with an emotional virtual character. Different traits of the character, such as being easily scared, can be discovered through the embodied action of the participant.

The work was an equal contribution between the writers. The author of this dissertation had a major role in the conceptual framework and designing the tool of the system for authoring the rules. He also contributed in conducting the experiments and writing the paper.

Publication II: "Spatial rendering of audio-tactile feedback for exploration and object interaction in virtual environments"

The paper describes an implementation of a system that integrates audio and tactile spatial feedback. The user is tracked using commodity hardware and sound is rendered through loudspeakers. We designed a wearable interface, a belt with actuators to convey tactile feedback when virtual sound objects approach and touch a user immersed in the virtual sonic environment. We created a demo to informally assess the multimodal audio-tactile display for spatial and natural interaction, and found the audio-tactile rendering technique able to render the tactile stimulus according to the sound position and with the desired intensity profile. The integration proposed is suitable for multimedia installations.

The author of this dissertation is the main contributor for the development of the system and the integration of audio and tactile rendering. The second author had a bigger part in the integration of the tracker data with the spatialization of the sound.

Publication III: "A qualitative evaluation of augmented human-human interaction in mobile group improvisation"

We presented a sonic augmentation of mobile music performance. On one hand, users can gesturally generate and control sound on their mobile phones individually. On the other hand, motion tracking is used for controlling sound parameters as a group during a music improvisation session with mobile phones, letting players continuously negotiate certain aspects of the sonic presentation created. We evaluated the experience using structured interviews plus a grounded theory approach to qualitative analysis. The results present the concepts emerging from the experience of collaborative scenarios of audio augmentation, such as the body-device relation in the music making activity, the shift-of-focus from music playing to a playful motion-based game and the contribution of familiarity with the instrument to this shift.

All the authors contributed equally to the work. The author of this thesis had a major role in the analysis of the results of the qualitative evaluation.

Publication IV: "Physically based sound synthesis and control of jumping sounds on an elastic trampoline"

A sonification of a trampoline using physically-based synthesis was proposed. The work included developing the sensing technique for interactively controlling the sound of jumping on a trampoline. The system simulated the sound of different realistic surfaces after the feet-trampoline contact. Results of a user study with subjective evaluation indicated an altered sensation of the tactile feeling of the surface when the sound of different physically-based synthesized materials are employed. The system proved robust and the interaction natural, thus suitable for VR and multimodal scenarios.

The author of this thesis carried out the experiment and provided the sensing solution used to control the physically simulated sounds.

Publication V: "ATSI: augmented and tangible sonic interaction"

This work presents a design and implementation of a system for tangible and sonic interaction with spatial sound. The design principle was to integrate tangible interaction with physical objects and spatial rendering of virtual sounds so that the user could dynamically attach sounds to real objects augmenting sonically the environment with a layer of sounds. A localization study showed that the rendering system is accurate enough for audio augmented reality applications. The interactive system contributes to the field of audio augmented reality by adding tangible interaction with sound without the use of special technologies and finds application in sonic interior design and music applications.

The author of this thesis is the main contributor of the work (implementation of the different modules and evaluation). The orientation tracking data and the binaural rendering integration were carried out by the second author.

Publication VI: "Sonic trampoline: how audio feedback impacts the user's experience of jumping"

This paper presents interactive sonification of a person jumping on a large trampoline. The technical implementation is based on audio and computer vision sensing techniques. The sonification strategy uses a parameter mapping scheme inspired by iconic game sounds. In a user study we measured the performance of users jumping on the trampoline and also asked them to assess their emotions while jumping by means of a questionnaire. Results indicated that both performance and experience can be augmented by using our system.

The author of this thesis is the main contributor of the work in all its aspects.

Publication VII: "Game experience when controlling a weak avatar in full-body enaction"

In this work we study the effect of controlling an avatar that does not mirror the player movement but lets the player control a weak body in full-body interaction. We designed a controller mechanism for motion-based games that invites players to adapt to the physical state of their avatar. In a mini-game projected on a screen, the player needs to pop bubbles controlling the body of a stick-figure avatar. Different postures and movement responses to the player movement are considered as the conditions for a user study. We show that the need to adapt to a weak avatar conveys to the player the sense of embodying a character with such features. The study is an exploration that poses the basis for future research about the process of identification with an avatar in full-body enaction.

The author of this thesis is the main contributor of the work in all its aspects.

1. Introduction

1.1 Why *interacting* ?

In a thesis where the word *interaction* will be repeated over and over again, one might pose a fundamental question :

Why do we interact ? and *What is interaction ?*

Human-computer interaction (HCI) has emerged in the early 1980's, initially as part of computer science including aspects of cognitive science and human factors [21]. The original target of HCI research was on personal productivity applications, but nowadays the field has expanded greatly outgrowing computer science and embracing aspects such as human factors, information systems, and library and information science (see [52] for an history of the HCI field and an overview of its branches).

Already in 1962 Engelbart envisioned computing as a mean for augmenting human intellect [38], that is "increasing the capability of a man to approach a complex problem situation, gain comprehension to suit his particular needs, and to derive solutions to problems". Part of human-computer interaction research has focused on simplifying work-related and discretionary tasks, reducing their complexity and making them more automatic. Interfaces are means to access functionalities that let the user accomplish these tasks. From this utilitarian perspective, interaction is a goal-oriented dialogue between the user and a computer aiming at solving well-defined problems.

Alternatively, we can think of a different kind of interaction, one at the core of a person's exploration of his environment, an interaction involved with encountering and experiencing the other; a dynamic condition that lets one discover the world and build so-called knowledge.

What is knowledge after all? Is it the ability of doing something? Is it the mental faculty of listing the steps involved in accomplishing a goal? Is it about knowing the rules of a game? Or being able to play a game?

This thesis is inspired by the relationship between a certain kind of technologically mediated interaction and exploration and sense-making.

To answer the previous question, for the purpose of this thesis interaction is defined as:

an action-driven process that lets a person discover, experience and make sense of the object of the interaction. In the case of living subjects, interaction is a dynamic coupling between the subject and the environment and among different subjects. Interaction is experience shaped by the human body.

But what about human-computer interaction?

Most of the applications presented in this thesis do not replicate existing functionalities of interactive systems based on a non- full-body interaction but rather augment activities that happen far from a computer screen. It is not about a dialogue with a computer, it is the creation of technologically mediated and possibly empowered experiences.

With these premises in mind, the body and senses of the agent will be at the very center of the case studies described. Moreover, with this focus on facilitating exploration and discovery, sound plays a central part in the designed interactions. Sound is a fluid, transient medium that invites exploration of the environment; sound conveys properties of the source originating it, being it externally caused or result of user interaction.

1.2 Scope of the thesis

This thesis is constructed as a collection of case studies. Each of them proposes an original design and an implementation of an interactive system. Interaction styles explored in this thesis are *spatial, multimodal and full-body.*

Spatial interaction (also referred as Spatial Augmented Reality [4] and [98]) stands for an interaction between user and virtual objects, in our case sounds, that happens in a 3D space where position, orientation and the relative positions of the user and the objects of interaction are meaningful factors of the media presentation.

Multimodal interaction in system design refers to ”interaction with the virtual and physical environment through natural modes of communication such as speech, body gestures, hand writing, graphics or gaze” [12]. Modes or channels are characterized by the combination of human senses (sight, hearing, touch, taste and smell) and effectors (limbs, fingers, eyes, head/face, body, vocal system). From a human-centered view, according to Charwat [22] modality is perception via one of the perception channels. I limit the scope of the explored multimodality to the senses of sight, hearing and touch and, as effectors, the limbs and the head.

Full-body interaction stresses the involvement of the whole body as channel of interaction. Rather than individual segregated body parts at work, e.g. the hand on a mouse interface or a tablet, full-body interaction relies on and requires the whole body to support the communication between the user and the system. Examples include virtual reality applications where the person is immersed in a virtual environment, motion-based games using motion-tracking technology such as Microsoft Kinect or dance performances where the dancer’s movement synthesize or transform the sound or visual elements like projections on stage.

The case studies presented can be grouped in the following categories:

1.2.1 Audio Augmented Reality and Multimodal Interaction

The physical environment is extended with virtual auditory cues and sound information. Sound can be reproduced through headphones or loudspeakers but in order to be more naturally perceived as blending with the environment, it should be heard spatially, as if it was coming from a certain position with respect to the position of the user. In addition to that, the case studies combine the audio modality with the tactile, either in the form of letting the user grasp tangible objects or augmenting the body with tactile cues of physical contacts with the virtual sounds. This category includes Publication II and Publication V .

1.2.2 Sonic Interaction

This category includes case studies of interactive sounds produced by the movements and results of the action of the user. More specifically two cases of sonification of a user jumping on a trampoline (Publication IV and Publication VI) and one case of sonic augmentation of a 3-person music improvisation session using mobile phones as instruments (Publication III) are discussed.

1.2.3 Full-body Interaction

Two cases are described: Publication I discusses an encounter with a virtual character projected in front of the participant and showing different behaviors as a result of the participant's actions; Publication VII presents a mini-game where a character, motion-controlled by the participant, portrays signs of weakness. The controller mechanism requires the participant to adapt his movement and posture to the constraint offered by the body of the avatar.

All these instances can be grouped into a framework for designing interaction, which is the embodied interaction and enaction paradigm. I will open up definitions and associated concepts for those terms in Chapter 2.

1.3 Evaluating interaction

Not a single standard approach alone for evaluating human-computer interaction could have been used throughout this thesis due to the different styles of interaction and various contexts of application explored. For one publication (PII) a formal evaluation is not provided but only a series of observations drawn from the experience of the authors. For the rest of the case studies different approaches to evaluate interaction were adopted to find the most suitable one for the context and the maturity of the development. This led the author to combine more traditional task-based user studies with free exploration sessions. Results were collected using objective measures of performance, self-reported quantities through questionnaires and even structured interviews analyzed with qualitative methods.

1.4 Research questions

The research has been focusing on the experience of novel interaction rather than the functionality of it. Many case studies presented in this work are not about goal-oriented interaction, digital tools or assistance for problem-solving. Rather, attention is given to the experience that interacting with this system brings forth without being aware of the system itself.

The goal of the studies has been to investigate the effects that the respective types of interaction have on the user experience. The experience of interaction is an elusive subject, as it tends to escape definitions but needs to be measured somehow in scientific research.

In this thesis I tried to assess the experience of interaction both from a performance point of view (how well did you perform a task) but also from a psychological point of view (how did it feel).

To summarize, three kinds of questions are posed after the realization of each interactive system. Firstly, I investigated the effect augmentation on performance, expecting that the augmented modalities might result in changes in behavior. This has been done for the publication PVI and as a localization task in PV. The effect on sensations and impressions, especially with respect to audio-tactile displays, was assessed in the evaluation of PIV and more informally in PII.

Another question arising from the fuzziness of the *experience* concept emerged: How does the interaction shape the experience to the extent that we conceptualize it differently (a game, an exercise, an encounter, a monitoring)? Publication III dealing with mobile sound performance dealt directly with the difficulty of evaluating an interactive system with predefined metrics. For that the authors used a qualitative approach based on semi-structured interviews combined with a grounded theory approach to understand how the participants conceptualized the experience of interacting with the system, individually and together with others.

Publication II to Publication VI all use sound as the main medium of interaction.

Publication I and PVII are about interaction with a virtual character and delve more directly in the concepts of human-avatar interaction, bodily behavior and identification. The purpose of these studies are to investigate how it feels to control a different body (PVII) or how an encounter with a virtual character a person can bodily interact with lets this participant recognize traits of the avatar behavior (PI).

Finally, the development of the systems shed some light on how to implement natural and multimodal interaction with off-the-shelf components and the technical challenges one might encounter. In particular, publications PII-PIV-PV-PVI deal with the integration of different sensing techniques and output displays (audio-tactile).

1.5 Thesis organization

Chapter 2 describes the theoretical foundation for the embodied and enactive paradigms used in this thesis and the implications for designing interaction. Chapter 3 presents the case studies for Interaction in audio augmented reality and the related works. In Chapter 4 I describe sonic interaction and interactive sonification, the related works and discuss our contribution to the domain. Chapter 5 presents cases of full-body interaction with virtual characters. A discussion of the technical implementation of the different systems can be found in greater detail in the individual publications, including used formulas, developed software and employed hardware. Finally, Chapter 6 draws some overall conclusions and outlines directions for further research.

2. Background

I provide in this chapter a theoretical background for the case studies implemented in this thesis. I introduce the concept of embodiment and address it as a context for the interaction discussed in the presented research. Gradually, I move towards discussing enaction as the theoretical framework informing the design of multisensory full-body experiences.

2.1 Embodied Mind

"Embodiment is the property of our engagement with the world that allows us to make it meaningful". [33, p. 126]

The concept stems from a philosophical discourse of the 20th century, the phenomenology tradition. Embodiment is a crucial notion for phenomenology, in that we do not only act through the body but that the particular shape and nature of one's physical, temporal and social immersion is what makes a meaningful experience possible [1].

Phenomenology is concerned with the structure of lived experience focusing on the link of action and perception in human sense-making. For Heidegger "being-in-the-world" is to be an agent coping with things. Our representation of the world is already the result of a certain grasp of the world that we have as agents in it [5, pp. 432-433]. For Merleau-Ponty perception and representation are structured by the embodied agent in the course of its ongoing purposeful engagement with the world [1, p. 104].

Quoting Dourish again: "Embodiment is a participative status, a way of being, rather than a physical property." [33, p. 125]

Embodied architectures for the mind have been proposed in different research disciplines. Varela et al. [119] bring the insights from phenomenology into cognitive science. The core concept is that cognition is NOT representation, and that experience is NOT passive sensing of pregiven attributes of the world. Instead, our perceived world is constituted through complex patterns of sensorimotor activity [119, p. 164]. The history of structural coupling between agent-environment is what makes the world what it is for the agent (*enaction*) and defines its lived experience.

Coming from linguistics research, Lakoff and Johnson [79] identify four aspects of embodiment upon which to ground advanced cognition: physiology; evolutionary history of the agent; practical activity; and socio-cultural situatedness. According to the authors, the fact that the mind is embodied does not only point at its neurological instantiation but equally importantly at "the dynamics of our perceptual and motor systems that play a foundational role in concept definition and in rational inference" [1, p. 105]. Reason is evolutionary, built on simple modes of inference and metaphorical mapping, and complex behaviors emerge as an agent-environment interaction. Concept formation is grounded in the acting body. Action is situated and its meaning is context specific.

Embodied cognition gives our body a central role in all cognitive activities. In this view the body shapes the mind to the extent that the way we think is a direct consequence of the body we have and the environment we live in [2].

Embodied theories in psychology suggest that the body is intimately involved in the processing of social and emotional information [89]. Several are the findings that the state of our body, in terms of posture or movement, affects information processing. For instance Duclos et al. [34] showed that participants that were put into various bodily positions associated with fear, anger and sadness experienced affects modulated by this postural conditions. In on-line embodied cognition or situated cognition [125], that is activity operating directly on a real-world environment, a human is involved in acquiring and modifying a repertoire of bodily responses to stimuli. This process of acquisition is building knowledge, a collection of stored embodied states later used not only in future experiences but also in cognitive activities not coupled with the real world environment, i.e. offline cognition. For instance when we think about an object without interacting with it or without even having it in front of us we activate embodied states as if we were engaging with it. This mechanism has profound consequences in rejecting models of cognition based on symbolic representations in high-order cognition [42].

An embodied view of the mind rejects the often too common metaphor of mind as a computer, based on the principles of that the software of the mind is independent of the hardware of the body and the brain ([10]; see [29] for different accounts of symbolic representation in neuroscience and cognitive science) and that high-level cognitive processes, such as inference, categorization and memory are performed using abstract symbols in language-like form.

Differences and specificities of theories of embodiment have emerged and the depth of the subject is such that I cannot get into more detail in here. I try to narrow down the scope of this theoretical review to the concepts that most inspired the approach and interests guiding the research presented in this thesis.

2.2 Embodied Interaction

What phenomenology offers to HCI research is a reminder: human-computer interaction does not happen in a vacuum, it is situated and includes aspect of social interaction. Secondly embodiment implies that beings are agents who interact with the world directly, they explore and make sense of the world being in contact with it, as a child who puts objects in his mouth to understand them. Humans do not represent reality in their mind and act accordingly, instead they construct reality through action. The phenomenological account of being-in-the-world reflects in the importance given to physical computing by the modern HCI research, an approach early advocated by Dourish [32].

Dourish refers to a trend that focuses on the gradual incorporation of a wider range of human skills and abilities into interacting with computers: it suggests to move computation and interaction out of the world of abstract cognitive processes and bring it to the same world where other kinds of ordinary tangible interaction take place, to involve different senses, direct manipulation of artifacts and the use of effective metaphors.

For the latter, advocates of embodied cognition from linguistics research have argued that metaphors are building blocks for ordering experience [79]. A metaphor, from the greek "metaphora" meaning *transformation*, is an interaction between a target and source domain linking two concepts. Rather than being the result of high-level abstract cognition, Lakoff and Johnson [78] argue that metaphors arise from experiential gestalt, named *embodied schemata*, relating the movement of the body, its orientation in space and interaction with objects. Metaphors that connect embodied schemata, sensorimotor registered states or stored embodiments as previously formulated, to abstract concepts are called *embodied metaphors*.

The verticality schema is one of those schemata on which many metaphors rely on supporting the association of UP IS MORE. For example HAPPY IS UP and SAD IS DOWN are embodied metaphors based on spatial experiences called orientational metaphors. They associate directionality with concepts, in this case with affective states. Consequently, expressions such as "feeling up" or "that goal lifts me up" are understandable.

Ontological metaphors connect an abstract concept to a physical object or substance. By doing so, it is easy to associate many physical attributes derived by the experience we had with objects and substances to intangible concepts. If stocks are entities like boxes, one can pile them up and say "stocks are going up" or "market is crashing".

2.3 Tangible Interfaces

When designing human-computer interaction and user interfaces, metaphors are used to represent elements of the interface, where icons and widgets visually hint at a functionality modeled as an action grounded in the physical world (e.g. picking and dragging, etc.).

In physical computing the mapping between action on the interface and functionality should be devised in a way that leverage on skills already accumulated in embodied experience and with an interaction technique based on embodied metaphor. Tangible User Interfaces (TUIs) are a family of interfaces that apply this criteria. Ishii and Ullmer [69] presented several instances of TUIs, which they defined as user interfaces that "augment the real physical world by coupling digital information to everyday physical objects and environments." [117].

Researchers have adopted metaphors and theories of embodiment to their design to different extents. A taxonomy for TUIs has been proposed by Fishkin [41]. The two axes proposed are *metaphor* and *embodiment*. Fishkin recommends to not confuse the embodiment used in his taxonomy with the concept of embodied interaction of Dourish. Fiskin's embodiment refers to what extent the user manipulating the object thinks of the states of the system as being "inside" the object they are manipulating. The metaphor axis, split into noun-like and verb-like metaphors, refers to the kind of metaphor underlying the analogy between functionality of the interface element with the appearance and user action needed to access it. Sometimes the link exists at the level of shape or position (windows and desktop are examples of noun-like), other times the action performed on the interface echoes the functionality of that action, like deleting a file by putting into the trash bin (verb-like). We can speak of a full-metaphor when the action on the TUI is the action in the physical world. This latter case is also referred to "Really Direct Manipulation" [40].

The TUI framework will be relevant for the interaction designed in Publication V described in Section 3.5 where I designed a system to dynamically create embodied and metaphorical sonic interaction with daily objects. My work has focused on exploring ways to exploit our physical and tactile skills, placing computing where the action already is and technology should not constrain but, in fact, leverage on physical interaction. The purpose of this thesis is not to demonstrate the superiority of embodied interfaces over those less embodied in every possible context. On the contrary, many tasks are better accomplished by sitting in front of a desktop computer than by exploring the environment where our body is situated. In fact, the goal of this thesis is to show potential scenarios where it would not make sense to employ more traditional GUI-based interaction and other techniques of interaction are more suitable and exploitable.

2.4 Enaction

The enactive approach helps us to link the theory of an embodied mind to the concept of experience. For the enactive approach, cognition is embodied action [28] and action is what makes an encounter a meaningful one. Cognitive systems "actively participate in the generation of meaning in what matters to them; they enact a world" [28, p. 488]. Drawing the enactive approach to perception, Nöe states: "Perception is not something that happens to us, or in us. It is something that we do." [92, p. 1].

To say that cognition depends on the body would not be very surprising or useful per se. With regard to the relation between action and perception, enaction and sensorimotor theories of cognition offer explanations on how the body shapes cognition and how action and perception are two sides of the same coin called sense-making. An example often used is that of perceiving the softness of a sponge [87]. This characteristic of the object can be understood only as the response of that object to my action of squeezing it with my hand and applying some pressure on it. The experience of softness is shaped by my body (for a small insect a sponge is as hard as a rock) and it is the outcome of probing the object with my motor skill and perceiving the reaction, an experience rather than an exercise of sensing.

This example lets us introduce the concept of sensorimotor contingencies in perception. In the case of visual cognition Nöe [96] proposes to treat vision as an exploratory activity mediated by lawful sensory changes produced by various motor actions. What distinguishes vision from audition or touch is the kind of structure of the rules governing the sensory changes. Different sensory domains are characterized by different variance properties. The eye for instance can move in ways (rotating) that produce certain changes in the retina stimulation that are lawful. For example looking at a straight line results in different curvature inside the eyeball depending on the fixation point. There is a lawful distortion rule governing the variation of the curvature of the line with the variation of the fixation point.

Auditory sensorimotor contingencies have a different structure being based on the movement of the head. Sound localization has been explained within the sensorimotor approach [3].

The implication of sensorimotor theory for cognition is that action is necessary for sense-making and goes as far as arguing that cognition is an exercise of mastering of sensorimotor contingency.

Sensorimotor contingencies are close to the concept of invariants in the ecological approach to perception [51], i.e. "properties that remain constant across transformation produced by self-motion" [18, p. 13]. As ecological approaches to perception, sensorimotor contingency theory acknowledges the relation between the body and the environment, in that perceptually relevant information is revealed by active interaction in the environment and the view that cognition arises from exploring the changes of the environment as a product of the body activity. Sensorimotor approaches to perception stresses that the mechanics of perception are based on coupling between the movement of an organism and co-occurring changes in sensory stimulation.

For the enactive paradigm every encounter between an organism and its environment is meaningful because organisms participate actively in the exploration of the world and they do that from their own perspective. In this respect the enactivist approach departs from the ecological view of cognition in that organisms do not retrieve invariants present in the world. Rather, organisms bring forth features of the environment that emerge from action and the dynamics of the environment, they are 'feature specifiers'. Their being in the world is inherently a sense-making activity.

Bruner [17], who developed theory of cognitive development and learning, distinguishes between symbolic, iconic and enactive knowledge. Symbolic knowledge relies on abstract reasoning, iconic on visual recognition and the capacity to compare and contrast while enactive knowledge is built on motor skills. Enactive knowledge, to be distinguished from symbolic knowledge, can be described as registered sensorimotor states accumulated through experiences. Experiences are built on sensorimotor loops and motor skills.

Quoting Nöe: "We enact our perceptual experience; we act it out" [92, p. 1].

2.5 Enactive Interfaces

A thesis that presents cases of enaction using mediation technology is ultimately concerned with interfaces that allow for the expression and transmission of enactive knowledge [94].

Enactive interfaces can be used to substitute certain modalities with others creating new sensorimotor coupling, which the user can master and leverage on in performing actions [121]. For instance, tactile sensory substitution by means of tactile feedback is the enactive torch [48], designed with infrared sensor mounting and vibrational motors attached to a Velcro wristband. The interface provides tactile feedback proportional to the vicinity of intercepted objects in axis with the torch.

The nowadays discontinued Tactaid auditory-tactile substitution device is an example of a wearable interface that translates into tactile feedback the audio captured by a microphone to aid people with impaired hearing, improving speech sentence understanding by about 10%. [102]

New sensorimotor contingencies can be learnt as in the case of the "6th sense" interface proposed by [88]. In one experiment adult subjects received orientation information, obtained by a magnetic compass, via vibrotactile stimulation around the waist. The results indicated that the sensory information provided by the belt is processed and boosts localization performance.

The main difference between TUI's of Section 2.3 and enactive interfaces lays in the form those interfaces represent knowledge. In TUI the interface is axplicit, it is a physical artifact instrumental to access digital

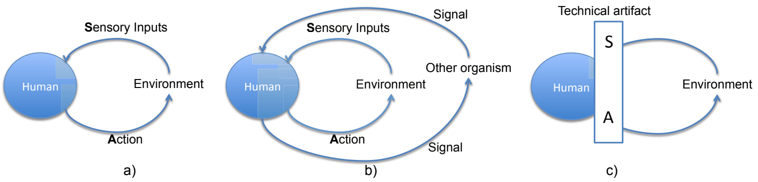


Figure 2.1. Basic schemes of sensorimotor coupling between an organism and its environment as in [109, p. 3]. The sensory inputs, S, guide the actions, A; the actions A modify the sensory returns. In a) human and environment are coupled through sensorimotor activity, which constitutes the basic of lived cognition. In b) auxiliary signals are exchanged between organisms that can create patterns of coordination of modulated sensory activities. In c) the mediation of technical artifacts and tools shapes the environment brought forth by the human when they become transparent.

information. Guiding principles informing the design of enactive interfaces should be to serve as a transparent medium for augmenting our natural skills of interaction with the world [48].

In some cases of technological mediation between human and environment the interface can become transparent, i.e. it is not experienced, it disappears from the awareness of the user in the interaction yet it hiddenly mediates it. This phenomenon is exemplified by the case of the white cane used by a blind person. Once the tool is integrated in the activity the person becomes one with the tool, which ceases to exist as a separate object of consciousness. As originally proposed by Heidegger [58], the tool becomes *ready-to-hand*.

With this thesis I contribute to the notion that designing an enactive interface is to implement a coupling that can be mastered through the exercise of motor skills and designing enactive experiences is to create technology-mediated situations that offer the potential for new knowledge through user action. In other words it is not enough to provide extra feedback to the user if the interface does not offer sufficient motor affordances to close the perceptual interaction loop [121, p. 59].

2.6 Enaction in this thesis

In Figure 2.1 three schematic representations of human-environment sensorimotor loops as originally presented by [109] are shown. The organism discovers a world that is built of the consequences of its action for its sensory inputs [109, p. 3]. The organism enacts the world.

When the organism encounters others, communication arises. In the context of sensorimotor coupling, communication is a coordination of action guided by signals sent between the organisms. The signal received modulates the action of one organism and symmetrically signals can be emitted to affect actions of others. Patterns of coordination of action can emerge from this coupling.

The role of the technology I designed is to extend the sensorimotor loops with auxiliary channels or create novel ones.

For the sonic interaction I will discuss cases where the sensory input is sound and the action is movement and tactile contact with the environment. Moreover those are cases of closed-loop interaction, the user is in contact with an interface that produces and modulates sound, that in turns constitutes a feedback that guides the action of the user.

The sensorimotor coupling can be altered by introducing an interface that modifies the outcome of the action creating a modified coupling, a different contingency that might affect perception and create illusions. For instance, sound feedback added to the tactile interaction with an object through contact can result in tactile illusions like pseudo-haptic effects [39].

Augmentation is a particular class of techniques for designing interaction that uses pre-existent sensorimotor coupling but associates the display and manipulation of digital information to those interactions. From this perspective augmented reality can be broadly considered an enactive interface.

In the discussion section of each chapter, the case studies will be abstracted and schematically represented in terms of the basic schemas of sensorimotor coupling model of enaction.

3. Multimodal interaction in audio augmented reality

This chapter describes the content of Publication II and Publication V and the related works. After a brief introduction about audio augmented reality, my contribution to this technology will be structured in three sections: Spatial rendering, Audio and haptics and Tangible interaction with virtual sounds. The first of these three topics is relevant for both papers while audio and haptics and tangible interaction are more specifically addressed in the individual publications PII and PV respectively.

3.1 Audio Augmented Reality

Audio Augmented Reality (AAR) belongs to the field of Augmented Reality and it is the sonic counterpart of the more common visual augmented reality in which computer-generated visual information are presented to the user through mobile screens, head-mounted displays, by overlaying artifacts or landmarks in the physical world. AAR systems aim to extend the physical environment with virtual auditory cues and sound information. The augmentation can take the direction of seamlessly mixing real and virtual sounds [56] or augment the environment with purely virtual sounds. Compared to visual displays, auditory displays can provide non-obstructing cues during the user exploration of data presented as sounds.

When using *earcons*, "abstract synthetic tones which can be used in structured combinations to create auditory messages" [16], auditory displays can present simultaneous information since they can be designed to have different auditory parameters such as pitch, timbre and rhythm (see [85] for guidelines on designing concurrent sounds in AAR). Also, following one of the principles of Auditory Scene Analysis [14], audio can be spatially presented from different directions and distances thus avoiding cluttering and potential grouping of sound sources into streams.

Another category of data-encoded sound are *auditory icons*, defined by Gaver [50] as "everyday sounds mapped to computer events by analogy with everyday sound-producing events". They have an advantage compared to earcons in terms of familiarity of the sound and the analogy between the everyday world and the model world of the computer displayed by the interface.

3.2 Spatial rendering

AAR systems require at least 2 components: the tracking of the user with respect to its environment, in order to activate, spatialize and transmit the sound information, and the actual rendering of the sound at the correct spatial location. The tracking mechanism of both Publication II and Publication V is based on a combination of positional and orientation tracking and it has been implemented with off-the-shelf components, using Microsoft Kinect for the positional tracking and inertial sensors for the orientation of the user's head. The fused tracked data are fed to a module that renders the individual sound sources that compose the audio scene.

The objective of the spatial rendering is to maintain the correct audio perspective that preserves the position of the sound objects in the audio scene while the user explores the environment by moving and directing their head in different directions. Sound localization is a complex task involving integration of multiple cues with motion.

The rendering implementation strictly depends on the audio system employed. In the case of loudspeaker reproduction a multi-channel system should be employed using spatial techniques such as vector base amplitude panning (VBAP) [99]. In VBAP only 2 or 3 adjacent speakers are active for a single source and the localization accuracy degrades gracefully as the listener is moving away from the center. An enhanced version of VBAP with distance cues was used in Publication II rendering the sound over an 8-channel ring setup on the horizontal plane.

While this solution might be used in certain contexts where the listening position is quite static, a binaural reproduction is more desirable in the context of exploration of an audio scene and interaction with virtual sounds. Different techniques for binaural rendering are available. The binaural renderer should be able to provide correct perceptual cues for each object in the scene. These cues are accurate inter-aural level and

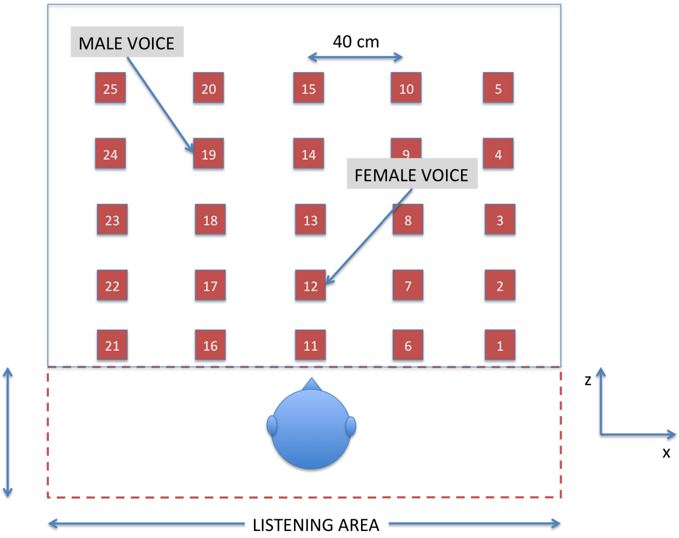


Figure 3.1. Setting for the localization accuracy user study: sounds are rendered at the placeholders positions, positioned on the same plane. The listener is free to move within the designated listening area.

time differences that relate to directional perception, and appropriate cues that relate to distance perception. The directional cues are encoded in Head- Related Transfer Functions (HRTFs), which is an effective choice for binaural panning [123]. Distance cues can be provided by modifying the ratio between the original sound and the reverberated version of it [129]. The closer the user is to the sound source the dryer is the mix between original and reverberated sound. This strategy was employed in both Publication II and Publication V.

In our system ATSI (Publication V), the user can augment physical objects with spatial sounds in a small scale environment, such as a normal-sized working space or living room and be able to move them around while listening to the sound attached to them as if it would be located in the object. The localization accuracy that can be achieved with the system was tested in a study where the listener needs to identify the correct position of a sound source, either a female or male voice (in a randomized order) among a grid of numbered place-holders, as shown in Fig. 3.1. These locations are 40 cm apart in both lateral and depth separation. Two conditions were presented: 1) only one source at the time was presented and 2) the two voices (male and female) were presented concurrently. The localization accuracy achieved in the user study is shown in Fig. 3.2.

The results show that the achieved accuracy is suitable for the audio

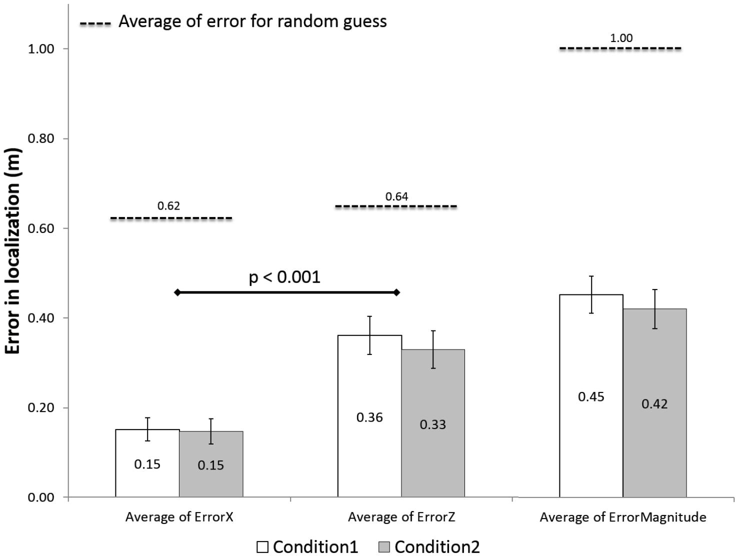


Figure 3.2. Average localization errors in the two conditions, with 95% confidence intervals and random guess average error.

augmented scenarios proposed and the concurrent presence of more than two sound sources did not significantly affect the performance of the localization. As expected the lateral error (ErrorX) is significantly smaller than the error in depth (ErrorZ). Localization on the azimuth plane is well rendered with non-individualized HRTFs while the vertical one is more problematic. The distance cue as a combination of amplitude decay with distance and the reverberant ratio mentioned before still provide cues that the user can rely on and learn to use achieving well below chance level error of localization.

3.3 The sense of touch

Touch and HCI have gone "hand in hand" due to the fact that touch is a common medium of interaction in the world and it mediates encountering and interaction with daily objects. From an enactive perspective to interface design, touch is a modality where the continuous interplay of action and perception is more evident. Concretely, touching an object is to feel the resistance offered as a result of the action of pushing, and a touch-based interface should take into account the existing sensorimotor abilities of the users.

In the case of virtual reality, every time we design an interface that operates by extending the user's perceptual-motor capabilities "into" these virtual environments, in some sense, the lack of tactile sensory feedback make users impaired in acting within the virtual environment [121].

Touch provides a rich set of information about the object we are interacting with such as texture, softness and viscosity, and provides dynamic feedback on our interactions with physical environments around us. The skin physiology is based on mechanoreceptive units embedded in the outer layers of the skin that provide tactile sensations (for an overview see [23]).

Tactile stimulation has been used to substitute other modalities such as vision and hearing when they were impaired or saturated, as in the case of enactive interfaces mentioned in Section 2.5.

3.4 Audio-tactile integration

Auditory displays present information encoded in the sounds rendered in a scene. Providing information with sound can be further enhanced if the virtual sounds present some sort of physicality. If sounds are physically embedded in objects, a visual and tangible cue for the position of the sound already exists, but in the case of a purely auditory display of spatialized virtual sounds, the user hears the sound without seeing where it originates from. If the sound is divorced from the localized sound, we encounter a situation that goes against an ecological approach to the perception of the auditory scene created.

The integration of the auditory and tactile modality is in-built in the physical world. Sounds can be produced by the collision or scrubbing of our skin against an object. Auditory feedback conveys a rich set of information, such as material, shape, size [49] but the tactile sensation can enrich the experience by providing information such as hardness or stiffness.

Our perceptual system integrates the combined audio and tactile stimuli in one perceptual unit and their interplay can drastically modify it. In fact, one modality can prevail over the other in case they are incongruent. Bi-modal perception in continuous contact interaction (i.e., scraping or sliding) has been studied in the case of bare skin exploration. The relative contribution of the auditory input in the perception of textural quality of material (roughness) was assessed by [82] showing that the tactile cue dominated the auditory cue in a surface texture discrimination task in the case of congruent stimuli.

Interestingly enough, the audio-tactile feedback contribution can be adjusted by a manipulation of the auditory input. Jousmäki and Hari [73] demonstrated that auditory inputs could affect tactile roughness judgments and later [54] showed the effect of auditory frequency manipulations on the perceived tactile roughness and moistness of surfaces. A review of experiments that show robust effects of auditory input on tactile frequency discriminations can be found in [47].

Tactile feedback has found many applications in virtual and mixed reality contexts. In the area of realistic multimodal simulation in VE, the haptic shoes of [114] provide audio-haptic integration and rendering of footsteps on different simulated surfaces. Tactile stimuli have been used to augment watching-movie experience [84] or guide a user during navigation tasks when the visual cue was saturated or impaired [113].

To provide a user with tactile sensations, a tactile display that renders a tactile stimulus is necessary. Tactile displays are created by means of grids of vibratory actuators and have been used to provide collision cues to a user immersed in a virtual environment [11], or proving guidance to the user [106]. Van Erp and his colleagues have created a vibrotactile vest to display orientation information for astronauts in zero gravity [118].

Giving physicality to virtual sounds through the interactive control of tactile feedback was explored in Publication II where an integrated system for the creation of a combined audio and tactile display was described. Several sound sources will move around the user and collide with her body. The collision event is presented not just in the auditory scene but also as a tactile sensation on the skin by means of a custom-built belt containing actuators.

With respect to the sound rendering, the relationship between physical and perceptual parameters of objects in contact have been studied [76] and different synthesis models for the sonic interactions between solid surfaces are available [25]. In order to model the vibration feedback that best integrates or belongs to the sound event, different strategies are possible. If the system incorporates measurements of contact or collision, then collision sounds or tactile signals can be synthesized from features extracted from the measured signals, as described by Turchet et al. [116]. If the environment is simulated then it is possible to calculate similar parameters and feed them to both the audio and the tactile rendering module. That is the case of Publication II where for each contact event between the virtual objects and the user an envelope is obtained by sam-

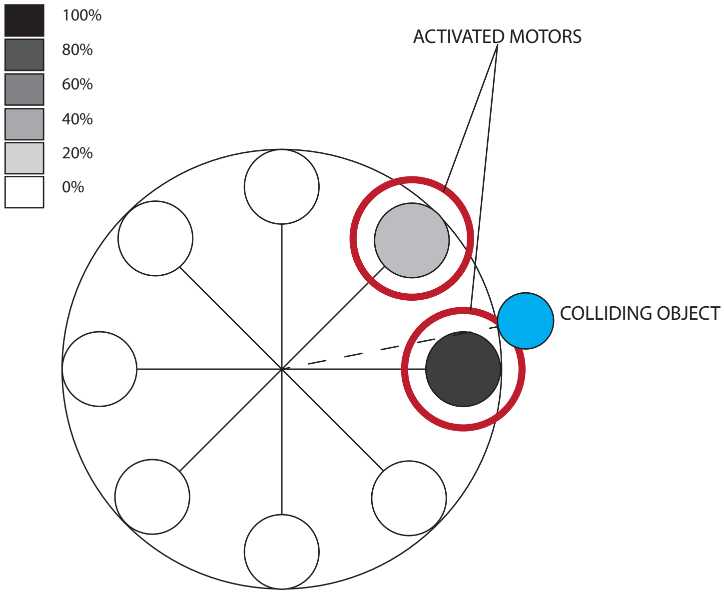
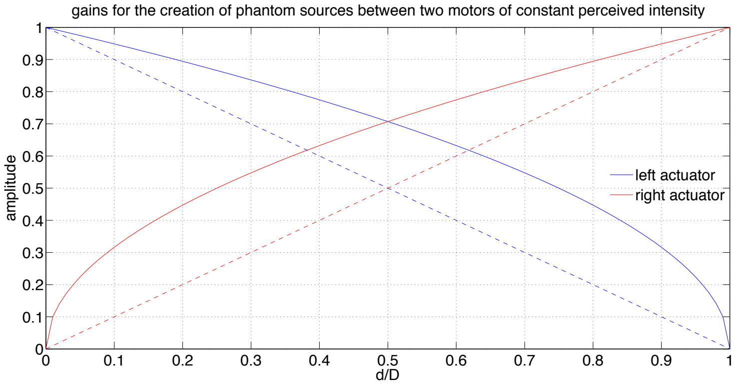


Figure 3.3. Spatialization of the vibration feedback on the belt. The contribution of the individual objects are summed together and encoded in the eight motors.

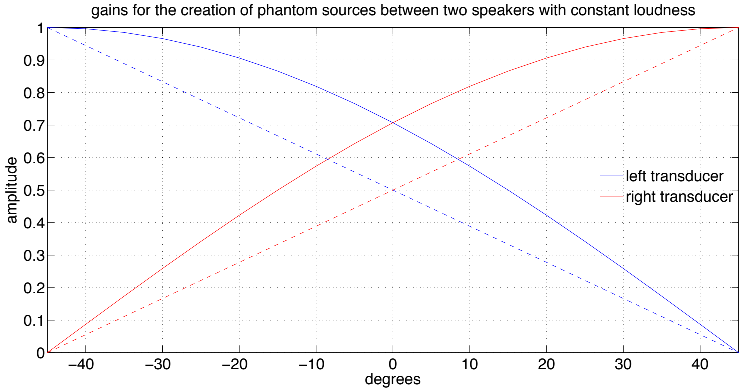
pling the magnitude of the speed of the object relative to the speed of the participant in the moment of the collision. This amplitude is then fed to a leaky integrator (a first-order recursive filter), with controllable decay characteristics. This obtained envelope is used as a gain factor to control the amplitude of the vibration.

The design took advantage of tactile illusions (see [57] for a taxonomy), in particular a phantom sensation also known as "funneling", created by two stimuli applied on the skin at a certain distance. The phantom sensation is such that the user would not perceive two actuators, but rather a single actuator moving between them [19]. The position of the phantom source can be controlled by modifying the ratio between the amplitude of the two vibrations applied using a tactile rendering scheme inspired by the stereo panning law. In Figure 3.3 the spatialization of the sensation for the collision of the virtual object with the skin is shown. The colliding object hits the belt between two actuators that get activated and vibrate to provide the tactile feedback. The gain of the motor is calculated using the profile shown in Figure 3.4. The reader can notice the similarity between a stereo panning law profile (4.3b) and the one used for the rendering of the funneling illusion (4.3a).

A tactile display that manipulates the gain of the actuators according



(a) Actuators gain profile



(b) Sound gain profile

Figure 3.4. Amplitude gains for equal perceived intensity of phantom source in a) between two motors and b) between two loudspeakers.

to this principle can be used to provide a diffused surface for interaction and a correctly spatialized tactile sensation. With a belt embedding eight actuators as shown in Fig. 3.5, the sensation of continuous tactile strokes on the skin was achieved. This tactile illusion was also used in the case of an augmented actuated chair by Israr et al. [70] to create continuous tactile brushes by means of any grid topologies of tactile displays. While more general than the one proposed, it offers the limitation that the stroke needs to be pre-computed by the algorithm before being delivered to the user through the intensity and temporal controls for each actuator.

All in all, a multimodal system that combines motor affordances and audio-tactile feedback to close the perceptual interaction loop was created. With the belt interface and the tactile rendering it was made possible to actively touch the virtual sounds. The interface is enactive because it

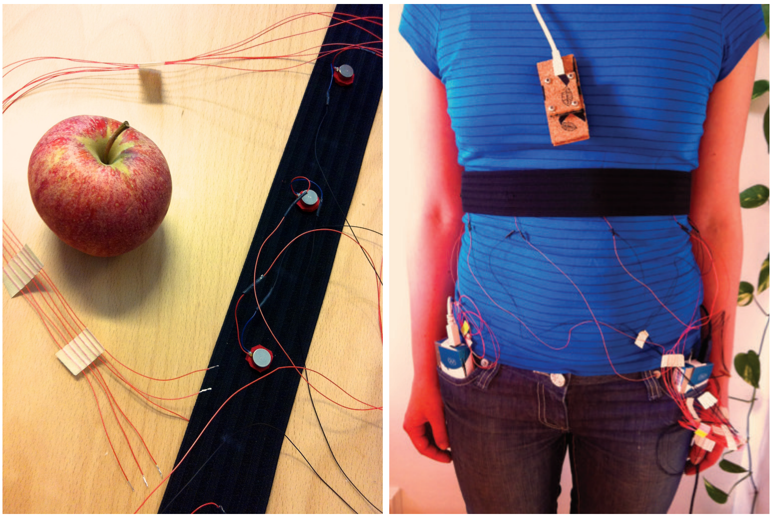


Figure 3.5. Elastic belt with eight vibration motors equally spaced in a ring and orientation tracker in the middle of the chest to provide torso orientation information to the system

enables contingencies to be explored through movement that are plausible because they are physically-informed by the simulation of the agent-environment coupling. More concretely, the user can provoke the collision and continuously perceive the result of their action (bouncing against the object, collision, audio-tactile feedback from the event, source goes away and might bounce against a wall or another object and bounce back, etc.).

The spatialized audio-tactile display was used to create a generative musical piece that follows an ecosystemic approach to music composition [13], [37]. *simpleLife* is a self-generative ecosystem of sound particles that emerges from and is governed by simple rules as attraction, repulsion, pro-creation and life span. This environment inhabitated by sound agents is perturbed by the presence of the visitor who bounces against the particles hearing them and feeling them on the skin in a multimodal embodied experience (Figure 3.6). The result is an installation with emphasis on spatialization similar to the research of Wilson [126], or the more general compositional framework by Schumacher et al. [107], but emerging from the physical exploration of the sonic and tactile environment.

3.5 Tangible Interaction with Virtual Sounds

In Publication V I presented another approach to physical interaction with sound in an auditory scene where physical objects are sonically aug-

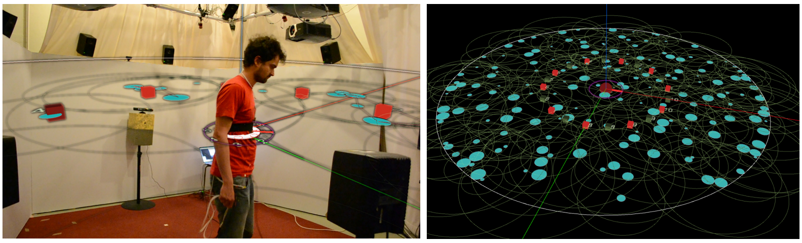


Figure 3.6. User wearing the belt and the rotation tracker of Publication II immersed in the sonic environment visualized on the right. A Kinect sensor tracks the position of the center of mass of the user in the simulated environment.

mented. Virtual sounds are associated with or embedded in them so that their presence becomes visible and the interaction with the sound is mediated or supported by a tangible artifact.

Other spatial sound systems designed for tangible interaction have been presented in the last two decades. Ishii et al. [68] used bottles as an interface to access digital information. Sounds are released by opening the bottles. The design is an example of a transparent interface that embodies sound (*full embodiment*) and uses a metaphor of freeing (*verb*) the sound as if it was an essence (*noun*).

Williams et al. [124] have proposed a system called SignalPlay where physical objects collectively controlled a soundscape by letting the users change their position and orientation in space, both individually and collectively experiencing and transforming the sonic environment. After installing this system in a "wild" setting such as a gallery space, the authors could collect evidence of the different modes of interaction with the augmented objects: *intrinsic interaction*, which comes from the intrinsic physical characteristics the object, *iconic interaction* which means interacting with the physical icon through the affordances of the object associated to it. Finally, *instrumental interaction* happens when the icon and the physical characteristics of the object are not in focus, but the user rather tries to deliberately control the effect of her action on the object and "play it" as an instrument.

This system as many others relies on augmentation of objects by means of sensors or markers to track and register the objects (see [75] for instance as an early example). Other systems, such as the one's in [31] or [90], let the user manipulate virtual sounds detached from any physical objects in an augmented environment by means of gestural interaction. Both systems require wearables such as special wands with markers or

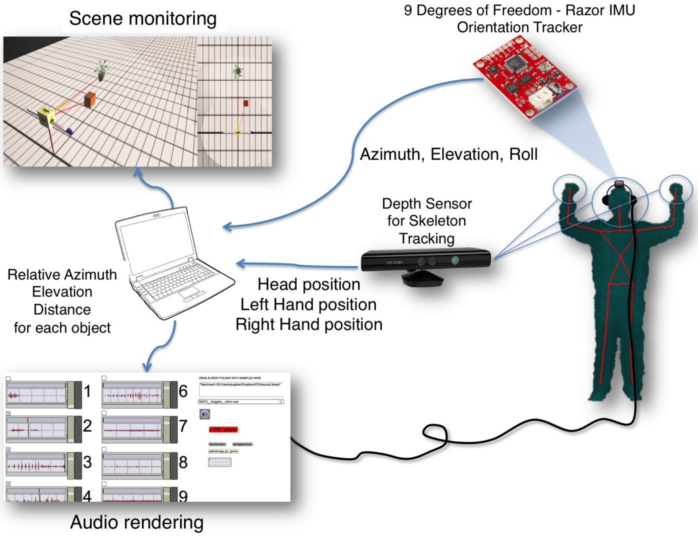


Figure 3.7. Modules of the ATSI system of Publication V

haptic gloves for the manipulation of the virtual sound sources.

To overcome these constraints, ATSI was developed, a new system that incorporates elements of audio augmented reality with tangible interaction. In ATSI the user does not need to wear any devices but headphones. Furthermore, (and this is a key-differentiating factor to any previous system) the physical objects do not need to be modified in any way, as the interaction and the auditory augmentation of the real scene is done by hands-tracking alone, offering a non-intrusive natural AR experience.

ATSI employs the binaural rendering discussed in section 3.2 and adds gestural interaction for assigning virtual sounds to physical objects. Tracking of the hands is provided by the Kinect sensor (see Figure 3.7). By means of intuitive gestures the user assigns the sounds to objects and afterward he/she is free to move the sound by moving the object attached to it. The process happens when exploring the visual/auditory scene while keeping the correct spatial auditory perspective inside the augmented scene during the actions of attaching sounds, picking and moving the sounding objects. The sequence of gestures is shown in Figure 3.8. The sound could also be manipulated with gestures and it is rendered by a flexible granular sampler that allows for a variety of continuous sound manipulation. To facilitate the memorization of the gestures and provide feedback on the correctness of their execution, auditory icons are associated to each of these actions and acknowledge the completion of each

Figure 3.8. A typical interaction scenario with the ATSI system: an object on the table is augmented with sound rendered from the perspective of the user who explores, selects, picks and re-positions the objects.

step.

The potential scenarios of application envisioned are auditory information display, soundscape preview for urban planning, sonic interior design and electroacoustic music performance. In the first scenario a user can dynamically attach information to objects that in turn can provide information about their status and call the attention of the user accordingly (a plant that needs water, food that needs to be eaten before the expiration date).

For sonic interior design one can imagine decorating an environment not only visually but also sonically. The spatial nature of the sound allows for the visitor of the decorated environment to explore and find an interesting and esthetically pleasing sonic location within the augmented space. For instance sitting close to the radio or close to the balcony could be associated to listening to different soundscapes.

Within musical applications both the sound preview of an instrument on display in a shop (imagine of approaching a rack full of guitars and being able to preview their sounds by passing in front of them), but also the more dynamic scenario of collective music playing are included. If non-acoustic instruments or controllers are used, it is possible to redirect the synthesized or amplified live sound to the point of interaction, such as positioning the amplified feed of an electroacoustic instrument to the location of the instrument. In this way the sound produced by each instrument could be heard by all the players wearing headphones coming from the correct location thus providing a very natural live-monitoring experience.

3.6 Discussion

This chapter has provided a description and contextualization of original case studies the author realized to illustrate methods and applications of embodied interaction with sound. The Publication II and Publication V

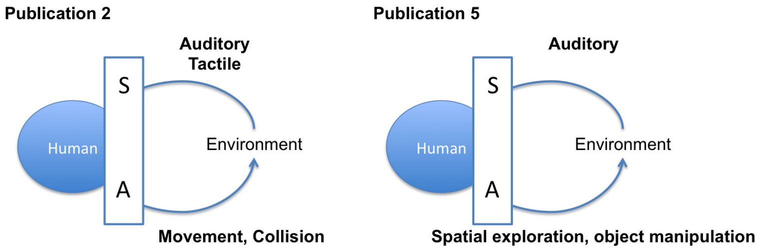


Figure 3.9. Schematic representation of the sensorimotor coupling designed in the publications PII and PV. In Publication II the human sensory-action loop is augmented by an enactive interface (an actuated belt) to integrate audio and tactile cues in a virtual environment. In Publication V an augmented auditory scene is presented and sound is heard and manipulated as if attached to physical objects. The sound is presented from the perspective of the person exploring the scene and sound is moved in space as tangible objects.

described systems for designing spatial interaction with sound. In Publication II, the design proved successful in the rendering of moving objects in contact with the participant and rotating around the torso thanks to the actuated belt. This design was an original example of audio-tactile integration in a virtual environment and used the phantom illusion with a low-cost tactile display to deliver continuous and correctly spatialized tactile brushes on the skin. This system could be used for installations and in augmented and virtual reality scenarios.

More focus on tangible interaction was given in Publication V. The localization accuracy of the virtual sound combined with the tracking technique used is adequate for sonic interaction in augmented reality. The gestural interaction devised allows for creating embedded and metaphoric associations between the objects present in the physical world and a personalized sound library in natural interaction. These two case studies complement each-other showing a combination of multimodal feedback from the collision with virtual sound sources and tangible interaction with virtual sounds mediated by physical objects. The two designs hint at an integration of augmented audio reality application with wearable interfaces for tactile feedback pointing towards a seamless integration of physical and virtual interaction across modalities.

In Figure 3.9 the two publications are represented in terms of the sensorimotor coupling model of enaction.

In Publication II, an enactive interface was designed by means of an actuated belt to integrate audio and tactile cues in a virtual environment. Here the rendering of the audio-tactile display is designed in accordance with the sensorimotor theory creating a physically informed relation arising every time there is a collision between the body of the person and virtual sound sources with rigid bodies. The sounds are spatialized creating a natural sensorimotor contingency to be exploited for making sense of the localization. Another lawful relation is that the tactile feedback happens on the body in a position consistent with the auditory scene. Lastly, the relation impact-tactile stimulus is extracted from the collision sound borrowing the invariances for the sound to inform those of the tactile display. The discovery of the property of the environment emerge from its physical exploration.

Publication V presents a case of tangible interaction in which physical objects are associated with sound. The auditory augmentation of the environment creates an additional layer in the world. While the manipulation of the sound is not enactive (the user cannot have a direct access to the source of sound and tangible manipulation of it), the auditory scene is accessed through the human enaction. The exploration of the environment is supported by the augmentation of the senses. The interaction with the added content happens through the same auditory-motion coupling that characterize everyday listening with the user being able to move and listen to sound from a first person perspective and reposition sounds as objects.

4. Sonic interaction

In the previous chapter I presented a series of case studies demonstrating the use of sonic augmentation and interaction in virtual environments and mixed-reality context. I focused on giving physicality to virtual sound with tactile feedback and to attaching information or musical content to physical objects. Also the sounds were spatialized to be perceived as objects in space.

This chapter describes the content of Publication III, Publication IV and Publication VI and the related works. I explore the possibilities offered by the use of sound as feedback during activities that involve motion of the body.

4.1 Interactive sonification

Sonification is defined as the use of non-speech audio to convey information [62]. At the intersection between HCI and sonification (see Figure 4.1) lies the field of designing systems that allow for direct manipulation of sound for data exploration: *Interactive sonification*. Interactive sonification [60] is a continuous auditory feedback created by the transformation of data into sound where the human user directly affects the synthesis process. The user is enclosed in an action-sound loop.

Hermann and Hunt [61, chapter 11, p. 275], broadly separate the ways humans generally interact with sounding objects in everyday life in two categories: physical manipulation, where sound is a highly informative byproduct of human activity, and the intentional use of objects to actively create sound, as the use of musical instruments. Interactive sonification belongs to the former, where the goal of the exploration and manipulation of the sound is data analysis. Playing music belongs clearly to the latter and it is a process where gestures are transformed into sound for the

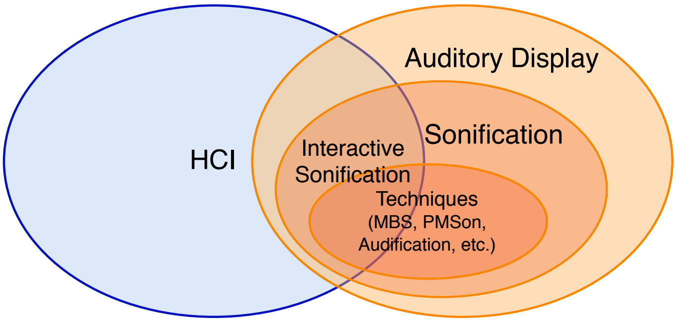


Figure 4.1. Topic Web for interactive sonification, from [61, p. 274]

purpose of expression.

This categorization closely follows the one introduced in Section 3.5 and helps us divide the contribution of this thesis to physical interaction with sound in Interactive Sonification (Publication IV and Publication VI) and instrumental interaction with sound in a musical context (Publication III). Nevertheless, the richness of interaction offered by musical instruments can inform the design of interfaces for interactive sonification, taking into account their core features such as real-time sonic response, a control of complex mapping and tactile feedback.

Interactive sonification finds a variety of application ranging from data exploration [97], to assistive technology for users with visual impairment [131] to application for sports (more in Section 4.3).

The possibilities offered by sonification builds on human capacity to extract information from an audio stream, perceiving multiple auditory objects within an auditory scene and being sensitive about the timing of sound events. These perceptual skills can be leveraged when designing alerting and monitoring systems or in ambient information contexts. Bringing interaction to sonification allows the user to access the data from a personal point of view including action and auditory feedback in the loop of sense-making.

Locomotion is a perfect example of physical interaction with sound that puts the person in the center of a sonic action-perception loop. From a perceptual point of view, walking sounds are self-generated and provide rich, multisensory information about the environment. They convey properties of the surface on which we walk, such as the material and the shape but also characteristics of the walker such as speed and stability. Also, the self-generated sounds inform the movement of the walker. From the walking sounds listeners can pick up several cues based on the features of the sound (for a discussion of perceptual factor and display technique, see [122]).

The interest for the interactive sonification of foot-floor interaction has grown (see for a review [63]). To give an example of application scenario, in VEs synthesized footstep sounds showed to affect users' behavior by stimulating more motion and increasing presence of the subjects [93]. Tajadura-Jiménez et al. [111] demonstrated that altering self-produced walking sounds led the subjects to change their perceived body weight, pattern of locomotion and also emotional state.

Sonification models for the simulation of footstep sounds have been proposed ([15], [116]). In floor-based interfaces sound has been used as a tool for rehabilitation on treadmills [112], and the contribution of auditory feedback to the perceived compliance of the floor and the effect on the walking pace has been studied by [115].

This thesis has contributed to the field of interactive sonification specifically by designing new foot-surface interfaces in Publication IV and Publication VI. In both publications I have developed and evaluated a system for the interactive sonification of jumping sounds on a trampoline. However, the strategies and the research objectives pursued were very different and demonstrated the wide range of interactive sonification.

4.2 Sound synthesis of jumping sounds

To our knowledge interactive sonification in the scenario of jumping on a trampoline has never been studied before.

Publication IV focused on the foot-floor interaction and created an interface that modifies the sound produced by the contact between the feet of the jumper and the trampoline surface. The goal of this study was to recreate the impact sound with different materials and study the effect on the jumper in terms of felt hardness. It was motivated by previous findings relative to haptic sensation modulation through visual modification, termed as pseudo-haptic effects [81].

The system developed is shown in Figure 4.2. The proposed sonification scheme was based on the following three steps: 1) detection of the foot-floor interaction by means of a contact microphone attached to the trampoline's membrane; 2) processing of the detected signal to obtain a control

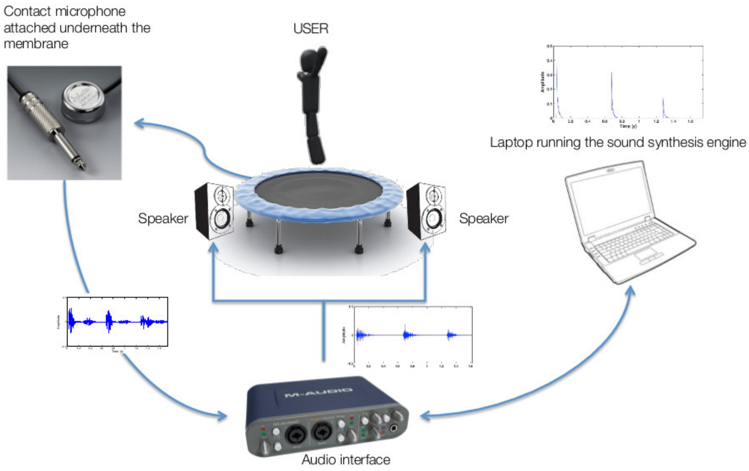


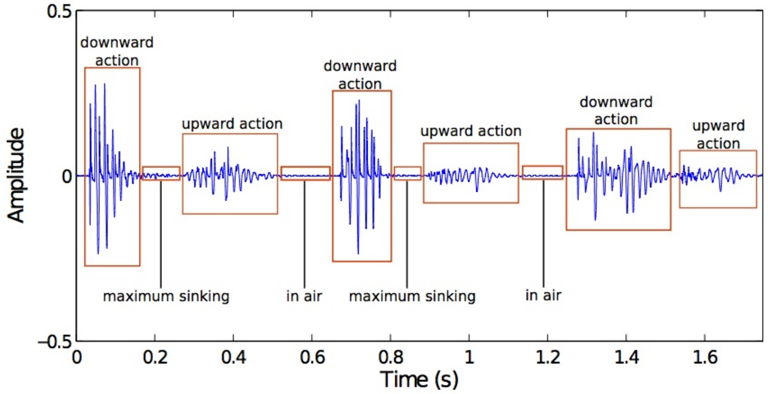
Figure 4.2. Schematic representation of the overall architecture developed

signal that accounts for the involved impact force and is used to drive the jumping sounds synthesis engine ; 3) sound synthesis and display.

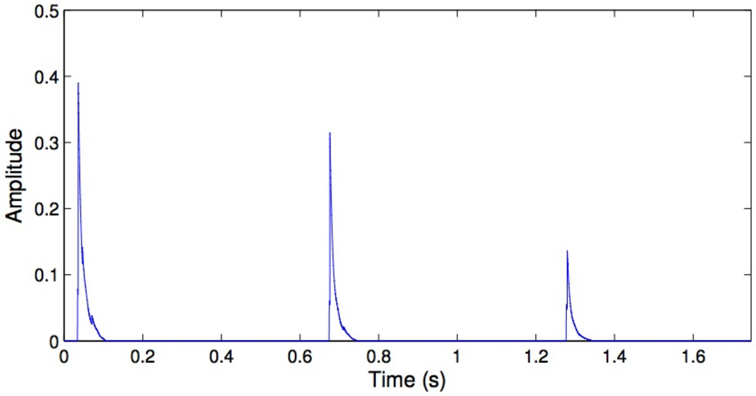
The signal processing involved is shown in Figure 4.3. A contact microphone (Schaller Oyster External Pickup 723) is attached to the periphery of the surface of the trampoline. The signal captured by the contact microphone was used as the exciter to feed the synthesis engine and simulate the wanted surface material. The models for impact, friction and crumpling, present in the Sound Design Toolkit [86], were utilized to create sounds of wood, gravel and puddle of water. Those synthesis algorithms are based on cartoonification, an approach to sound design that simplifies the sound models while retaining and exaggerating perceptual invariants from an ecological perspective [104]. The system showed a negligible latency of less than 5 ms.

In a user study the naturalness of the system and the self-reported user perception of the trampoline surface materials in the different conditions were evaluated. Participants rated their impressions (e.g. "how hard did the surface feel during the jumping exercise?") on a Visual Analogue Scale (VAS) . It was expected the different simulated surfaces to influence the user's behavior and to alter the haptic perception of the elasticity of the membrane of the trampoline.

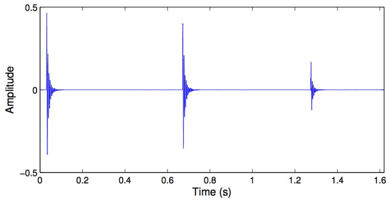
The overall results confirmed our hypotheses: participants correctly identified the types of the material (solid, liquid or aggregate) and to a certain extent correctly specified which material was synthesized. They reported that the sound conditions significantly influenced the way they jumped compared to the condition without added sound. Also, their im-



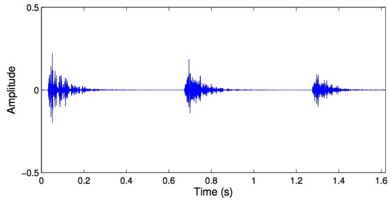
(a) Captured signal



(b) Exciter signal



(c) Synthesized wood sound



(d) Synthesized gravel sound

Figure 4.3. Sonification process: from the signal captured by the contact microphone to the synthesis of jumping sounds on different surfaces. a) the waveform of the signal detected by the microphone corresponding to three jumps with different dynamics. b) the signal utilized as an exciter simulating the dynamics of a typical foot-floor interaction corresponding to a jump, modulated in amplitude by the maximum value of the envelope derivative of each of the three jumps. Bottom figures are the corresponding synthesized sounds for wood (c) and gravel (d).

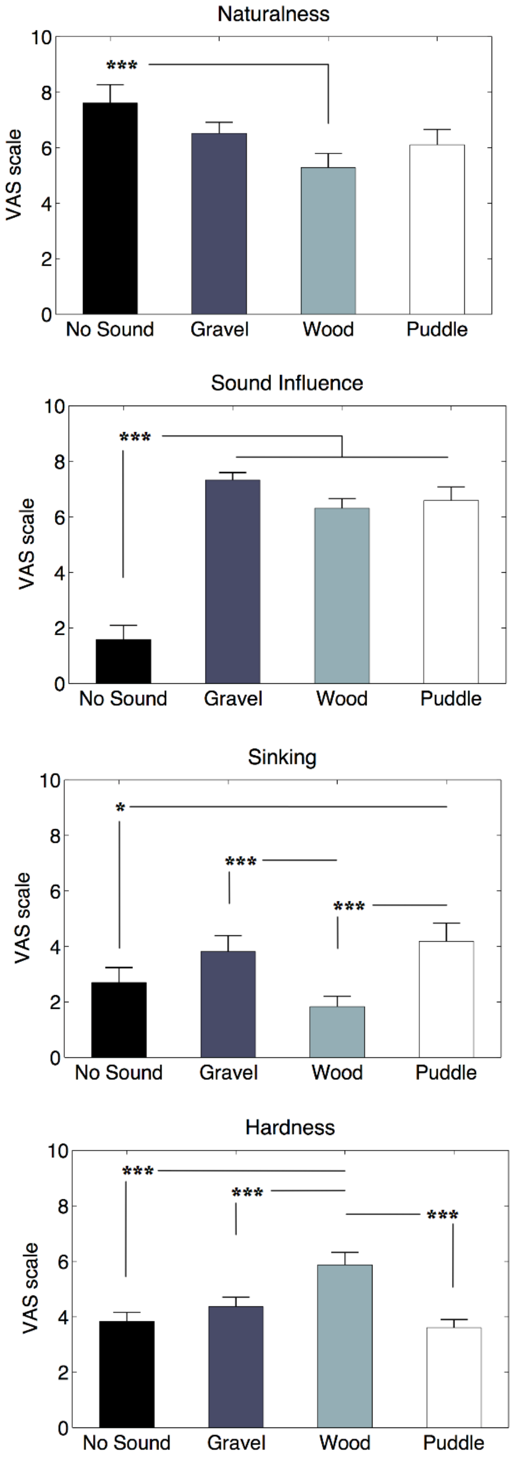


Figure 4.4. Graphical representation of the mean and the standard deviation for participants' answers to the questionnaire expressed on a Visual Analogue Scale (VAS). Legend: * represents $p < 0.05$, ** $p < 0.01$, and *** $p < 0.001$.

pression of sinking into the surface of the trampoline and the perceived hardness were affected by the sound conditions, with differences in trend with the hardness of the real material (see Figure 4.4)

Overall the system proved successful in simulating the different surfaces for the exercise of jumping and could be used in the future for studies on multisensory perception involving auditory and foot-haptic modalities.

One limitation of this system is that it approximates the impact force acting on the trampoline during the interaction between the feet and the mesh with the Ground Reaction Force extracted from the signal captured by the microphone [108]. While this is valid in the case of a solid floor, the trampoline system is rather a coupled vibrating system together with the jumper. Accounting for the real displacement of the membrane caused by the jumper might result in even more natural interaction when simulating aggregate surfaces and have further influence on the performance. This study did not measure performance (average tempo or highest point reached) and it would be worth considering this measure as a dependent variable of the auditory feedback but also as independent variable to evaluate the influence on the felt compliance. In order to measure the displacement force acting on the mesh, an inertial measurement unit (IMU) could have been attached to the membrane. Nevertheless, when simulating a solid surface one always need to decide at which depth of the displacement the solid surface should be and to sample at that level the value of the force to be used as excitation signal for the synthesis. In our case we decided to place the intended surface on the same level as the surface of the trampoline at rest.

4.3 Movement sonification

When interactive sonification is applied to motion data we can speak of movement sonification. The use of sound as an aid for the completion of tasks that involve body movement has found application in exercise, play and sports (see [64] for a review). The first is mainly concerned with health and rehabilitation, the second with the use of sound in gaming and entertainment in general, while in sports the main goal is to improve the performance of athletes.

The benefits of movement sonification are apparent in enhancing the performance of the human perceptual system [36]. Effenberg shows that performance accuracy can be improved using movement sonification and that this latter can play an important role in motor learning and control. Following the ecological approach to auditory perception, the sonification is based on kinetic parameters of the movement and relying on the natural connection between the kinetic event and the produced sound. The sonification process is a kinetic-acoustic transformation that can also provide auditory feedback to naturally silent movements.

Here we can draw a link between the use of movement sonification and the enactive interface based on sensorimotor contingencies described in Section 2.4. Indeed movement sonification constitutes an extension of active perception through the introduction of new audio-motor dependencies. The user can feel extra control of the sonified movement, because of the bodily engagement in patterns of sonically augmented sensorimotor contingencies, analogous to the sensorimotor account of visual consciousness [96].

4.4 Sonic trampoline

After successfully experimenting with physically-based synthesis of jumping sounds, I wanted to apply movement sonification of a person jumping on the trampoline and find out whether sport games developed using non-visual, audio-only information for action control could be motivating for the user. As a consequence further sensing technology and a different strategy for the sound synthesis was needed.

Other systems like AcuMotion [59] have been developed to design and evaluate sonification-based sport games. AcuMotion is based on a tangible sensing device that is tracked, a dynamic computer simulation model where the game happens and a sonification engine that presents all information about the position of relevant virtual objects to the user via sound. The system has been used to develop a version of badminton to be played without visual information called One-Player-Blindminton.

Another system called MotionLab Sonify: A Framework for the Sonification of Human Motion Data [35] was used to sonify breast strokes of swimming motion data, transforming calculated motion parameters into sound. The motions of wrists and ankles were used to modulate frequency and amplitude of the sound.

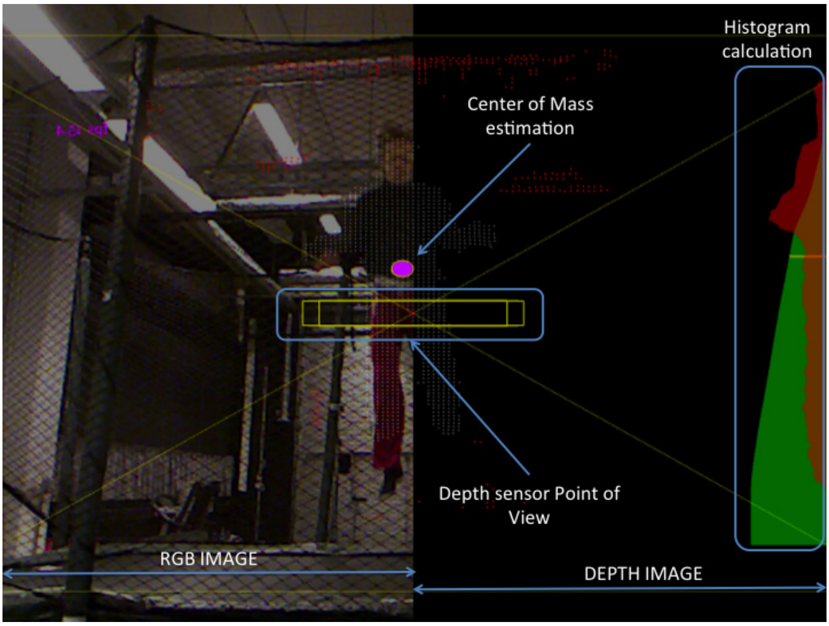


Figure 4.5. Depth map analysis and estimation of the center of mass from the histogram and original RGB image

4.4.1 Sensing technology

Our apparatus consists of an Acon Air Sport 16 trampoline with a safety net (169" x 96" x 114"), a contact microphone, loudspeakers and a laptop running the synthesis engine. The architecture of the system is similar to the one of Publication IV and the same onset detection algorithm is employed to capture the moment of feet-surface contact, later used to isolate the landing event.

To track the motion of the jumper and to calculate kinematic parameters for sonification, a solution based on commercially available depth sensors like Microsoft Kinect or Prime Sense was developed. By analyzing the depth image within a restricted volume, it is possible to isolate the jumper from the trampoline and obtain an image that only contains the user's body. From this image, an histogram function of the vertical position is obtained by counting all the points for each horizontal slice (red curve in Fig. 4.5). The vertical position where the integral of the histogram (green curve) is half of the overall area provides an estimation of the vertical position of the Center of Mass (CoM) in the depth image. The vertical position of the CoM and the derived vertical speed (Fig. 4.5) are used to analyze each jumping cycle and frame the events of interest (Fig. 4.6): the landing moment and the in-air phase of the jump.

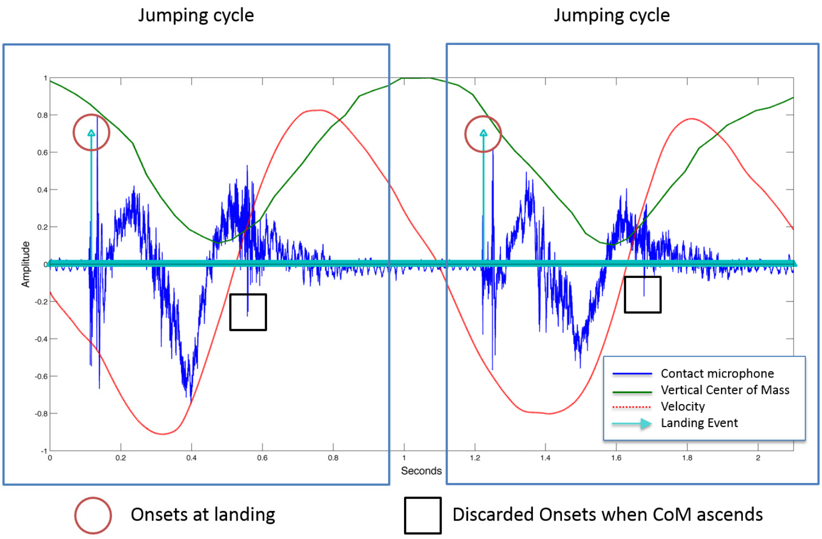


Figure 4.6. Analysis of the input signals to detect the phases of one jump cycle.

4.4.2 Sound Synthesis

Compared to Publication IV in which a realistic rendering of the sound produced by the contact of the feet with the trampoline was proposed, in the case of the sonic trampoline sound is more directly used to provide an auditory feedback of the motion of the jumper. The goal was to induce a direct effect on motor behavior.

The interactive sonification technique proposed here belongs to parameter-mapping sonification (PMSon), "which involves the mapping of data features onto acoustic parameters of sonic events (such as pitch, level, duration, and onset time)" [61, p. 289].

As such the interactive sounds can be freely designed as long as the action-sound mappings stays consistent. The importance of parameter mapping in electronic instrument design has been thoroughly investigated within the musical context [66]. Also, if the mapping is based on familiar metaphors, the feedback might result more effective to guide the action. Such metaphors can be informed by physics (the more energy is put in the movement the louder the sound) or be embodied metaphors [80], i.e. supporting high level abstract concepts by extending low level body schemata (see also Section 2.2).

The potential benefits of incorporating embodied metaphors in the design of an augmented audio environment have been shown by [2]. The authors demonstrated how conditions in which embodied metaphors were used in the mapping layer movement-sound parameter led to higher performance and accuracy. Their recommendations indicate that the use of embodied mapping can provide some advantages to learning and usability aspects of an interface.

With the sonic trampoline I wanted to explore different mapping strategies for the interactive sonification of jumping simulating the iconic 8-bit jumping sound from the Nintendo game *Super Mario Bros*, but made it interactively generated and controlled during each jumping cycle. A periodic wave resembling the original waveform used in the game was employed and the verticality of pitch [130] as one mapping dimension in the range 98 - 880 Hz explored. The pitch was linearly mapped to the vertical position of the tracked CoM. The speed of the CoM was analyzed to determine the moments of maximum height reached in each jump and then used to shape the envelope of the sound and force it to be silent during the downward motion. Also, this nominal range was compared with exaggerated and compressed versions of the ranges of pitch for the same jumping height. The 1:0.5 mapping resulted in a compression of 50 percent of the pitch range while the 1:1.5 mapping in an expansion centered in the fixed frequency.

In the user study the participants were let to freely explore the sonic interaction during the different conditions without assigning any tasks. Results showed that different mapping conditions were responsible for the different jumping behaviors measured in terms of achieved height and different emotional responses among no-sound versus with-sound conditions. The amplified and ascending pitch mapping gave the highest measured performance (Figure 4.7) and most pleasant and exciting self-reported emotional involvement (Figure 4.8). Results showed that interactive sonification motivates the user to jump higher and that different mapping conditions and pitch range alter the jumping height. On top of these behavioral effects, the experience of jumping was more exciting owing to the designed interactive sonification and game inspired sound choices. The use of game sounds to augment the trampoline is an example of a sonically extended commodity as proposed by Hug [65] in the sound design oriented typology of interactive commodities, which is also taking advantage of a strong *brand* such as the one of the famous platform game.

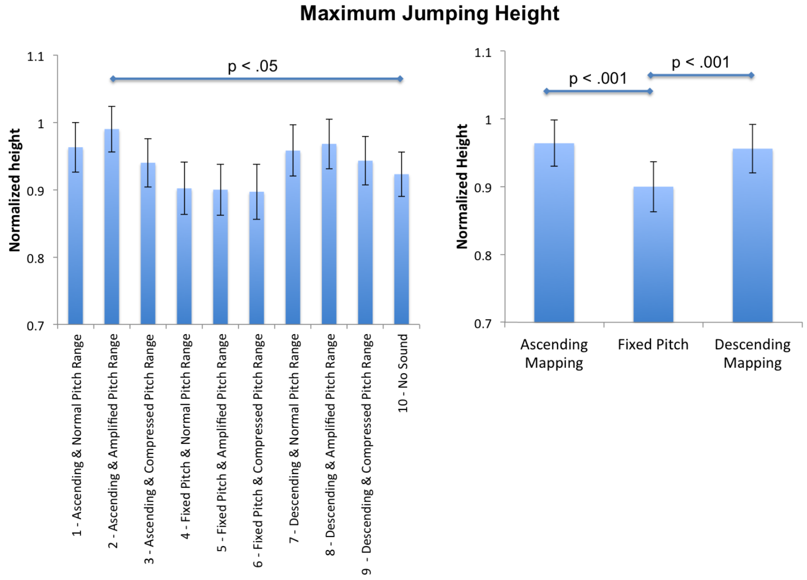


Figure 4.7. Average of the peak jumping heights in different conditions. Error bars represents 95% confidence intervals.

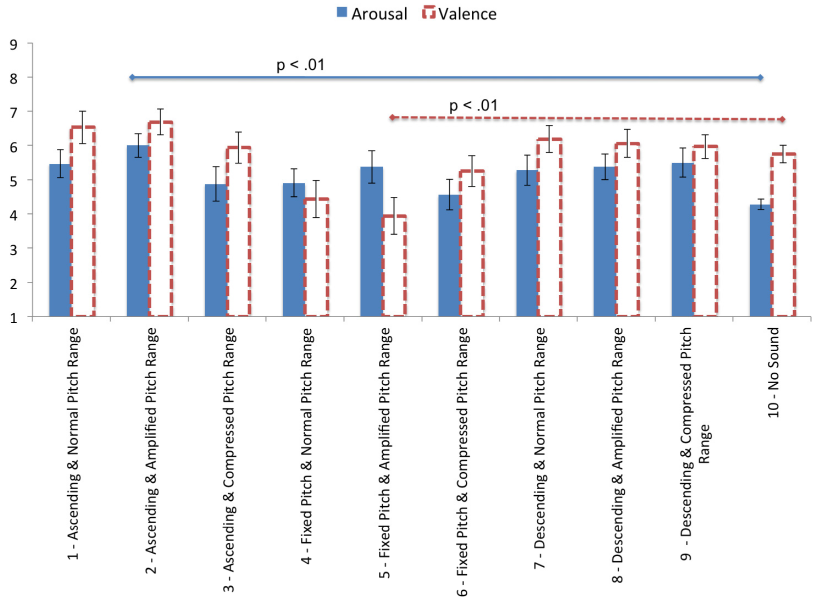


Figure 4.8. Self-assessed Arousal and Valence in the 10 conditions of the user study. Error bars represent standard deviation.

4.5 Mobile-music augmentation

Sometimes introducing sonic interaction to a familiar scenario of human-human interaction can transform people's experience to the extent that it changes the way they conceptualize the experience itself. As in the example of the sonic trampoline of Publication VI, interactive sonification can bring an extra fun-factor to the activity turning training into a game. The ability to make taxing tasks more engaging and interesting to learn already justifies the use of movement sonification, but what about a scenario of collective music improvisation?

This research question inspired the system development and user study of Publication III in which sound is used to augment the social dimension of collaborative music making. Players wear speakers on their chest and use mobile phones as musical instruments. The music controller designed is a mix of a gestural controller and sonification of movement. Players can create sound independently from other players by shaking their phone and the sound is reproduced through the speaker they wear. Nevertheless, some sound parameters are the result of relational parameters that provide an indication of the group dynamics and social interaction while playing together.

In the user study the distance between the players was taken into consideration. By tracking the horizontal position of the participants in the space, the perimeter of the triangle formed by the participants inversely controls the distortion amplitude applied to the instrument. This aspect of the sound is thus democratically and bodily controlled by the group and needs to be negotiated while being involved in a collective action-sound loop. The scenario described is a minimal realization of a participatory enactive sonic interaction.

4.5.1 Qualitative evaluation

Qualitative methods are often employed for evaluation when the phenomena under consideration are particularly elusive and measuring observable quantities or asking specific questions will not be able to represent it fully. Also, asking specific questions already poses constraints on the conceptualization of the experience since it defines the dimensions along which the experience is articulated and decomposed in order to be measured. Music-making in its richness, expressivity and collaborative nature is one of these scenarios and researchers have proposed a variety

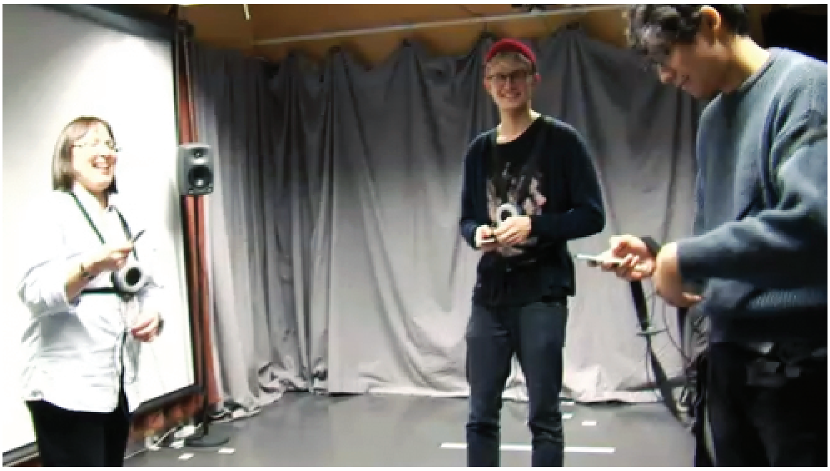


Figure 4.9. A still from the footage recorded in one user study session

of techniques and methods to assess new musical interfaces. In particular the approach proposed by Johnston [72] was followed and most of his methods was borrowed, namely the group-session, the semi-structured interview and the data analysis. The approach proposed here is a mix of qualitative data gathering and a Modified Stimulated Retrospective Think-Aloud (MSRTA), a usability method that collects the verbalization of a user's performance after the performance is over [53], which has been recommended in the musical context by [95].

Data analysis was conducted using techniques from the grounded theory tradition [26]. The aim of a grounded theory is "to build a theoretical explanation by specifying phenomena in terms of conditions that give rise to them, how they are expressed through action/interaction, the consequences that result from them, and variations of these qualifiers." [26, p. 9].

First, all the footage, that is the data of the experiment, is analyzed and all the relevant incidents are labeled (open-coding phase). The analyst prunes redundant entries and reaches a smaller number of concepts associated with the data. The second phase, axial-coding, is about finding relations among the data. The analyst searches for relationships among categories and sub-categories. Through continuous comparison of incidents in the data, the process tends to converge towards a theory.

Under the core-category of *experience*, the categories emerging are presented in Figure 4.10 that shows that one of the concepts encountered is the missing link and, for some participants, the dichotomy between music-making and game. This theory emerging from the data would probably

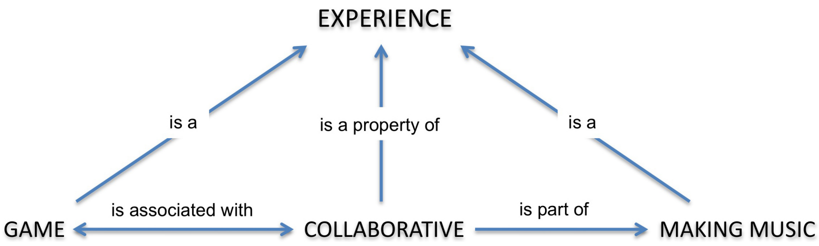


Figure 4.10. The experience conceptualized by the participants resulting from the analysis of the interviews. Core categories and their relationship are presented.

have remained unnoticed in a standard questionnaire. Further relations among categories can be inferred from the data and these are showing the responsible factors of this *shift-of-focus* from music-making to bodily and situated engagement (Fig. 4.11).

Our sonic augmentation design approach to use relational descriptors to create musical affordances proved successful in collectively engaging the players. Further studies could focus on similar scenarios where players are already proficient with their own musical instrument. This might help merging a collective game experience with the musical one.

4.6 Discussion

Sound has been used as feedback to sports activities in Publication IV and Publication VI in the case of jumping on a trampoline. The contribution of the first publication is generalization of feet-trampoline interaction results already found in the case of physically-based synthesis of contact and walking sounds. I found that similar techniques can indeed affect the perception of compliance of the trampoline surface. This original result could be used to manipulate the experience of jumping on a trampoline during an exercise activity.

A different approach inspired by sonic interaction design and gamification was used in Publication VI. I created an original system to track the position of the jumper and analyzed the jumping cycle. The sonification scheme devised proved successful in motivating the user to jump higher by mapping the pitch of the sound with the vertical position of the Center of Mass, in this way showing the impact of interactive sonification on performance. Also the emotions of the jumper were positively affected by the presence of interactive sounds. Along the line of current research in interactive sonification, motion-controlled sounds open up a rich scenario

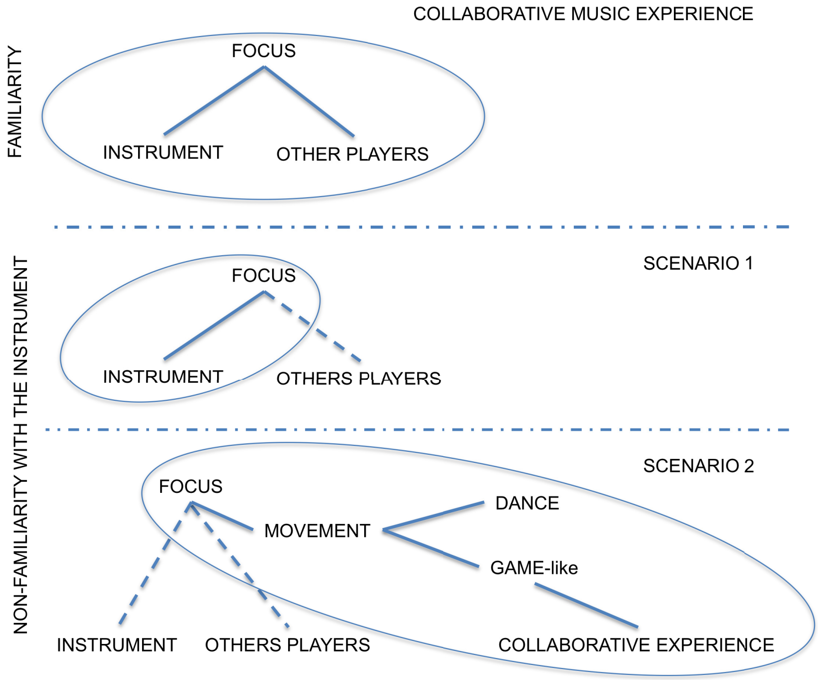


Figure 4.11. Relationship between shift of focus in activity and conceptualization of the experience. Once the player is familiar with the instrument, the focus during the activity is managed across the action on the instrument and coordinating with other players by focusing on their activities. If the player is presented with an instrument he is nor familiar with, two scenarios occur: either the person focus on his own music playing and the experience ceases being collective. Alternatively the player focuses on the movements of other players and the experience becomes more similar to dancing or game-like, but still remain collaborative and based on social interaction.

of opportunities where task performances can be aided by using appealing sound design to motivate the user and to keep their interest high during the task.

The most musical of the case studies was the one on augmentation of mobile music improvisation in Publication III. A scenario of participatory sonic interaction was designed using mobile phones as gestural music controllers. Some aspects of the sound were left to be controlled collectively. For this a Kinect-based tracking system controls the distortion applied on the sound as a function of the distance among the players letting them negotiate this sound parameter as a group. In this way the players explore mediated social interaction while making music together. The research contributed to designing sonic interaction mainly from a methodological point of view. Indeed, the focus of the paper was neither the sound synthesis nor the system for collectively controlling some sound parameters but rather an investigation of the experience emerging from playing together avoiding a priori assumptions or imposed constructions. The qualitative methodology proposed can easily be applied to other contexts of HCI and embodied interaction where it might be too risky to reduce the whole phenomenon to few measurements or a generic questionnaire.

In Figure 4.12 the three publications presented in this chapter are abstracted in terms of the sensorimotor coupling model of enaction.

The concept of perceptually guided action is crucial for all instances of interactive and movement sonification as in Publication IV and Publication VI. These two publications present cases of mediated closed auditory action-perception loop. The sonification directly and implicitly shapes the user's perceptual activity in a jumping task.

Enaction in social interaction is the context of Publication III. With respect to the schematic representations of Figure 2.1, we can think of a sonic environment participants create with their actions. Musicians are involved in dynamics of coordination that they can facilitate or disrupt with their bodily activity. In a non-augmented scenario of music making, the sound (signal) guides the action of every other agent but also the action of the individual becomes a signal for the other participants. The presence of technologically mediated musical instruments creates a new sensorimotor coupling between the participants since the value of certain sound parameters are the results of collective action. In the case of mediation presented here, the influence of one participant's action on other participants' becomes more apparent because the dynamics of group af-

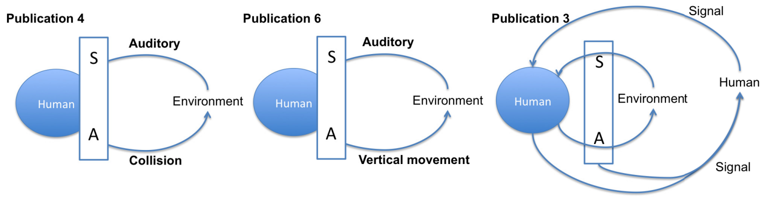


Figure 4.12. Schematic representation of the sensorimotor coupling designed in the publications PIV, PVI and PIII. Publication III is a case of participatory enaction (human-human interaction) in the context of music making with mobile phones. The mediating technology in the form of sounds manipulated as a group adds another layer of sensory-action contingencies in the human-environment loop. Patterns of coordination are influenced by the added structural coupling among participants (their relative distance) of the music playing. Publication IV and Publication VI are instances of interactive and movement sonification applied to the scenario of trampoline jumping, where enaction is present in the form of added action-sound coupling.

fects the sound of the individual's instrument resulting in new patterns of coordination of human movement and sound.

5. Avatars and Mirrors

This chapter describes the content of Publication I and Publication VII and the related works. The research presented in these publications is concerned with visual feedback projected in front of the participant. The case studies are instances of embodied interaction without interfaces; only the body is used as the communication channel and the quality of the motion and posture function as explicit or implicit signals input to the system.

The first publication explores a scenario of full-body interaction between a human and a virtual character where the only communication channel is the perception of each other's action and the reaction to these. The second publication investigates body mediation when a player controls an avatar. The virtual character does not only mirror the player's movement but also shows different postures and a slow response to player's movement. The mechanism requires the player to adapt to the avatar's body in order to control a game.

5.1 Full-body interaction with virtual characters

In Publication I I wanted to examine the role of bodily interaction with a virtual character that shows a repertoire of behaviors activated in reaction to the behavior of the participant. This two-way flow of action and reaction is an enactive loop sustained by two agents, the human and the virtual character, by bodily movement.

I have already mentioned the difference in information processing during an embodied experience compared to a mere observation of a phenomena in terms of attitudes, social perception, and emotion in Section 2.1. I have also referred to participatory enaction where sense-making emerges from the interaction among individuals.

To experiment with the enactive loop a system was created where a motion-captured participant moves freely in front of a screen where a virtual character is represented as a stick-figure (see Figure 5.1). The action of the participant are perceived by the virtual character that can in turn react performing basic motions such as walking, jumping or waving hands according to authored rules. As opposed to scripted behavior and turn-based interaction [120], the system is designed for continuous synthesis of action response of the virtual character.

Previous work in the field of expressive gesture in music and dance performances [20] has described motion in terms of Quantity of Motion (QoM), contraction index and many other descriptors. The experiment was limited to the QoM and the distance between the human and the virtual character. Within this two-dimensional space of features the rules are implemented as a mapping between input motion features extracted from the human motion and output motion features (see Figure 5.2). Using the participant's motion observed by the virtual character as input, the system produces as result the motion that best approximate the desired output features. The result motion is chosen from a database of recorded animations including walking, jumping, and running, segmented into one-second long clips.

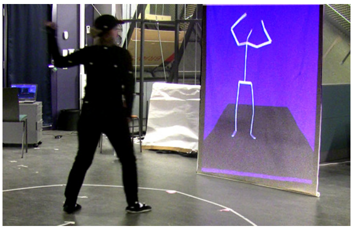


Figure 5.1. Live enaction from Publication I

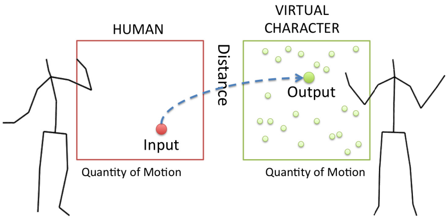


Figure 5.2. An example of authored rules for the virtual character

The rules were manually annotated defining input-output point pairs in the space of QoM and distance. This process has been extended with more motion descriptors and automated by defining the rules through recorded actions and reactions [43]. The motion is created using graph-based motion synthesis ([77]) and searching for the recorded motion best fitting the desired values of the motion descriptors.

Even with a limited amount of motion descriptors the system allowed for the creation of virtual characters exhibiting different behaviors and attitudes. Seven conditions were designed by specifying point-pairs in the space of QoM and distance labeled for the sake of convenience as the following: "cry baby", "easily scared", "calm imitation", "regular imitation", "hyperactive imitation", and "random". The user study consisted of two minutes of free interaction with each character, after which the participants were asked to evaluate the behavior of the virtual character using a predetermined set of adjectives, such as "scared", "shy", "aggressive", "childish", "curious", "bored", "joyful", "indifferent". Results showed that participants interpreted as "scared" a character that was stepping backwards if the participant would approach it with fast movements. Another class of characters mentioned was the aggressive one, which would show mainly movements with high QoM and a tendency to approach the participant. Also, participants clearly imitated the behavior of the virtual character and engaged more in bodily interaction when the virtual character showed more active behavior and less when the virtual character acted more passively. While a more quantitative and rigorous methodology should be adopted to study the phenomenon in more depth, a simple case of enaction was successfully designed.

5.2 Impersonating a different character

Research in media psychology has shown the impact of impersonating an avatar in a Virtual Environment (VE), a digital representation of ourselves, on the feeling of oneself and on one's action in that environment. A situation when the user's virtual self-representation is dissimilar to the physical self [127], for instance a more or less attractive version of oneself, a thinner or heavier, a taller or shorter, can produce different attitudes towards others [105]. Also known as the Proteus Effect, this phenomenon can induce alterations of traits of the personality that extend outside the digital world of the avatar and also promote virtuous behaviors for instance more healthy eating behavior or exercising more ([45], [46]).

Related to the Proteus Effect are concepts such as self-presence and identification. Self-presence is defined as the experience of feeling one-self in a VE ([83], [71],[101]). Self-presence is further analyzed by Ratan [100] and encompasses three layers: core-presence, which is the body-level of self-presence or "the extent to which a mediated self-representation is integrated into body schema", proto-presence that is "the extent to which mediated interactions between a self-representation and mediated objects cause emotional responses" and extended self-presence, which accounts for the relation between aspects of personal identity and the self-representation.

According to Klimmt et al. [24], "from the perspective of social psychology, identification is defined as a temporary alteration of media users' self-concept through adoption of perceived characteristics of a media person". The same authors affirm that an interactive setting like a game can boost this phenomenon since the game character is not an entity completely separate from the players themselves but rather a merging of their own self and the protagonist of the game. Both self-presence and identification express the degree of relation and bonding between a person and an avatar; the higher the self-presence and identification the more the player can be influenced by the behavior of her virtual self's and infer her own features from that (self-perception theory [6]), affecting their self-image and also attitudes and future behavior [44].

Once again I am interested in what the embodiment brings to this process of identification. For the Proteus effect, embodiment produces significantly larger behavioral changes than mere observation of the same visual stimuli [128]. But what would happen if a person needed to actively adapt to the features of the avatar? What if the body of the character would offer constraints and weaknesses the person does not possess? What if the person needed to move differently than she would ordinarily do in order to control the body of the avatar? Does adapting to another body and getting control over it affect our feelings and behavior inferred from the movement capabilities of the new body?

I am interested in designing a situation where the goal can be achieved only by enacting the body of the avatar, and where the player understands the avatar she embodies through enaction.

5.3 Enacting a weak character

Publication VII continues investigating the role of the body in mediating the experience and it does so by offering the player the possibility of controlling in full-body interaction a character that does not follow themselves completely. More specifically, the character tends to lag behind and the player needs to adapt to this different body by matching the character pose if she does not want to lose control of it. The avatar follows the movement of the player only as long as the player waits for the avatar

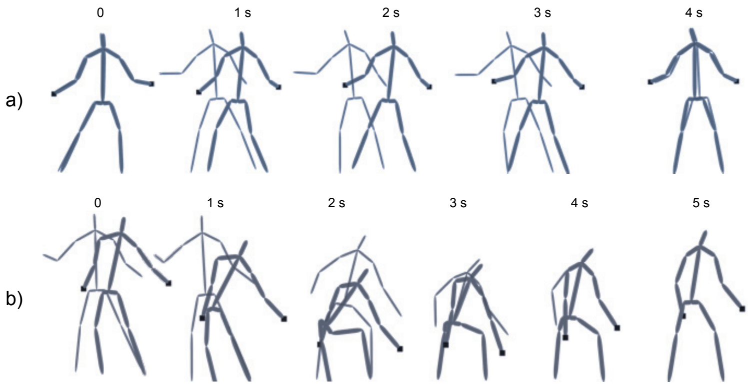


Figure 5.3. Original (thin limbs) and avatar (tick): in a) the control of the avatar is partially lost due to a lateral fast movement. The player needs to wait for the avatar. In b) the avatar starts falling down and the player needs to adapt to the crouched posture.

Figure 5.4. Subject playing the bubble-popping mini-game (left) and screen-grab from gameplay (right).

to follow the movement. This process stimulates the adaptation of the player to the pace of the avatar in a continuous feedback of push-pulling the avatar, which is a simple instance of enaction where the interface is the body of the avatar controlled by the player.

A physics-based virtual character was created with joints affected by attraction forces pulling towards the position of the motion-tracked player, forces that push the character down to a crouched position and forces that make the character follow the player but only within a certain distance from it (Figure 5.3). By combining these forces one can design weak characters that the player can impersonate and embody.

Previous studies have considered the influence of posture on attitude and behavior [103] and more recently in the game context by [8]. In this study I am interested in the dynamic coupling between player's and avatar's bodies. As opposed to a mirror case, conditions that alter the coupling were designed: firstly by statically changing the posture of the avatar into a crouched position and secondly by modulating the coupling (tighter and looser) with the distance between the player's and avatar's joints position. This second technique is directly informed by sensorimotor theory of experience [96] and particularly addresses sensorimotor coordination, which is a contingency described by co-dependencies between sensory and motor activity that reliably contribute to functionality [18]: the controller mechanism is an added sensorimotor contingency the player learns during the gameplay and will ultimately master, i.e. the player knows the laws that govern what happens when she does all the things she can do when controlling that character. Following Nöe [91], we can say that the feel of being that character is "to be ‘attuned’ to the ways in which one’s movements will affect the character of input".

For the experiment a motion-controlled mini-game was designed where a character on the screen needs to pop soap bubbles as fast as possible using its hands only (Figure 5.4). Six conditions were created to explore the effect of the posture of the avatar and the avatar’s response to the player’s movement. Self-presence, identification and game experience using the Game Experience Questionnaire [67] were chosen as dependent variables together with the general impression or features the player saw in the avatar: "Which of the following best describes your impression of the avatar? (weak, strong, angry, sad, happy, other)".

While self-presence and identification did not show higher score for conditions where coordination and adaptation of the player to the avatar were needed, the designed enaction-based game brought forth interesting results.

Players moved faster in conditions where they felt more in control, spent more times in a crouched position in the conditions where the character risked to fall into a crouched posture if not controlled and adapted their movement to counteract the response of the controlled body.

The different conditions offered, as expected, different levels of challenge but the flow, i.e. the state of optimal experience characterized by the balance between challenge offered and skill required to accomplish a task ([27], in computer games research [110]) did not score that differently in the conditions where the character was falling if the player was not deliberately adapting her movement to the avatar’s body. Also, the condition that required stronger adaptation was more often associated with a weak character. This suggests that the effort required for controlling the character induces an experience different from just observing or controlling a character with different posture.

Overall the controller mechanism proposed fulfills the intended design of portraying a weak character that can be enacted by the player in full-body interaction.

5.4 Discussion

In this chapter two designs of full-body enaction were presented. Our approach extends the interaction with virtual agents to full-body embodied action and can find application in motion-controlled narrative games and implementations of interactive agents.

In Publication I a framework and implementation of interaction with a virtual character that is able to react according to authored rules to the action and motion style of a human was presented. The implementation showed limitations from an animation standpoint as the virtual character was not responsive at all times due to constrained clip length. The reaction time was consequently harming the flow of interaction. Also, the design space of authoring rules was limited to two dimensions that by themselves are far from representing the richness of a real encounter with another being.

However, these shortcomings could easily be overcome using more sophisticated animation techniques and a higher dimensionality of the extracted motion features while keeping the concept of action-reaction rules present. By combining and applying meta-rules that let the virtual character shift from one rule to another we can increase the complexity of the logic used by the character. The future of gaming based on affordable motion controllers like Kinect could be leveraged on a more physical AND emotional involvement with game characters "reading" the intention of the player from their bodily attitudes and responding meaningfully to that. As a step beyond that, the personality of a game character is revealed during the embodied interaction of the player in the game world, less scripted and emergent from the situation.

The second publication presented in this chapter brings the space of interaction with the virtual character closer to the skin of the player, providing an avatar with different physical abilities that can be embodied and controlled. While other studies of motion-controlled games have been focusing on engagement and found high correlation of this with the player's motion ([9], [7]), I have focused on a controller mechanism able to convey and transfer features and attitudes of the avatar controlled. In so far only exaggeration of a player's skills and consequent empowerment has been proposed ([55], [74]). Our approach aims at promoting identification with different, possibly impaired characters - a weak avatar in Publication VII, obtaining a strong psychological bond mediated and informed by the body involvement of the player. These could be extended to a variety of physical and psychological conditions the player could identify with. The results are promising but still preliminary. In the studies investigating the Proteus effect ([45], [46]), the experiments aimed also at validating the hypothesis that certain behavior observed in the avatar will indeed transfer to the participant, such as changes in eating behaviors or will to exercise. The experiment did not include a measure of how the experience of different conditions extends into different behavior post-game: will the player feel weaker because of the avatar controlled? This question remains open and will probably require a much richer game scenario where goals, player motivations and the consequences of actions are at stake.

In Figure 5.5 the two publications are represented in terms of the sensorimotor coupling model of enaction.

A situated full-body involvement in an encounter with a virtual character is described in Publication I. The action of the person lets them discover lawful relations at the basis of a dynamic coupling between her actions and the reactions of the virtual character. This is still far from the autonomous emergent behavior that an enactive account of social interaction prescribes [30] but puts emphasis on a continuous flux of action and perception in interaction.

Publication VII implements a case that lies in between an augmentation of the person's body and enaction with another agent. The coupling between the two bodies follows a non-trivial physics-based rule that requires the person to actively search for the "grip" of the body of the avatar in order to control it. The experience oscillates between the case of perfect coordination and full embodiment and the case of unstable coordination with a separate agent through adaptation.

Finally, a common reductionist approach was used in both studies: the avatar is a visual object and only the visual modality is used as a channel of interaction. Embodiment is the result of multimodal sensorimotor coordination, thus quite obviously a character with more senses should be implemented. Already creating an auditory action-perception loop for the character (a character turning towards sounds coming from the player's living room or being able to react to the atmosphere of a soundscape) would make its behavior more lively and believable.

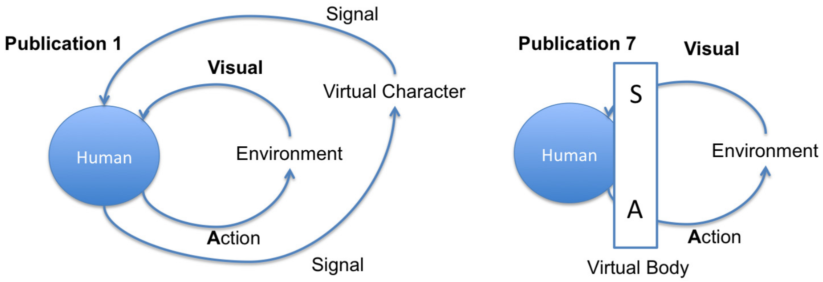


Figure 5.5. Schematic representation of the sensorimotor coupling designed in the publications PI and PVII. Publication I presents the case of a human full-body involvement in an encounter with a virtual character. In Publication VII the human enacts a character with different movement capabilities (weak character).

6. Conclusions and future direction

In this thesis I have demonstrated original case studies of embodied interaction and enaction considering multiple modalities such as visual, auditory and tactile. I have placed emphasis on natural interaction based on embodied metaphors and putting the body and its capabilities at the center of the interaction. The research behind this thesis has produced real-time systems that allowed for evaluating different scenarios of full-body interaction. I focused on sonic interaction, bridging the field of auditory display with HCI and contributing to it with the use of tangible, spatial and also tactile interaction. In interactive sonification I contributed to incorporating sound in physical exercise showing how this can improve perception and experience. In gaming I have shown ways of letting the body participate in the game experience both physically and psychologically by designing embodied interaction and enaction.

In building up this series of case-studies I came to learn that enactive technology can be used to naturally guide the participant during the experience. Rather than focusing on given functionalities needed by the user, a technology that manipulates the pre-existent natural sensorimotor loop of the person in the environment can lead to improved physical performance. This has been shown in the case of interactive sonification and audio-tactile feedback and this line of research is currently explored by the interactive sonification and sonic interaction community. The case studies I proposed can be adopted in the context of sport or rehabilitation in aiding patients to reeducate motor skills.

But a full-body interaction approach can also act as facilitator of psychological bond among people and virtual characters. Identification with a virtual character is a powerful way to induce positive behavior in humans or just to create more dramatic or memorable experiences in entertainment. Full-body enaction scenarios, as opposed to indirect controlling of avatars using graphical user interfaces, should be studied further to understand how the body mediates the sense of being that character in the virtual world and the transfer of emotions among the participant and the avatar.

Virtual agents that are more aware of the person's avatar action are a current mission of the community researching embodied virtual agents. We can envision an interaction paradigm where more communication among agents arises from the presence of an active human body in the virtual world. If we can design virtual agents able to respond with believable and sensible behavior in both time and space we might make their mission of aiding humans more successful.

I can conclude that if enaction is perceptually guided action, enaction technology is mediation technology capable of guiding the participant at a physical, psychological and emotional level. It is not that surprising that all these layers of experience can be accessed since our main hypothesis is that cognition is embodied and our senses bring forth the experience of the world.

It is necessary to pursue a multimodal approach, in which the extended world accessible through enaction needs to involve the different senses in a way that possibly all the channels of perception (auditory, tactile, visual, if not olfactory or gustatory) are together at work in the enactive loop. At least with respect to common human modalities like auditory, tactile and visual this thesis shows concrete implementations to build upon full-body interaction.

The consequence of referring to the embodied cognition and enaction approach is to start designing interaction from the body for the body, immediately thinking about the motor skills of the participant, the senses at work and the action-perception loops involved in the activities.

The concepts inspiring this thesis are not new: several are the research communities investigating themes such as Embodied Interaction and Tangible Interaction (TEI), embodied agents, movement and sound, sports and games. The more specific, in-depth and narrow research is carried out, the more it needs a general framework that does not let one lose the big picture.

This thesis contributes to the effort of combining modalities, integrating technologies and interaction techniques coming from different contexts and looks at the human as a holistic multisensory agent in exploration. I have used enaction as a common framework to guide the design and engineering of the future of media experience. I have shifted the focus from designing interaction for a *user* to creating experiences for a *participant* with a living body and contributed to cases of mediating technology where the interface becomes incorporated in the human mind.

A large part of the research presented here revolves around sound and sonic interaction. The use of sound in interaction has found establishment in the last decade but it is still confined to the field of research. Also, even though present in human-computer interaction design, I feel it is often a non-core feature of the artifacts designed. Sound is often an overlooked medium of interaction that naturally connects the visual and tactile modalities; it can convey information and emotion, it can guide actions and increase presence. Sonic interaction will surely find an increasingly important role in the domain of designing virtual agents, animation and games.

References

[1] Michael L Anderson. Embodied cognition: A field guide. *Artificial intelligence*, 149(1):91–130, 2003.

[2] Alissa N Antle, Greg Corness, and Milena Droumeva. What the body knows: Exploring the benefits of embodied metaphors in hybrid physical digital environments. *Interacting with Computers*, 21(1):66–75, 2009.

[3] Murat Aytekin, Cynthia F Moss, and Jonathan Z Simon. A sensorimotor approach to sound localization. *Neural Computation*, 20(3):603–635, 2008.

[4] Ronald Azuma, Yohan Baillot, Reinhold Behringer, Steven Feiner, Simon Julier, and Blair MacIntyre. Recent advances in augmented reality. *Computer Graphics and Applications, IEEE*, 21(6):34–47, 2001.

[5] Kenneth Baynes and James Bohman. *After philosophy: end or transformation?* Mit Press, 1987.

[6] Daryl J Bem. Self-perception theory. *Advances in experimental social psychology*, 6:1–62, 1972.

[7] Nadia Bianchi-Berthouze. Understanding the role of body movement in player engagement. *Human–Computer Interaction*, 28(1):40–75, 2013.

[8] Nadia Bianchi-Berthouze, Paul Cairns, Anna Cox, Charlene Jennett, and Whan Woong Kim. On posture as a modality for expressing and recognizing emotions. In *Emotion and HCI workshop at BCS HCI London*, 2006.

[9] Nadia Bianchi-Berthouze, Whan Woong Kim, and Darshak Patel. Does body movement engage you more in digital game play? and why? In *Affective Computing and Intelligent Interaction*, pages 102–113. Springer, 2007.

[10] Ned Block. The mind as the software of the brain. *Science Fiction and Philosophy: From Time Travel to Superintelligence*, pages 170–185, 1995.

[11] Aaron Bloomfield and Norman I Badler. Collision awareness using vibrotactile arrays. In *Virtual Reality Conference, 2007. VR'07. IEEE*, pages 163–170. IEEE, 2007.

[12] Marie-Luce Bourguet. Designing and prototyping multimodal commands. In *INTERACT*, volume 3, pages 717–720. Citeseer, 2003.

[13] Oliver Bown. A framework for eco-system-based generative music. *Proceedings of the SMC*, pages 195–200, 2009.

[14] Albert S Bregman. *Auditory scene analysis: The perceptual organization of sound*. MIT press, 1994.

[15] Roberto Bresin, Anna de Witt, Stefano Papetti, Marco Civolani, and Federico Fontana. Expressive sonification of footstep sounds. In *Proceedings of ISon 2010: 3rd Interactive Sonification Workshop*, pages 51–54, 2010.

[16] Stephen Anthony Brewster. *Providing a structured method for integrating non-speech audio into human-computer interfaces*. PhD thesis, Citeseer, 1994.

[17] Jerome S Bruner. The course of cognitive growth. *American psychologist*, 19(1):1, 1964.

[18] Thomas Buhrmann, Ezequiel Alejandro Di Paolo, and Xabier Barandiaran. A dynamical systems account of sensorimotor contingencies. *Frontiers in psychology*, 4, 2013.

[19] Harold E Burtt. Tactual illusions of movement. *Journal of Experimental Psychology*, 2(5):371, 1917.

[20] Antonio Camurri, Barbara Mazzarino, Matteo Ricchetti, Renee Timmers, and Gualtiero Volpe. Multimodal analysis of expressive gesture in music and dance performances. In *Gesture-based communication in human-computer interaction*, pages 20–39. Springer, 2004.

[21] John M Carroll. *Human computer interaction-brief intro*. The Interaction Design Foundation, 2013.

[22] Hans J Charwat. Lexikon der mensch-maschine-kommunikation. 1994.

[23] VG Chouvardas, AN Miliou, and MK Hatalis. Tactile displays: Overview and recent advances. *Displays*, 29(3):185–194, 2008.

[24] Klimmt Christoph, Hefner Dorothée, and Vorderer Peter. The video game experience as “true” identification: A theory of enjoyable alterations of players’ self-perception. *Communication Theory*, 19(4):351–373, 2009.

[25] Perry R Cook. Physically informed sonic modeling (PHISM): Synthesis of percussive sounds. *Computer Music Journal*, pages 38–49, 1997.

[26] Juliet M Corbin and Anselm Strauss. Grounded theory research: Procedures, canons, and evaluative criteria. *Qualitative sociology*, 13(1):3–21, 1990.

[27] Mihaly Csikszentmihalyi and Mihaly Csikzentmihaly. *Flow: The psychology of optimal experience*, volume 41. HarperPerennial New York, 1991.

[28] Hanne De Jaegher and Ezequiel Di Paolo. Participatory sense-making. *Phenomenology and the cognitive sciences*, 6(4):485–507, 2007.

[29] Daniel C Dennett. *Content and consciousness*. Routledge, 2002.

[30] Ezequiel A Di Paolo, Marieke Rohde, and Hanne De Jaegher. *Horizons for the enactive mind: Values, social interaction, and play*, chapter in Enaction: Towards a new paradigm for cognitive science, pages 33–87. MIT Press Cambridge, MA, 2010.

[31] Daniel Dobler, Michael Haller, and Philipp Stampfl. ASR: augmented sound reality. In *ACM SIGGRAPH 2002 conference abstracts and applications*, pages 148–148. ACM, 2002.

[32] Paul Dourish. Embodied Interaction : Exploring the Foundations of a New Approach to HCI. 1999.

[33] Paul Dourish. *Where the action is: the foundations of embodied interaction*. MIT press, 2004.

[34] Sandra E Duclos, James D Laird, Eric Schneider, Melissa Sexter, Lisa Stern, and Oliver Van Lighten. Emotion-specific effects of facial expressions and postures on emotional experience. *Journal of Personality and Social Psychology*, 57(1):100, 1989.

[35] Alfred Effenberg, Joachim Melzer, Andreas Weber, and Arno Zinke. Motionlab sonify: A framework for the sonification of human motion data. In *Information Visualisation, 2005. Proceedings. Ninth International Conference on*, pages 17–23. IEEE, 2005.

[36] Alfred O Effenberg. Movement sonification: Effects on perception and action. *Multimedia, IEEE*, 12(2):53–59, 2005.

[37] Arne Eigenfeldt and Philippe Pasquier. A sonic eco-system of self-organising musical agents. In *Applications of Evolutionary Computation*, pages 283–292. Springer, 2011.

[38] Douglas C Engelbart. Augmenting human intellect: a conceptual framework (1962). *PACKER, Randall and JORDAN, Ken. Multimedia. From Wagner to Virtual Reality. New York: WW Norton & Company*, pages 64–90, 2001.

[39] Mikael Fernstrom, Eoin Brazil, and Liam Bannon. Hci design and interactive sonification for fingers and ears. 2005.

[40] Kenneth P Fishkin, Anuj Gujar, Beverly L Harrison, Thomas P Moran, and Roy Want. Embodied user interfaces for really direct manipulation. *Communications of the ACM*, 43(9):74–80, 2000.

[41] KennethP. Fishkin. A taxonomy for and analysis of tangible interfaces. *Personal and Ubiquitous Computing*, 8(5):347–358, July 2004.

[42] Jerry A Fodor. *The language of thought*, volume 5. Harvard University Press, 1975.

[43] Klaus Förger, Tapio Takala, and Roberto Pugliese. Authoring rules for bodily interaction: from example clips to continuous motions. In *Intelligent Virtual Agents*, pages 341–354. Springer, 2012.

[44] Jesse Fox. *The use of virtual self models to promote self-efficacy and physical activity performance*. PhD thesis, Stanford University, 2010.

[45] Jesse Fox, Jeremy Bailenson, and Joseph Binney. Virtual Experiences, Physical Behaviors: The Effect of Presence on Imitation of an Eating Avatar. *Presence: Teleoperators and Virtual Environments*, 18(4):294–303, August 2009.

[46] Jesse Fox and Jeremy N Bailenson. Virtual self-modeling: The effects of vicarious reinforcement and identification on exercise behaviors. *Media Psychology*, 12(1):1–25, 2009.

[47] John J Foxe. Multisensory integration: frequency tuning of audio-tactile integration. *Current biology*, 19(9):R373–R375, 2009.

[48] Tom Froese, Marek McGann, William Bigge, Adam Spiers, and Anil K Seth. The enactive torch: a new tool for the science of perception. *Haptics, IEEE Transactions on*, 5(4):365–375, 2012.

[49] William W Gaver. What in the world do we hear?: An ecological approach to auditory event perception. *Ecological psychology*, 5(1):1–29, 1993.

[50] William W Gaver. Auditory interfaces. *Handbook of human-computer interaction*, 1:1003–1041, 1997.

[51] James J Gibson. *The Ecological Approach to Visual Perception: Classic Edition*. Psychology Press, 2014.

[52] Jonathan Grudin. A moving target: The evolution of hci. *The human-computer interaction handbook: Fundamentals, evolving technologies, and emerging applications*, pages 1–24, 2008.

[53] Zhiwei Guan, Shirley Lee, Elisabeth Cuddihy, and Judith Ramey. The validity of the stimulated retrospective think-aloud method as measured by eye tracking. In *Proceedings of the SIGCHI conference on Human Factors in computing systems*, pages 1253–1262. ACM, 2006.

[54] Steve Guest, Caroline Catmur, Donna Lloyd, and Charles Spence. Audiotactile interactions in roughness perception. *Experimental Brain Research*, 146(2):161–171, 2002.

[55] Perttu Hämäläinen, Tommi Ilmonen, Johanna Höysniemi, Mikko Lindholm, and Ari Nykänen. Martial arts in artificial reality. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 781–790. ACM, 2005.

[56] Aki Härmä, Julia Jakka, Miikka Tikander, Matti Karjalainen, Tapio Lokki, Jarmo Hiipakka, and Gaëtan Lorho. Augmented reality audio for mobile and wearable appliances. *Journal of the Audio Engineering Society*, 52(6):618–639, 2004.

[57] Vincent Hayward. A brief taxonomy of tactile illusions and demonstrations that can be done in a hardware store. *Brain research bulletin*, 75(6):742–52, April 2008.

[58] Martin Heidegger. *Being and time: A translation of Sein und Zeit*. SUNY Press, 1996.

[59] Thomas Hermann, Oliver Höner, and Helge Ritter. Acoumotion–an interactive sonification system for acoustic motion control. In *Gesture in Human-Computer Interaction and Simulation*, pages 312–323. Springer, 2006.

[60] Thomas Hermann and Andy Hunt. An introduction to interactive sonification. *IEEE multimedia*, pages 20–24, 2005.

[61] Thomas Hermann, Andy Hunt, and John G Neuhoff. *The sonification handbook*. Logos Verlag Berlin, GE, 2011.

[62] Thomas Hermann and Helge Ritter. Listen to your data: Model-based sonification for data analysis. *Advances in intelligent computing and multimedia systems*, 8:189–194, 1999.

[63] John M Hollerbach. Locomotion interfaces. *Handbook of virtual environments: Design, implementation, and applications*, pages 239–254, 2002.

[64] Oliver Höner. Aiding movement with sonification in "exercise, play and sport". In Thomas Hermann, Andy Hunt, and Neuhoff John G., editors, *The Sonification Handbook*, pages 525–553. Berlin: Logos, 2011.

[65] Daniel Hug. Towards a hermeneutics and typology of sound for interactive commodities. In *Proc. CHI Workshop on Sonic Interaction Design*, pages 11–16, 2008.

[66] Andy Hunt, Marcelo M Wanderley, and Matthew Paradis. The importance of parameter mapping in electronic instrument design. *Journal of New Music Research*, 32(4):429–440, 2003.

[67] Wijnand IJsselsteijn, Wouter van den Hoogen, Christoph Klimmt, Yvonne de Kort, Craig Lindley, Klaus Mathiak, Karolien Poels, Niklas Ravaja, Marko Turpeinen, and Peter Vorderer. Measuring the experience of digital game enjoyment. In *Proceedings of Measuring Behavior*, pages 88–89. Maastricht Netherlands, 2008.

[68] Hiroshi Ishii, Ali Mazalek, and Jay Lee. Bottles as a minimal interface to access digital information. In *CHI'01 extended abstracts on Human factors in computing systems*, pages 187–188. ACM, 2001.

[69] Hiroshi Ishii and Brygg Ullmer. Tangible bits: towards seamless interfaces between people, bits and atoms. In *Proceedings of the ACM SIGCHI Conference on Human factors in computing systems*, pages 234–241. ACM, 1997.

[70] Ali Israr and Ivan Poupyrev. Tactile brush: drawing on skin with a tactile grid display. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 2019–2028. ACM, 2011.

[71] Seung-A Annie Jin and Namkee Park. Parasocial interaction with my avatar: Effects of interdependent self-construal and the mediating role of self-presence in an avatar-based console game, wii. *CyberPsychology & Behavior*, 12(6):723–727, 2009.

[72] Andrew Johnston. Beyond evaluation: Linking practice and theory in new musical interface design. *interpretation*, 14:105, 2011.

[73] Veikko Jousmäki and Riitta Hari. Parchment-skin illusion: sound-biased touch. *Current Biology*, 8(6):R190–R191, 1998.

[74] Raine Kajastila, Leo Holsti, and Perttu Hämäläinen. Empowering the exercise: a body-controlled trampoline training game. *International Journal of Computer Science in Sport*, 2014.

[75] Hirokazu Kato, Mark Billinghurst, Ivan Poupyrev, Kenji Imamoto, and Keihachiro Tachibana. Virtual object manipulation on a table-top ar environment. In *Augmented Reality, 2000.(ISAR 2000). Proceedings. IEEE and ACM International Symposium on*, pages 111–119. Ieee, 2000.

[76] Roberta L Klatzky, Dinesh K Pai, and Eric P Krotkov. Perception of material from contact sounds. *Presence: Teleoperators and Virtual Environments*, 9(4):399–410, 2000.

[77] Lucas Kovar, Michael Gleicher, and Frédéric Pighin. Motion graphs. In *ACM transactions on graphics (TOG)*, volume 21, pages 473–482. ACM, 2002.

[78] G Johnson Lakoff. M.(1980) metaphors we live by. *ChicagoLondon: University of ChicagoPress*, 1985.

[79] George Lakoff and Mark Johnson. *Philosophy in the flesh: The embodied mind and its challenge to western thought*. Basic books, 1999.

[80] George Lakoff and Mark Johnson. *Metaphors we live by*. University of Chicago press, 2008.

[81] Anatole Lécuyer. Simulating haptic feedback using vision: A survey of research and applications of pseudo-haptic feedback. *Presence: Teleoperators and Virtual Environments*, 18(1):39–53, 2009.

[82] Susan J Lederman. Auditory texture perception. *Perception*, 8:93–103, 1979.

[83] Kwan Min Lee. Presence, explicated. *Communication theory*, 14(1):27–50, 2004.

[84] Paul Lemmens, Floris Crompvoets, Dirk Brokken, Jack Van Den Eerenbeemd, and G-J De Vries. A body-conforming tactile jacket to enrich movie viewing. In *EuroHaptics conference, 2009 and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems. World Haptics 2009. Third Joint*, pages 7–12. IEEE, 2009.

[85] David K McGookin and Stephen A Brewster. Understanding concurrent earcons: Applying auditory scene analysis principles to concurrent earcon recognition. *ACM Transactions on Applied Perception (TAP)*, 1(2):130–155, 2004.

[86] Stefano Delle Monache, Pietro Polotti, and Davide Rocchesso. A toolkit for explorations in sonic interaction design. In *Proceedings of the 5th Audio Mostly Conference: A Conference on Interaction with Sound*, page 1. ACM, 2010.

[87] Erik Myin. An account of color without a subject? *Behavioral and Brain Sciences*, 26(01):42–43, 2003.

[88] Saskia K Nagel, Christine Carl, Tobias Kringe, Robert Märtin, and Peter König. Beyond sensory substitution—learning the sixth sense. *Journal of Neural Engineering*, 2(4):R13, 2005.

[89] Paula M Niedenthal, Lawrence W Barsalou, Piotr Winkielman, Silvia Krauth-Gruber, and François Ric. Embodiment in attitudes, social perception, and emotion. *Personality and social psychology review*, 9(3):184–211, 2005.

[90] Matti Niinimäki and Koray Tahiroglu. Ahne: a novel interface for spatial interaction. In *CHI'12 Extended Abstracts on Human Factors in Computing Systems*, pages 1031–1034. ACM, 2012.

[91] Alva Noë. The origin of "feel". In *From Animals to Animats 7: Proceedings of the Seventh International Conference on Simulation of Adaptive Behavior*, volume 7, page 27. MIT Press, 2002.

[92] Alva Noë. *Action in perception*. MIT press, 2004.

[93] Rolf Nordahl. Increasing the motion of users in photo-realistic virtual environments by utilising auditory rendering of the environment and ego-motion. In *PRESENCE 2006: The 8th Annual International Workshop on Presence*, pages 57–63, 2006.

[94] Peter Bennett Sile O'Modhrain. Towards tangible enactive-interfaces. In *4th International Conference on Enactive Interfaces 2007*, page 37, 2007.

[95] Sile O'modhrain. A framework for the evaluation of digital musical instruments. *Computer Music Journal*, 35(1):28–42, 2011.

[96] J Kevin O'Regan and Alva Noë. A sensorimotor account of vision and visual consciousness. *Behavioral and brain sciences*, 24(05):939–973, 2001.

[97] Sandra Pauletto and Andy Hunt. Interactive sonification of complex data. *International Journal of Human-Computer Studies*, 67(11):923 – 933, 2009. Special issue on Sonic Interaction Design SI: Sonic Interaction Design.

[98] Ivan Poupyrev, Desney S Tan, Mark Billinghurst, Hirokazu Kato, Holger Regenbrecht, and Nobuji Tetsutani. Developing a generic augmented-reality interface. *Computer*, 35(3):44–50, 2002.

[99] Ville Pulkki. Virtual sound source positioning using vector base amplitude panning. *Journal of the Audio Engineering Society*, 45(6):456–466, 1997.

[100] Rabindra Ratan. Self-presence, explicated: Body, emotion, and identity. *Handbook of Research on Technoself: Identity in a Technological Society: Identity in a Technological Society*, page 322, 2012.

[101] Rabindra Ratan, Michael Santa Cruz, and Peter Vorderer. Multitasking, presence, and self-presence on the wii. In *Proceedings of the 10th Annual International Workshop on Presence*, pages 167–190, 2007.

[102] Charlotte M Reed and Lorraine A Delhorne. Current results of a field study of adult users of tactile aids. In *Seminars in Hearing*, volume 16, pages 305–315. Copyright© 1995 by Thieme Medical Publishers, Inc., 1995.

[103] John H Riskind and Carolyn C Gotay. Physical posture: Could it have regulatory or feedback effects on motivation and emotion? *Motivation and Emotion*, 6(3):273–298, 1982.

[104] Davide Rocchesso, Roberto Bresin, and Mikael Fernström. Sounding objects. *MultiMedia, IEEE*, 10(2):42–52, 2003.

[105] Robin S Rosenberg, Shawnee L Baughman, and Jeremy N Bailenson. Virtual superheroes: Using superpowers in virtual reality to encourage prosocial behavior. *PloS one*, 8(1):e55003, 2013.

[106] Jonghyun Ryu and Gerard Jounghyun Kim. Using a vibro-tactile display for enhanced collision perception and presence. In *Proceedings of the ACM symposium on Virtual reality software and technology*, pages 89–96. ACM, 2004.

[107] Marlon Schumacher and Jean Bresson. Spatial sound synthesis in computer-aided composition. *Organised Sound*, 15(03):271–289, 2010.

[108] Stefania Serafin, Luca Turchet, and Rolf Nordahl. Extraction of ground reaction forces for real-time synthesis of walking sounds. *Audio Mostly*, pages 99–105, 2009.

[109] John Stewart. Foundational issues in enaction as a paradigm for cognitive science: From the origin of life to consciousness and writing. *Enaction: Toward a new paradigm for cognitive science*, pages 1–31, 2010.

[110] Penelope Sweetser and Peta Wyeth. Gameflow: A model for evaluating player enjoyment in games. *Comput. Entertain.*, 3(3):3–3, July 2005.

[111] Ana Tajadura-Jiménez, Maria Basia, Ophelia Deroy, Merle Fairhurst, Nicolai Marquardt, and Nadia Bianchi-Berthouze. As light as your footsteps: Altering walking sounds to change perceived body weight, emotional state and gait. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, pages 2943–2952, New York, NY, USA, 2015. ACM.

[112] Andrés Villa Torres, Viktoria Kluckner, and Karmen Franinovic. Development of a sonification method to enhance gait rehabilitation. In *In Proc. of the Interactive Sonification Workshop*, volume 4, 2013.

[113] Koji Tsukada and Michiaki Yasumura. Activebelt: Belt-type wearable tactile display for directional navigation. In *UbiComp 2004: Ubiquitous Computing*, pages 384–399. Springer, 2004.

[114] Luca Turchet, Rolf Nordahl, Stefania Serafin, Amir Berrezag, Smilen Dimitrov, and Vincent Hayward. Audio-haptic physically-based simulation of walking on different grounds. In *Multimedia Signal Processing (MMSP), 2010 IEEE International Workshop on*, pages 269–273. IEEE, 2010.

[115] Luca Turchet, Stefania Serafin, and Paola Cesari. Walking pace affected by interactive sounds simulating stepping on different terrains. *ACM Transactions on Applied Perception (TAP)*, 10(4):23, 2013.

[116] Luca Turchet, Stefania Serafin, Smilen Dimitrov, and Rolf Nordahl. Physically based sound synthesis and control of footsteps sounds. In *13th International Conference on Digital Audio Effects (DAFx-10)*, pages 161–168, 2010.

[117] Brygg Ullmer and Hiroshi Ishii. Emerging frameworks for tangible user interfaces. *IBM systems journal*, 39(3.4):915–931, 2000.

[118] Jan BF Van Erp and HAHC Van Veen. A multi-purpose tactile vest for astronauts in the international space station. In *Proceedings of eurohaptics*, pages 405–408. Dublin, Ireland: ACM, Press, 2003.

[119] Francisco J Varela, Evan Thompson, and Eleanor Rosch. *The embodied mind*. CogNet, 1992.

[120] Hannes Vilhjálmsson, Nathan Cantelmo, Justine Cassell, Nicolas E Chafai, Michael Kipp, Stefan Kopp, Maurizio Mancini, Stacy Marsella, Andrew N Marshall, Catherine Pelachaud, et al. The behavior markup language: Recent developments and challenges. In *Intelligent virtual agents*, pages 99–111. Springer, 2007.

[121] Yon Visell. Tactile sensory substitution: Models for enaction in HCI. *Interacting with Computers*, 21(1-2):38–53, 2009.

[122] Yon Visell, Federico Fontana, Bruno L Giordano, Rolf Nordahl, Stefania Serafin, and Roberto Bresin. Sound design and perception in walking interactions. *International Journal of Human-Computer Studies*, 67(11):947–959, 2009.

[123] Elizabeth M Wenzel, Marianne Arruda, Doris J Kistler, and Frederic L Wightman. Localization using nonindividualized head-related transfer functions. *The Journal of the Acoustical Society of America*, 94(1):111–123, 1993.

[124] Amanda Williams, Eric Kabisch, and Paul Dourish. From interaction to participation: configuring space through embodied interaction. In *UbiComp 2005: Ubiquitous Computing*, pages 287–304. Springer, 2005.

[125] Margaret Wilson. Six views of embodied cognition. *Psychonomic bulletin & review*, 9(4):625–636, 2002.

[126] Scott Wilson. Spatial swarm granulation. In *Proceedings of the 2008 International Computer Music Conference*. SARC, ICMA, 2008.

[127] Nick Yee and Jeremy Bailenson. The proteus effect: The effect of transformed self-representation on behavior. *Human communication research*, 33(3):271–290, 2007.

[128] Nick Yee and Jeremy N Bailenson. The difference between being and seeing: The relative contribution of self-perception and priming to behavioral changes via digital self-representation. *Media Psychology*, 12(2):195–209, 2009.

[129] Pavel Zahorik. Direct-to-reverberant energy ratio sensitivity. *The Journal of the Acoustical Society of America*, 112(5):2110–2117, 2002.

[130] Lawrence M Zbikowski. Metaphor and music theory: Reflections from cognitive science. *Music theory online*, 4(1):1–8, 1998.

[131] Haixia Zhao, Catherine Plaisant, Ben Shneiderman, and Jonathan Lazar. Data sonification for users with visual impairment: a case study with georeferenced data. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 15(1):4, 2008.

This thesis contributes to the field of embodied and multimodal interaction by presenting the development of different original interactive systems. Using a constructive approach, a variety of real-time user interaction situations were designed and tested, two cases of human-virtual character bodily interaction, two interactive sonifications of trampoline jumping, collaborative interaction in mobile music performance and tangible and tactile interaction with virtual sounds.
While diverse in terms of application, all the explored interaction techniques belong to the context of augmentation and are grounded in the theory of embodiment and strategies for natural human-computer interaction (HCI). The cases have been contextualized within the umbrella of enaction, a paradigm of cognitive science that addresses the user as an embodied agent situated in an environment and coupled to it through sensorimotor activity.

Cover photo: *PuShy*, Roberto Pugliese, 2011.

9 789526 065533

ISBN 978-952-60-6553-3 (printed)
ISBN 978-952-60-6554-0 (pdf)
ISSN-L 1799-4934
ISSN 1799-4934 (printed)
ISSN 1799-4942 (pdf)

Aalto University
School of Science
Department of Computer Science
www.aalto.fi

BUSINESS +
ECONOMY

ART +
DESIGN +
ARCHITECTURE

SCIENCE +
TECHNOLOGY

CROSSOVER

DOCTORAL
DISSERTATIONS