DIS06 Dissertationen

DIS06

DIS06 Dissertationen

Apprimus

Natural Interaction with Audio Playback: Tapping Physical Skills

As humans, we possess inherent perceptual and motor skills: Our stereoscopic vision provides us with depth information and our spatial hearing can localize sounds around us with high accuracy. Our hands are capable of controlling motion with high precision and speed which allows us to write, draw, or play an instrument. Throughout our history we have developed and shaped physical tools that extend and leverage these skills. One important tool of our time, the personal computer equipped with mouse and keyboard, however, is not particularly fit to tasks outside the office domain and falling short of fully leveraging our natural skills. Among the types information we manage with computers, time-based media like audio recordings, is a comparatively young form of information. Since their debut in the 19th century, audio recording and playback interfaces were always designed along technical constraints. The aim of this thesis is to create audio playback interfaces that leverage our natural skills to a larger extend. To increase the interaction bandwidth, we systematically augmented each of the three modalities involved in the interaction: haptic, visual, and auditory.

In der vorliegenden HCI Reihe werden Schriften zum Thema Mensch-Computer Interaktion veröffentlicht, die am HCI Center der RWTH Aachen University entstanden sind. Die Themen behandeln Fragestellungen aus dem Schnittpunkt zwischen Architektur, Informatik, Psychologie, empirischen Sozialwissenschaften und adressieren aktuelle Herausforderungen der Integration neuartiger Technologie im Spektrum vom Mensch, Medien, Raum und Gesellschaft.

9 783863 594619

HCIC - Human-Computer Interaction Center | RWTH AACHEN UNIVERSITY

Florian Heller Natural Interaction with Audio Playback: Tapping Physical Skills

Florian Heller

Natural Interaction with Audio Playback

Tapping Physical Skills

Herausgeber:
HCI Center der RWTH Aachen

HCIC - Human-Computer Interaction Center | RWTH AACHEN UNIVERSITY

Natural Interaction with Audio Playback: Tapping Physical Skills

Von der Fakultät für Mathematik, Informatik und Naturwissenschaften der
RWTH Aachen University zur Erlangung des akademischen Grades
eines Doktors der Naturwissenschaften genehmigte Dissertation

vorgelegt von

Diplom-Informatiker
Florian Heller
aus Aachen

Berichter: Professor Dr. Jan Borchers
Professor Stephen Brewster, PhD

Tag der mündlichen Prüfung: 24. Juni 2016

Florian Heller

Natural Interaction with Audio Playback:

Tapping Physical Skills

Herausgeber:
HCI Center der RWTH Aachen

Band DIS06

Bibliografische Information der Deutschen Nationalbibliothek
Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über http://dnb.ddb.de abrufbar.

Florian Heller:

Natural Interaction with Audio Playback: Tapping Physical Skills

1. Auflage, 2016

Gedruckt auf holz- und säurefreiem Papier, 100% chlorfrei gebleicht.

Apprimus Verlag, Aachen, 2016
Wissenschaftsverlag des Instituts für Industriekommunikation und Fachmedien
an der RWTH Aachen
Steinbachstr. 25, 52074 Aachen
Internet: www.apprimus-verlag.de, E-Mail: info@apprimus-verlag.de

Printed in Germany

ISBN 978-3-86359-461-9

D 82 (Diss. RWTH Aachen University, 2016)

Contents

Abstract **xv**

Überblick **xvii**

Acknowledgements **xix**

1 Introduction **1**

1.1 Thesis Statement 5

1.1.1 Research Question 6

1.1.2 Contribution 7

1.2 Thesis Structure 9

2 Augmenting Haptics: Wearable Controls for Audio Playback **13**

2.1 Portable music player controls 14

2.1.1 From Analog to Digital 15

2.1.2 Wearable Interfaces 17

2.2 Pinstripe 19

2.2.1 System Design 21
Continuous Value Input 22
Menu Navigation 23
2.2.2 Implementation 23
2.2.3 Evaluation of the Pinstripe Interaction 26
2.2.4 Results 28
2.2.5 Embedded Prototypes 31
Evaluation of the Base Fabric 33
Results 34
2.3 Intuitex 36
2.3.1 Limitations 39
2.4 Fabritouch 39
2.4.1 Textile Touchpads 40
2.4.2 The Fabritouch Prototype 40
2.4.3 Evaluation: Support Surface Rigidity 42
2.4.4 Evaluation: Usage Posture 44
Results and Discussion 45
2.4.5 Design Implications 47
2.5 Conclusion 48

3 Visual Augmentation of a Digital Vinyl System 51

3.1 Introduction 51

3.1.1 From Analog to Digital 53

3.1.2 Terminology 56

3.2 Related Work 57

3.3 The DiskPlay System 60

3.3.1 Evaluation 62

3.4 Integration Into a Commercial DVS 65

3.4.1 Technical Setup 67

3.4.2 Evaluation: Acceptance 68

3.4.3 Evaluation: Mixing Task 69

3.4.4 Feedback 70

3.5 Changing the Spectator Experience 71

3.6 Future Directions 72

3.7 Conclusions 74

4 Audio Augmented Environments **75**

4.1 Introduction 75

4.2 Foundations of Spatial Hearing 77

4.2.1 Head Related Transfer Functions . . . 78

Localization Accuracy 79

4.2.2 Audio Augmented Reality 80

4.3 Related Work 81

4.4 Orientation Measurement in Audio Augmented Reality 83

Technical Setup 84
4.4.1 Conditions & Methodology 87
4.4.2 Results 88
4.4.3 Discussion 91
4.4.4 Orientation Measurement and Perceived Presence 93
Technical Setup 94
4.4.5 Conditions & Methodology 94
4.4.6 Results 95
4.4.7 Discussion 97
4.5 Impact of Elevation on Audio Augmented Reality 98
4.5.1 Experiment 99
4.5.2 Technical Setup 101
4.5.3 Results 101
4.5.4 Discussion 104
4.6 Smartphones as Platform for Audio Augmented Reality 105
4.6.1 Implementation 106
4.6.2 Evaluation 107
4.6.3 Results 109
4.6.4 Discussion 110
4.7 Conclusion 111

5 Summary and Future Work **113**

5.1 Contributions 115

5.1.1 Haptics 115

5.1.2 Visual 116

5.1.3 Auditory 117

5.2 Future work 118

A Questionnaires **121**

A.1 Pinstripe Questionnaire 121

A.2 DiskPlay Questionnaire 123

A.3 Presence Questionnaire 129

Bibliography **131**

Index **143**

Own Publications **145**

Papers 145

Posters 146

Workshop papers 147

Magazine articles 147

List of Figures

1.1 How the computer sees us. 2

1.2 Play/Pause interaction is similar on turntables and MP3 players 4

1.3 Cutting audio material in analog and digital form 7

2.1 The Diamond Rio PMP300 portable music player. 16

2.2 Pinstripe usage principle 20

2.3 Pinstripe provides different levels of granularity 22

2.4 The initial Pinstripe prototype. 24

2.5 Pinstripe connection matrix 26

2.6 Different Pinstripe materials 27

2.7 Pinstripe ratings 29

2.8 Different Pinstripe materials 31

2.9 Simplified sensing algorithm 33

2.10 Material ratings 35

2.11 Preferred Material ratings 35
2.12 Intuitex actuation. 37
2.13 Intuitex 38
2.14 Clipping mechanism for quick prototyping. . 38
2.15 Layered architecture of Fabritouch. 41
2.16 Fabritouch integrated into a pair of trousers. 42
2.17 Fabritouch layer materials. 42
2.18 Experiment setup: path. 45
2.19 Movements and according plots on the prototype 47

3.1 Physical structure of a vinyl record. 52
3.2 Traktor Scratch Pro user interface. 54
3.3 The DVS setup 55
3.4 The physical features of a timecode record . . 56
3.5 DiskPlay uses top-projection to augment a turntable 60
3.6 DiskPlay visualization details. 61
3.7 DiskPlay test setup. 63
3.8 Visualization in the second iteration of DiskPlay. 66
3.9 Visualization of a passing cue point 67
3.10 DVS controllers 73

4.1 ILD & ITD. 77

4.2 Head-Related Transfer Function. 79

4.3 The historic Coronation Hall. 83

4.4 Three reference systems. 84

4.5 Orientation: Experimental setup 85

4.6 Low-pass filter curve 86

4.7 Frequency spectrum of the beacon sound. . . 88

4.8 Recorded paths. 89

4.9 Mean head-yaw per source 91

4.10 Head-yaw over task time. 92

4.11 Mean ratings for the different compass placements on the presence questionnaire. . . 96

4.12 Tilting the head changes relative elevation. . 98

4.13 Elevation study: experimental setup 100

4.14 Elevation study: RMS angles by condition. . 103

4.15 IMU comparison. 106

4.16 Audioscope study setup. 108

5.1 A depiction of the modality space. 118

A.1 Pinstripe Questionnaire 122

A.2 DiskPlay Acceptance Questionnaire 124

A.7 Presence Questionnaire 130

List of Tables

2.1 Pinstripe tasks for the qualitative study . . . 28

2.2 Effects of posture and direction on the gestures 46

4.1 Source recognition rates 104

Abstract

As humans, we possess inherent perceptual and motor skills: Our stereoscopic vision provides us with depth information and our spatial hearing can localize sounds around us with high accuracy. Our hands are capable of controlling motion with high precision and speed which allows us to write, draw, or play an instrument. Throughout our history we have developed and shaped physical tools that extend and leverage these skills. One important tool of our time, the personal computer equipped with mouse and keyboard, however, is not particularly fit to tasks outside the office domain and falling short of fully leveraging our natural skills. Among the types of information we manage with computers, time-based media like audio recordings, is a comparatively young form of information. Since their debut in the 19^{th} century, audio recording and playback interfaces were always designed along technical constraints. The aim of this thesis is to create audio playback interfaces that leverage our natural skills to a larger extend. To increase the interaction bandwidth, we systematically augmented each of the three modalities involved in the interaction: haptic, visual, and auditory.

First, we increased the haptic interaction bandwidth for mobile audio players through different wearable interfaces. They build on the affordances of fabric and allow us to use our fine manual motor skills to control playback parameters.

Second, we built on an existing interface for audio playback with an already high haptic interaction bandwidth: the DJ turntable. We extended its visual bandwidth to re-create visual cues for navigation that were lost during the process of digitalization. We could thereby re-locate the haptic input and visual output to a single device.

Third, we increased the auditive interaction bandwidth by integrating a spatial component to recorded audio. This let us create engaging audio augmented reality experiences. However, the increased audio interaction bandwidth also provides much more parameters to be controlled. We evaluated the use of different metaphors to control these parameters in a natural way.

Überblick

Als Mensch besitzen wir uns eigene Wahrnehmungs- und motorische Fähigkeiten. Unser räumliches Sehen erlaubt es uns Entfernungen abzuschätzen und unser räumliches Hören kann den Ursprung eines Geräusches sehr genau orten. Mit unseren Händen können wir schnelle und präzise Bewegungen ausführen welche es uns erlauben zu schreiben, zu zeichnen oder ein Instrument zu spielen. In unserer Geschichte haben wir Werkzeuge entworfen die diese Fähigkeiten nutzen und erweitern. Ein wichtiges Werkzeug unserer Zeit, der Computer, nutzt mit Tastatur und Maus ausgestattet diese Fähigkeiten jedoch wenig. Ursprünglich für die Verwaltung von Text und Tabellen entworfen, sind diese Eingabemöglichkeiten oft hinderlich wenn wir andere Medien wie z.B. Fotos, Musik und Videos betrachten.

Um den Computer für Aufgaben wie z.B. zeichnen zu nutzen, können wir auf Jahrhunderte Entwicklung zurückgreifen und entsprechende Metaphern verwenden. Zeitbasierte Medien sind eine vergleichsweise junge Form von Information, und die bisherigen Interaktionsformen sind technischen Gegebenheiten folgend entstanden. Ziel dieser Arbeit ist es, die Schnittstellen zum Abspielen von Tondokumenten zu überarbeiten damit sie unsere natürlichen Fähigkeiten besser nutzen. Dazu haben wir die Bandbreite der drei betroffenen Interaktionsmodalitäten (haptisch, auditiv, visuell) erhöht.

Zuerst haben wir die Eingabebandbreite von tragbaren Musik-Abspielgeräten über die Nutzung von textilen Schnittstellen erhöht. Da wir hier nicht auf Jahrhunderte alte Eingabeformen als Metapher zurückgreifen konnten, haben wir den natürlichen Angebotscharakter von Stoffen genutzt.

Im zweiten Abschnitt haben wir ein Abspielgerät mit einer bereits hohen haptischen Interaktionsbandbreite genutzt, um die Erweiterung der visuellen Bandbreite zu untersuchen. Dazu haben wir auf einem DJ-Schallplattenspieler Informationen angezeigt die auf klassischen Schallplatten sichtbar sind, die bei der Transition ins digitale Zeitalter aber verloren gingen. So konnten wir haptische Eingabe und visuelle Ausgabe wieder an einem Ort vereinen.

Zuletzt haben wir auch die auditive Interaktionsbandbreite erweitert. Obwohl wir in der Lage sind Schalquellen sehr genau zu orten, ist unser alltägliches Musik-Hörerlebnis eher passiver Natur. Wenn wir Musik über Kopfhörer hören nehmen wir den Ton als "im Kopf" war, und die relative Position der Schallquellen wird zum Zeitpunkt der Aufnahme festgelegt. Wir haben die Möglichkeiten des räumlichen Audio-Simulation genutzt, welche es ermöglicht, die wahrgenommene Position einer virtuellen Schallquelle zu verändern. Da diese Erweiterung der Ausgabebandbreite mehr Steuerungsparameter bietet, haben wir uns gleichzeitig der Eingabebandbreite angenommen um über die Nutzung einfacher Metaphern ein natürliches Hörerlebnis zu generieren.

Acknowledgements

First of all, I want to thank Prof. Jan Borchers for providing me the opportunity to work in the highly interesting research field of HCI. The environment he created was ideal to bring up all these crazy ideas, which after numerous iterations and brainstormings, eventually turned into research projects.

I also want to thank Prof. Stephen Brewster, who was kind enough to agree to be my second advisor, for opening the field of multimodal interaction and interaction with spatial audio. His numerous publications of the past 20 years are a great inspiration to dig further in the area of spatial audio interaction.

I want to thank all my colleagues at i10 for being such a supportive team. Working with people I have know for such a long time, such as Simon Völker and Moritz Wittenhagen made the start of my research career feel like a departure to great adventures. I want to thank Thorsten Karrer and Malte Weiss for being great mentors all the way long. I want to thank Chat Wacharamanotham for teaching me the statistical aspect of HCI and for the crazy paper writing sessions! It was a lot of fun.

Thanks to all the students who worked hard to make my ideas become actual prototypes!

I want to thank my family for always being supportive and believing in me.

Finally, I want to thank Ines and Björn for their endless love, patience, and support.

Chapter 1

Introduction

As humans, we possess inherent perceptual and motor skills: Our stereoscopic vision provides us with depth information and our spatial hearing can localize sounds around us with high accuracy. Our hands are capable of controlling motion with high precision and speed which allows us to write, draw, or play an instrument. Throughout our history we have developed and shaped physical tools that extend and leverage these skills. While these tools do not originate from a human centered design process, but were based on available materials, the course of evolution simply rejected unusable variants. One important tool of today, the personal computer, as introduced by the Xerox Star in the early 1980s and largely unchanged, primarily uses a mouse and a keyboard as input devices, and a screen and loudspeakers as output devices. The use of such general purpose I/O devices has contributed to the success of the PC as it allows to easily control a large number of different applications from various contexts. However, this flexibility comes at the price of not necessarily being particularly fit to tasks outside the office domain and falling short of fully leveraging our natural skills (Figure 1.1).

Humans have developed tools that leverage their natural skills.

The standard PC interface does not make particular use of our natural skills.

The course of evolution in human-computer interaction has brought up technologies trying to address this specific issue. Ivan Sutherland's Sketchpad from 1962 made use of our pointing capabilities using a light pen for two-dimensional manual input, but keeping one's arm lifted to

manipulate objects on the vertical plane of a screen was too exhausting for extended use. The computer mouse was simple and cheap to construct, while being easy to learn and showing a great performance in pointing tasks, which are at the core of the WIMP (windows, icons, menus, pointers) interaction metaphor. However, it does not use the fine motor control of our finger tips, but only part of the hand's mobility. Instead, our fingers are used to press binary buttons. On the output side, our eyes stare at a flat surface, and audio feedback is given through simple loudspeakers which leaves much of these senses unused.

Interfaces that leverage our manual skills have been demonstrated very early in the development of the computer.

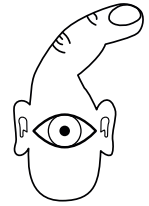


Figure 1.1: How the computer sees us. Only few of our natural skills are used in interaction with computers. (Taken from [O'Sullivan, Igoe, 2004])

Considering the advantages of digital documents, e.g., undo, (nondestructive) editing, copies and backups, versioning, and distribution, the reduction of input and output bandwidth has been considered acceptable in many cases. Imagine the hassle having to search for a specific term in a pile of handwritten notes compared to the simple full-text search on your computer's hard drive, or having to write an entire PhD thesis with a fountain pen. That said, a common observation in the transition to digital documents is the reduction in expressiveness of input controls, which largely abstracts the personal touch of the document. While one can argue if omitting the handwriting from a document is a loss, as it might be more legible in typeset form, only little indication of manual skill of the author remains.

The digitalization of media has abstracted their individual touch.

To appropriately digitalize tasks where manual skill is an essential part of the editing process, such as drawing, input devices with a higher bandwidth are necessary. The "analog" ancestors, the tools for these tasks, are the re-

sult of an evolution over centuries, which means that they have reached a mature and well-established level. Using these, we actually manipulate several parameters simultaneously without thinking much about it. In the case of drawing, for example, stroke width and shape depend on the pressure and angle applied to the brush. Thus, to digitalize tasks with well established analog methods and tools, we can easily revert to the "analog" interaction and make it accessible as input for the computer. Today, we have graphic tablets that measure position, pressure, tilt-angle and, through exchangeable tips, allow the system to closely simulate the feeling of various pens. Such physical tools allow us to take advantage of working with digital documents, but at the same time have rich manual input capabilities.

For some tasks, we can revert to long evolved metaphors to create a computer interface.

INTERACTION BANDWIDTH:
"The use of multiple sensory channels increases the bandwidth of the interaction between the human and the computer, and it also makes human–computer interaction more like the interaction between humans and their everyday environment, perhaps making the use of such systems more natural." [Dix et al., 2004]

Definition: *Interaction Bandwidth*

The personal computer has become the universal tool to handle all sorts of media, not only static text and spreadsheets. How can we apply this physical computing approach to other types of media such as audio and video? Time-based media is a comparatively new form of information that appeared in the second half of the 19th century. While video is basically a sequence of still images which we could also interpret one by one, audio only exists in the time-domain, thus its existence is dependent on an apparatus to record it and play it back. With Edison's phonograph and wax cylinder recordings, audio recordings quickly became popular in the 1880's, but their production was difficult, which is why 30 years later, the shellack disk replaced it as the dominant recording format. In contrast to drawing, the interaction with the medium is not based on human skills, but closely related to the apparatus. To play back sound on a gramophone, the user has to handle a circular disk, the stylus, and a crank; an interaction which has a

Time-based media is a young form of information.

The interfaces for audio playback follow technical constraints.

Figure 1.2: While pressing the button to play/pause a track is essentially the same between a turntable and an MP3 player, the navigation inside a track has changed.

physical component. Navigation to specific tracks and even specific parts of a track are possible by placing the stylus at the appropriate position, which can be perceived visually by differences in the groove pattern. Disk records were the preferred medium to distribute audio until the end of the 20^{th} century.

There is no way for an end-user to record audio on a vinyl record, which led to the development of magnetic tape. The controls we see in current audio players and tools have their origin in the controls used on tape recorders like the Revox A77 from the 1970s. In contrast to vinyl records, audio tape is a purely linear access medium, as it does not allow you to jump to a certain position in the audio signal without winding the tape to the physical location of the recording. The typical buttons for play/pause, stop, fast forward (FFWD), rewind (RWD) provide direct access to motor control and the physical position of the playhead (on or off the tape), and are totally unrelated to the content stored on the medium. However, being well known and simple to implement, their general design was retained on players for random access digital media such as the CD. The semantics changed slightly as FFWD and RWD are now used to skip from one track to another, while seeking within a track is linearized on most players. It is only with the timeline slider of software players that the random access of digital audio recordings became easily accessible. The problem with the timeline slider is its disconnection from the semantic content of the recording; thus if we want to navigate to a certain break or chorus, we do not know where to click. A waveform display is a useful hint for such

Today's audio playback interfaces are not much different than those from the 1970's.

Modern audio playback interfaces have a low interaction bandwidth.

a task as it reveals the structure of the recording through its volume, however, it is seldom used in simple audio player software.

We are thus facing two problems: First, during the evolution of interfaces for time-based media, controls have become more shallow in terms of their use of physical skills. Second, in contrast to drawing, when designing interfaces for time-based media we do not have a century old "analog" predecessor to which we can revert, but only a series of technically derived controls with only a very low interaction bandwidth.

How can we create natural interfaces for audio playback.

1.1 Thesis Statement

The main question behind this thesis is how to create interfaces for audio playback that provide a more natural access to the controls. Human interaction is multimodal and of high bandwidth: People express themselves through gestures, mimics, voice and sound, and perceive the world through five distinct senses. As seen in the history of playback interfaces, "the language between people and machines has been determined mainly by technological constraints, and humans had to adapt to such language" [Valli, 2008]. Building on the research framework and directions presented by Jacob et al. [1993], we seek to increase the interaction bandwidth between human and computer with the goal to make computers a better fit to humans for the specific case of audio playback. More precisely, we assert that an increase in bandwidth along each of the three modalities used in interaction with audio playback control — visual, haptic, auditory — results in more natural user interfaces to manipulate a set of playback parameters.

Increasing interaction bandwidth results in more natural user interfaces.

Current end-user audio playback controls only have a very limited bandwidth as they mostly consist of a series of buttons, but manufacturers have worked on interaction bandwidth increase. In the case of seeking to a certain point in an audio recording, for example, we can only progress at a fixed rate by keeping the FFWD button pressed. The iPod jog wheel solved this issue by giving access to var-

The iPod jog wheel increased interaction bandwidth.

ious parameters in a way that users could easily control these at varying speeds. It was used to control volume and manipulate timeline sliders by mapping the rotation to a horizontal movement. Furthermore, its universal design allowed a meaningful mapping of clockwise and counterclockwise rotation to up and down movement of selection in the menus of the iPod software interface. Smartphones with touchscreens or computers increase the interaction bandwidth through timeline sliders that allow you to quickly jump to a certain position with a single touch. This, first, requires visual attention, which might be disturbing in a mobile context, and second, still does not provide much information as the timeline slider is disconnected from the content of the recording. Thus, increasing bandwidth without providing a comprehensible connection to the media will not result in improved user interfaces. In the example of the timeline slider, combining the control with a waveform display would provide for some simple form of semantic navigation.

Touchscreens provide rich visual output, but only little haptic feedback.

Finally, we can think of radically different perspectives on playback control. Modern audio rendering technology allows us to think beyond mere reproduction of a stereo recording by providing means of integrating spatial information into the audio stream. While this substantially increases the interaction bandwidth between the user and the medium, it builds on our naturally given capability of our auditory sense to localize the source of a sound, resulting in only minimal additional cognitive load.

1.1.1 Research Question

In order to achieve our goal of natural interfaces for audio playback we have to answer the following general questions:

How can we increase the interaction bandwidth in interfaces for audio playback? As seen in the examples above, the idea of increasing the interaction bandwidth has been implemented in various ways. The iPod jog wheel or faders

How to increase interaction bandwidth?

Figure 1.3: Cutting audio material in the analog and digital world. The tools in the digital workflow are metaphors of the once physical activities. (Left image from phizyx.com)

and rotary controls to adjust volume increase the haptic interaction bandwidth, while waveform displays help to visually navigate through the content of an audio file. Instead of following a human centered design, however, these approaches are mostly driven by technological possibilities, which directly leads to the following question:

How to leverage our natural skill?

How can we leverage our natural skills in such augmented interfaces for audio playback? This particularization is the focus of this thesis, as it shifts attention from technology to human capabilities. In the case of drawing, evolution has sorted out unusable interaction forms over time, thus we can safely revert to these as metaphor for a computer interface. Since time-based media are comparatively young, this evolution has not happened yet and we have to look for adequate replacement in form of metaphors that are easy to understand and learn.

1.1.2 Contribution

We perform a systematic analysis of the modality space.

The contribution of this thesis is to show how to increase the input and output bandwidth of audio playback controls by systematically augmenting each of the three modalities involved in the interaction: haptic, visual, and auditory. As audio playback controls are used in different contexts from accompanying sports activity to professional setups, we show how this process can be used in a variety of applications.

Haptics: As mentioned before, pressing binary buttons does not fully take advantage of our haptic skills, which is why we fist look at increasing the haptic interaction bandwidth. Mobile audio players are used in a context where precise manipulation of buttons may be difficult and visual perception is already occupied for a primary task such as walking. Performing double or triple presses on the small remote inline the headset cable while jogging is difficult, as one needs to find the remote first and then keep the hand steady relative to the body to not pull the earplugs out of the ear. We build on the natural affordances of fabric to create a controller with high haptic interaction bandwidth. The user grabs a fold in a piece of cloth to control continuous linear values with varying granularity. This does not require visual attention as the sensor is covering a large area and only senses relative changes to the initial point of interaction. Next, building on the established knowledge of gesture input on tablets and touchscreens, we created a two-dimensional wearable touchpad to enter basic commands without facing the problem of involuntary activation. While participants successfully used both interfaces, and their acceptance of our wearable interfaces was high, achieving high precision proved to be difficult. Thus, increasing the input bandwidth in these cases leverages our natural skills, but the best results are achieved in conjunction with adapted software interfaces. Coming back to our comparison with the drawing interface, increasing the haptic interaction bandwidth is similar to the step from drawing with a mouse to drawing with a digital pen.

How can we use our fingers for more than pressing buttons?

Visual: In the second step, we take an audio playback interface that already has a very high haptic interaction bandwidth, and increase its visual output bandwidth. In this case, we explore the space in the context of professional audio playback interfaces. Since playback control is the primary task, specialized interfaces exist, mostly providing dedicated buttons and sliders. The turntable, however, provides such a unique haptic feeling that special systems were built to use it as a controller for digital audio. During this process, however, some of the unique features of vinyl records got lost. We will thus first restore information that was present in the “analog” ancestor but was lost during

How can we use visuals to improve interaction with digital audio?

the digitalization process, and second, create an embodied visualization that does not require split attention between the location of haptic input and the separate visual output. In the drawing analogy, this is a pen display which allows you to draw and immediately see the result below the tip of the pen. In contrast to a screenless graphics tablet, you do not have to do any spatial mapping between input and output.

Auditory: Finally, we explore the possibilities of spatial audio as an augmentation of the common stereo recording. We analyze a spatial audio display that is controlled through position and orientation tracking of the user. The spatial audio display has an increased output bandwidth since it allows communicating and modifying the spatial arrangement of the different sound sources. The location and orientation tracking, depending on the technology used, provides three to six different degrees of freedom, which seems complex. If this information, however, is mapped to the according parameters in the spatial audio rendering algorithm, using such an auditory display does not require any additional attention as the system simulates the natural, spatial perception of sound. Even simple simulations of spatial sound allow successful navigation within an arrangement of virtual sound sources. Sensing the user's head orientation using additional hardware can be approximated using the built-in sensors of a modern smartphone. Modern, high-end rendering algorithms allow to precisely tell apart different proximate sources from a distance. In the drawing example, this would be a pen display with a surface that feels like paper, or any other appropriate drawing substrate.

How can we make use of our spatial auditory perception?

1.2 Thesis Structure

In chapter 2, we present three different wearable controllers and how they can be used in conjunction with a portable MP3-player or smartphone. We explore the space of wearable controllers by implementing different specific charac-

teristics. We propose a wearable interface for simple gesture input to replace buttons with an unobtrusive, always available interface that does not have the problem of involuntary activation. We propose a textile controller for one-dimensional continuous values that allows one to change, e.g., the volume at different granularities. We extend this interface to cover two dimensions which can be used in conjunction with circular spatial auditory menus.

In chapter 3 we analyze a highly specialized haptic interface, the DJ turntable, and augment it with a visual overlay that shows additional information on the track currently playing directly on the vinyl record. In the era of vinyl records, DJs created a new form of art around the manipulation of the record and the turntable. Instead of simply playing the record from start to end, they spin it back and forth, adjust playback speed, and combine two parallel tracks to form a new one. Mastering the art of scratching takes years of practice, and when switching to digital media, the DJs do not want to start over. Digital vinyl systems use a special vinyl record to control audio playback on a computer, thereby allowing the DJ to build on his skills and take advantage of digital media playback and storage. However, the special control record does not provide the same amount of visual information as the traditional ones, thus spatially separating visualization and control. By building an augmented DJ turntable, we bring back these visual cues, creating an embodied unit for rich haptic audio playback control.

In chapter 4, we study the impact of different sensor platforms and sensor placements on the experience of using a spatial audio display. By sensing head and body orientation in up to six degrees of freedom and feeding this data into a spatial audio rendering algorithm, we create an embodied experience, which, since it behaves very similar to our natural perception, generates only minimal additional cognitive load.

Chapter 5 summarizes the insights gained through the thesis and relates these back to the research questions defined in section 1.1. Some examples will show how the approach taken in this work can be applied to other forms of media

to generate useful ideas for future controllers. We will discuss limitations of this work and, based on these, point out further research directions.

Chapter 2

Augmenting Haptics: Wearable Controls for Audio Playback

"...But I haven't figured out an iPod yet."

—Harrison Ford

The first modality we consider for bandwidth augmentation is the haptic one. We will look at haptic interaction with audio playback controls in a context where the visual sense is already occupied and thus, should be exempt from additional load. When using portable music players, the visual sense is used to perceive our surroundings, so looking

Publications: Pinstripe was first published as Poster at UIST '10 [Karrer et al., 2010] and as full paper at CHI '11 [Karrer et al., 2011] for which the author was responsible for the hardware. Jan Thar worked on Pinstripe as subject of his bachelor thesis under the supervision of the author of this thesis [Thar, 2013]. Intuitex has been published as poster at MuC '15 [Heller et al., 2015]. The author of this thesis contributed the text and illustrations and supervised the hardware development. Fabritouch was the subject of the Master's thesis of Stefan Ivanov [Ivanov, 2012] under the supervision of the author of this thesis and has been published as short paper at ISWC '14 [Heller et al., 2014a] which was written by the author of this thesis as a main author.

When handling a portable music player, the visual sense is already occupied.

at a display or a set of buttons distracts from that primary activity. Well designed interfaces on such devices allow one to reach the controls to the most relevant functions eyes-free, but that requires to remember their physical arrangement. Additionally, these controls are prone to involuntary activation when carried in the pocket, which is typically solved by introducing a slider locking the user interface, which can make the interaction even more complicated.

How do we create a natural interface for a portable music player?

We do not have an "analog" ancestor to revert to.

We will use the natural affordances of fabric to make clothing a ubiquitous interaction surface.

We are thus facing two problems: First, constructing an interface that can easily be manipulated eyes-free which means that the number of controls should be minimal to avoid having to memorize a spatial layout. Therefore, the controls need to have a higher interaction bandwidth to provide access to equivalent functionality. Second, the controls should be designed in a way that the risk of involuntary activation is reduced to a minimum. As described in the previous chapter, no analog ancestor with a natural interface exists that would provide metaphors we could use for a portable music player. Hence, we looked at affordances of possible interaction surfaces that are available in a mobile context, and opted for clothing which is a nearly ubiquitous interaction surface. In this chapter, we explore the space of fabric interfaces, focusing on how to utilize their natural affordances to control a portable music player. By creating textile controls with a higher interaction bandwidth, we allow to control continuous values using a single point of input and we replace binary command buttons with simple gestures on easily reachable input surfaces.

2.1 Portable music player controls

Portable music players had the same basic transport controls as their stationary counterparts.

The Sony Walkman, introduced in 1979, was the first portable music player small enough to be a permanent companion while on the go and as such, defined a new device category. As with its stationary counterparts, the transport controls were designed around the medium, meaning that *fast forward* winds the tape at a faster speed, resulting in a high-pitched playback or without any audio feedback at all, depending on whether the playhead remains on the tape or not. Being a linear medium, content on the cassette

can only be parsed in a time-based manner, since direct access to specific tracks is not possible.

2.1.1 From Analog to Digital

With the introduction of random access media, like the CD, users could easily switch from one track to another, or to a specific time in a track. The controls of portable CD players are those of tape players adapted to the new functionality, i.e., *fast forward* was replaced by *next*. Since users mostly wanted to skip tracks, not search in them, the seek-functionality was only available as secondary function when keeping the skip button pressed. Well designed devices allowed eyes free control through the tactile design of their physical buttons, and to allow the player to be stowed away, small remote controls inline with the headphone cable appeared, providing access to essential functions. With a CD able to hold around 74 minutes of audio, the available controls can be considered appropriate for managing this amount. Despite these advantages over tape, the standard CD player controls still require the user to memorize the order of songs on the CD, as navigating through the tracks is linearized by the use of *skip*-buttons.

Portable CD players just adapted the tape controls.

In the late 90's and early 2000's, the MP3 file format radically changed how we manage music. Suddenly, large music collections were not distributed on physical media and could all be stored on a single hard drive. This made it possible to easily carry large amounts of music and audio on a portable device. The first models like the successful Diamond Rio PMP300 (Figure 2.1), however, just copied the controls from portable CD players, which only allowed linear navigation within the music collection stored on the device. With increasing storage capacity, it became apparent that this is not practicable. Apple's iPod introduced a large planar scroll wheel which increased the input bandwidth enough to make large musical collections manageable on a portable device.

MP3 players required new controls to manage the large music collections they could store.

With progressing evolution, the dedicated portable MP3-player and the mobile phone eventually merged into one

Figure 2.1: The Diamond Rio PMP300 was the first commercially successful portable MP3 player. Although users were now able to carry large music collections with them, the controls were still the same as on other portable music players (Image courtesy of John Fader).

Music players on smartphones do not provide haptic feedback and require visual attention.

single device: the modern smartphone. The large touchscreen allows for more adaptive interfaces than the fixed set of hardware buttons available on dedicated players, which makes it possible to easily manage large collections of music. However, these software interfaces do not provide any haptic feedback and require constant visual attention.

In most use cases, listening to music is an auxiliary activity to a primary task, with which interacting with the player should interfere as little as possible. If we think of situations like sports, handling a touchscreen mostly means interrupting the primary activity. To solve this problem, manufacturers often integrate a minimal remote control into the headphones, providing quick access to functions like play/pause, next or previous track, and volume control. In contrast to the cable remotes used with portable CD players, these also contain a microphone for voice input, which restricts their size and weight because they need to

be placed such that they can record a clean voice signal. This reduces the number of buttons to three, which limits the expressivity of these controls and results in some functions only being accessible through time-based commands. On the iPod, a single press toggles playback, a double press jumps to the next track, and pressing three times jumps back to the previous track. Interacting with these remotes requires a steady hand as otherwise you easily pull out the earplug. Such fine motions are difficult during physical activity. Larger headphones offer the possibility to integrate more dedicated controls, either by using simple buttons, or more complex capacitive sensors which enable gestures on the earcups (such as on the Jabra Revo Wireless[1] or the Intelligent Headset[2]). However, such headphones are not suitable for physical activity due to their size and weight.

Headset remotes only have a very limited interaction bandwidth.

2.1.2 Wearable Interfaces

Our own body provides a relatively stable frame of reference, in which we can reach points on the arms and upper body successfully even while in motion [Wagner et al., 2013]. Proprioception helps us compensate for the relative movement between the hands and the touch target, thus, using the body as input potentially alleviates the issues we face using inline remotes. EarPut [Lissermann et al., 2014] is a small array of capacitive proximity sensors placed behind the pinnae that augments the human ear with touch functionality. iSkin [Weigel et al., 2015] generalizes this idea by using printed patterns of conductive ink on a flexible substrate which are then connected to a microcontroller. With this technology, a number of additional dedicated input controls can be added to any part of the body, independently of the size of the earphones. The disadvantage is that there is no physical hint where the different control areas are located, which, in the case of EarPut, limits the number of precisely distinguishable control areas to four. In addition, placing the sensitive areas on exposed areas of the body increases the risk for involuntary activation.

Proprioception helps us when using the body as input.

[1] jabra.com
[2] intelligentheadset.com

Instead of creating potentially obtrusive tactile cues on the human body, is to use the natural segmentation of the human body to delimit different functions. Skinput [Harrison et al., 2010] and the system by Mujibiya et al. [2013] use the propagation of sonic waves inside the human body to distinguish between touches at different positions of the lower human arm. This allows one to differentiate taps on the different fingers, and as such, provides clearly separated input surfaces for different commands. Depending on the situation, for example when wearing gloves in cold weather, this approach might fail due to the dampening effect of additional layers of material.

These controls are not always accessible, and involuntary activation is an issue.

Looking for potential surfaces to increase the input bandwidth for interaction with a mobile music player, we investigated textile interfaces, as these can easily be integrated into everyday clothing. The idea to augment clothing with additional functionality to control electronic systems has been around for over a decade [Rantanen et al., 2002]. A closer look at the small number of commercial products like the Rosner mp3blue[3], however, reveals that the interaction concepts of these wearable controls are basically a direct transfer of known concepts, e.g., buttons, to a wearable context, and do not use any of the natural affordances of fabrics. Furthermore, depending on the technology used to implement these buttons, operating these still requires visual attention. A capacitive button registers a touch the moment you reach it, meaning that you cannot stroke over a series of buttons and count the tactile landmarks, since passing over them would already trigger actions. If placed at an exposed location, buttons also have the problem of involuntary activation, which is hard to solve without interfering with their regular use. The appropriation of regular clothing components, such as cords or the fabric itself, has many advantages. They provide large surface areas for interaction, allowing continuous input in potentially many dimensions. Schwarz et al. [2010] built and evaluated several prototypes of cord-based controllers which could sense touch location, pulling force, and rotation. The preferred interaction dimensions are twisting for continuous input, e.g., changing the volume, and pulling for toggle actions, auch as play/pause. These results demonstrate that using

Clothing is a nearly ubiquitous interaction surface.

[3] rosner.de

the intrinsic affordances of textiles leads to easy to use interfaces. But what if the personal clothing style does not include cords?

We will now present textile input controllers that can be integrated nearly everywhere into everyday clothing.

2.2 Pinstripe

When we interact with the fabrics that surround us everyday, we explore these with our hands to determine their properties by folding, crumpling, and caressing them. The idea behind the *Pinstripe* textile input controller is to leverage two of these characteristics of interaction with fabric: grasping and deforming . The user pinches a fold into the textile and rolls it between her fingers to control an continuous value.

The underlying technology of Pinstripe is very simple: A number of parallel stripes of conductive thread are sewn into the garment. Once the user grabs a fold, some of these lines get connected, which can easily be sensed by a microcontroller (cf. Figure 2.2). By determining which of the lines are interconnected, we can measure the size of the fold and its movement. The fact that the amount of technology needed to create such an interface is minimal, allows it to fulfill a number of design requirements that are posed on to smart clothing which can roughly be categorized as follows.

Pinstripe consists of parallel lines of conductive thread used to measure the size and position of a fold.

Wearability commonly refers to the demand that integrating electronic components should not influence the primary functionality of the clothing itself. This, for example, includes body temperature distribution and insulation, breathability of the fabric, and wearing comfort in general [Marculescu et al., 2003; Martin et al., 2007; McCann et al., 2005]. A detailed analysis of the influencing factors on wearability can be found in Gemperle et al. [1998].

Wearable interfaces should not interfere with the primary function of the clothing.

Fashion compatibility should be maintained by minimizing the influence the electronics have on the visual aesthetics

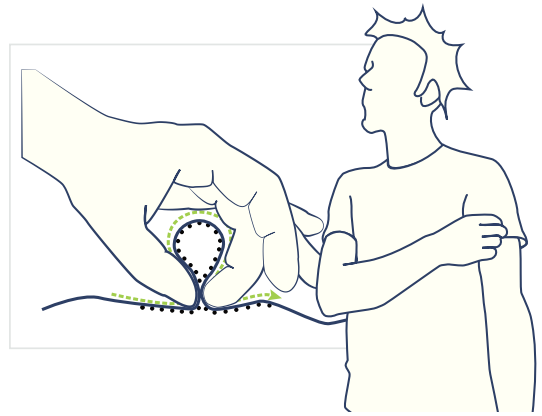


Figure 2.2: Pinstripe senses the movement and size of a fold that the user pinches into the garment and moves between her fingers (Adapted from [Karrer et al., 2011]).

of the garment, or to be invisible altogether [Holleis et al., 2008; O'Donnell, 2003; Marculescu et al., 2003; Toney et al., 2003]. While the electronics like LEDs can be an integral part of the visual appearance of a piece of clothing, a visible wiring on the outside of the garment make it look like a technical prototype and not like something that would be worn in daily use.

Durability is a crucial quality for the success of basically any clothing. Just as regular garments, smart clothing should be engineered to withstand many cycles of washing and drying without limitation in functionality [Linz et al., 2005].

Interaction with wearable controls in itself already poses a series of challenges [Martin et al., 2007]. While the manipulation of the control should obviously work as intended, wearable interfaces are much more subject to external influences than their desktop counterparts for example. They should not activate involuntary [Komor et al., 2009], e.g., by body contact with other people in a crowded area, and since their interaction area is limited, they should be easy to

detect on the clothing through visual or haptic cues [Holleis et al., 2008].

2.2.1 System Design

When worn, clothing either exhibits loose folds in different areas, or, if it is made from stretchable material, a fold can easily be grabbed into it. Pinstripe lets wearers provide input by pinching a fold between thumb and another finger into their clothing, and then rolling this fold between their fingers (see Figure 2.2). This movement changes the relative position of the two sides of the fold, which results in connections between the conductive lines sewn into the fabric being closed and opened, which is measured and interpreted to a continuous change in value.

Pinstripe senses the size and position of a fold to control a continuous linear value.

This design addresses the *interaction* problems mentioned above: Pinstripe is operated one-handed, and since it is not necessary to create the fold at an exact specified location (control is non-local), Pinstripe is well-suited for eyes free operation. This non-local control also results in a high robustness against the garment shifting relative to the body while worn. While a button on the outside of a sleeve may easily move from its position on the back of the wrist to its side when the sleeve of the garment twists, Pinstripe works equally well, no matter where on the interactive surface the fold is located. Pinstripe does not directly activate when being touched, making it robust against involuntary activation, since pinching a fold in the fabric is rarely done accidentally. Nevertheless, body areas like joints where garments fold naturally while moving, are, of course, not a suitable position for Pinstripe.

Pinstripe avoids involuntary activation.

Since the active areas do not need to be highlighted on the garment, and the conductive threads run on the inside of the textile, the impact on *fashion* is minimal. The only part of the systems that impacts *wearability* is the sensing microcontoller which can be mounted on a flexible printed circuit board (PCB) and miniaturized to reduce the non-flexible components to a minimum. The conductive threads are spaced apart enough to maintain breathability and pre-

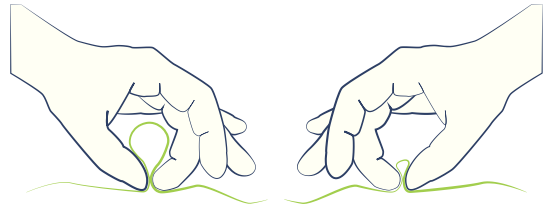


Figure 2.3: By pinching and rolling deeper or smaller folds into the textile, the user can define different granularities of control. (Adapted from [Karrer et al., 2011])

serve other engineered properties of the fabric. Modern conductive yarns, flexible PCBs and their interconnection become increasingly robust and can withstand many washing cycles [Linz et al., 2005], such that Pinstripe can be constructed as a *durable* textile interface

Continuous Value Input

Control elements for continuous values usually exhibit an inherent problem of domain scaling. A GUI slider, for example, can only provide as many distinct values as it occupies pixels on the screen. Hürst et al. [2004] proposed the use of an adaptive scale which can be varied by moving the mouse orthogonally to the primary axis of the slider. With increasing distance to the slider the resolution gets finer, which allows a quick coarse navigation followed by a fine tuning of the location at the end. This principle is implemented, for example, in Apple's iOS music and video players. Pinstripe allows a similar interaction by taking the size of the fold into account. Pinching a large fold results in coarse control while a grabbing a small fold yields more fine-grained control (Figure 2.3).

The user can control values at varying granularities.

Menu Navigation

Besides providing control over a linear value, such as the volume of your MP3 player, Pinstripe can also be used to navigate linear or nested menu structures, e.g., the music player's playlist. In this setting, scrolling the fold through your fingers moves the current selection across the list. The size of the fold indicates the level of scrolling, e.g., a small fold scrolls though the list of tracks of an album while a large fold scrolls at the level of albums. Several ways to issue confirmation and cancellation commands are thinkable. Some which can easily be recognized by the microcontroller are *dwell time*, *activate-on-release*, and *grab-and-crumple*. *Dwell time* confirms a selection if the current selection is held for a certain amount of time and cancels it if the fold is released before reaching that time. As timeout-based interfaces are known to negatively affect the usability, we chose not to use this method in our prototypes. *Activate-on-release* issues an implicit confirmation when the user releases the fabric, which means that an additional cancel item has to be added to the list. Furthermore, depending on the use context, an uninterrupted use cannot be guaranteed, which would result in a number of unintended selections. *Grab-and-crumple* is triggered by the user grabbing the fold in her fist, which results in a large amount of the stripes being connected to one another. This makes this gesture easy to distinguish from the normal gesture, both for the user and the microcontroller.

To activate a selection, the user crumples the sensor.

2.2.2 Implementation

We evaluated the concept with a series of increasingly sophisticated prototypes. The initial implementation was built from 18 parallel lines of conductive thread sewed at 2 mm intervals on the inside of a sleeve of a t-shirt (Figure 2.4). Every line was connected to a single digital I/O of a LilyPad Arduino[4]. The digital pins of the LilyPad were initialized as inputs and their state set to HIGH by activating the internal pull-up resistors. The actual measure-

[4] lilypadarduino.org

Figure 2.4: The initial Pinstripe prototype. 18 parallel lines of conductive thread are sewn into the sleeve of a T-shirt and connected to a LilyPad Arduino (Image taken from [Karrer et al., 2011]).

The first prototype had 18 stripes connected to a LilyPad.

ment was performed by iterating over the stripes and setting one of them as an output at LOW-level, and checking the state of the remaining stripes. If one of these is pulled to LOW-level, we know that there is a connection to the output stripe. The result of these measurements is stored in a matrix which is then sent to a remote computer over a serial connection for further processing. Since the matrix is symmetric, except for measurement errors and outliers, we only consider the upper right triangular matrix for our search for a connected area of positive entries. The position and size of this area describes the two main features of the fold: Its position along the primary diagonal represents the position of the fold, while the position on the secondary diagonal indicates the size of the fold (see Figure 2.5). When the user rolls the fold, the area of positive entries moves along the primary diagonal. The values were low-pass filtered before sending them off to the application to be controlled, which can interpret them, e.g., to control the volume of a portable MP3 player, to adjust the temperature of a garment with built-in heating or cooling, or to navigate through graphical or auditory menus on a device.

While in this first implementation the data processing entirely happened on a computer, we built several iterations of prototypes, up to an autonomous version, to further explore the concept. To test the applicability of the concept in a real world scenario, we built an improved version based

```
input : stripes: An array containing the mapping from
        stripe numbers to I/O pins
output: connections[# of stripes][# of stripes]: An upper
        right triangular matrix containing the connection
        information of the stripes
set all pins as input
enable all PullUps
for (i < |Stripes|) do
    set stripes [i] as output;
    set stripes [i] to LOW
    for (i < j < |Stripes| do
        if (stripe == LOW) then
            connections [i][j] = 1;
        else
            connections [i][j] = 0;
        end
    end
end
```

Algorithm 1: The original Pinstripe sensing algorithm

on the findings of our previous study [Karrer et al., 2011]. Instead of sewing single lines of conductive thread into the sleeve, we used a piece of fabric with conductive patches (Figure 2.6a). These patches are connected in vertical direction but not in horizontal direction and thus form the stripes of our sensor. The larger patches create more reliable connections with each other which results in lower noise in the sensor readings and makes them easier to interpret. To connect the fabric with the electronics, we first sewed lines of conductive thread into the stripes on the one side and into a small piece of copper foil on the other end, allowing for a soldered connection of a wire going to the microcontroller. We could increase the number of stripes to 24 and adapt the size of the sensor patch accordingly. Since this required a larger number of available digital I/O pins, we replaced the LilyPad Arduino by an Arduino Mega based on an Atmel ATmega 1280 microcontroller. Running with a faster CPU clock speed (16 MHz instead of 8 MHz), we could also increase the sample rate from 2.8 kHz to more than 5 kHz to compensate for the lag introduced by the filtering. To remove jitter we used a simple median-of-three filtering and smoothened the mean loop position and size using an exponential low-pass filter.

We built several prototypes with different conductive fabrics.

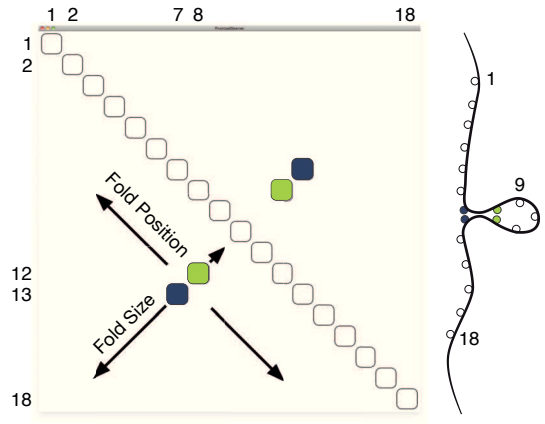


Figure 2.5: The connection matrix (filtered to remove outliers) and a corresponding fold. Entries in the connection matrix show which lines are currently connected through pinching. The 'blob' of all connection entries in the matrix indicates the size and placement of the fold across the conductive threads. Here, lines 7 and 13 (blue) as well as 8 and 12 (green) are connected; the user has formed a small fold for fine-grained control. Note that the matrix is always symmetric. (Adapted from [Karrer et al., 2011])

2.2.3 Evaluation of the Pinstripe Interaction

We evaluated Pinstripe using a music player application.

As a real-world application, we implemented a music player which was controlled using the Pinstripe sensor to either change the volume or the current track in a playlist. Every new pinching gesture resets the origin of the fold, meaning that the changes communicated for every subsequent rolling of the textile are relative to the value that is currently controlled (e.g., track number or current volume). When controlling the volume, the size of the fold was used as a scaling factor for the step size (Figure 2.5). The smallest detectable fold of our prototype was 2 mm in size, resulting in two neighboring stripes to be connected, while the largest fold was detected when the two outermost stripes

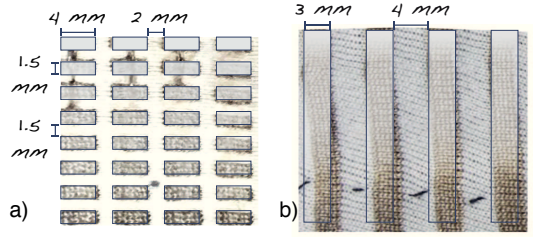


Figure 2.6: Dimensions of the Pinstripe base materials we used in different iterations. **(a)** The fabric consists of a series of conductive pads that are connected in vertical direction. **(b)** The material consists of conductive stripes.

connected (approx. 14 cm). For volume control, moving the fold 2 mm (distance between two threads) changes the volume by 1% for the smallest and 33% for the largest fold size. When switching through the list of tracks, 'next track' and 'previous track' commands are issued every time a predefined threshold is crossed, which happens approximately every 4 mm.

Participants had to change the volume and navigate through a playlist.

Our music player application copied the behavior of many mobile MP3 players that apply volume changes directly while they give a brief audio feedback in form of a beep before they switch over. Additionally, our application played a distinct sound when reaching either end of the playlist. Moving the selection mark while navigating in the graphical menu works similar to changing tracks, but without audio feedback. We adopted the 'grab-and-crumple' gesture described earlier to confirm the selection, which is triggered when 35% of all possible thread connections are active. All of these settings were derived from the results of a small pilot study performed beforehand with members of our lab.

We concluded the study with a standard SUS questionnaire [Brooke, 1996] that we extended to include the following questions specific to our project.

menu mode	1	move to the next item in the menu
	2	move back to the previous item
	3	move ahead 3 items in the menu
	4	skip to the last item in the menu
	5	go back to the first item in the menu
	6	select the item “flower” in the menu
	7	deselect the item “lightning” in the menu
volume mode	8	adjust the volume to a suitable value
	9	adjust the volume to minimum
	10	adjust the volume to maximum
	11	adjust the volume to a suitable value
playlist mode	12	move to the next item in the playlist
	13	move back to the previous item
	14	move ahead 3 items in the playlist
	15	go back to the first item in the playlist

Table 2.1: Tasks to be performed during the qualitative study.

- I felt that I could control the volume precisely.
- I felt that it was easy to switch between tracks.
- I had difficulties navigating the graphical menu.
- I would be uncomfortable to use a final version of the system in public.
- I would buy clothing with this functionality to control my portable music player.
- I would be willing to pay an extra ____ EUR for clothing that included this functionality.

2.2.4 Results

A total of 14 people (2 female) with an average age of 26 (range 21-31, SD=2.5) participated in this study with only one being left-handed. Most users experimented with touch input first when being introduced to the Pinstripe garment, but were able to successfully use Pinstripe after being shown by the experimenter and a short learning

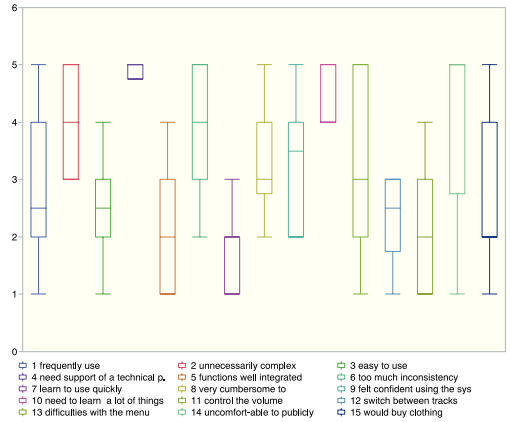


Figure 2.7: Median ratings of Pinstripe on a five point Likert scale (1: strongly agree, 5: strongly disagree).

Pinstripe received an average SUS score.

phase. Our system received an average SUS rating of 68.9 (SD=13.3) (Figure 2.7), which means it only has an average usability [Sauro, 2011] and there is room for improvement. While we did not perform a quantitative evaluation, we made some interesting observations while participants performed the assigned tasks.

Sensor location and angle: Consistent with the findings from our study on sensor placement [Karrer et al., 2011], the angle at which people pinched the sleeve was aligned with the stripes of the sensor and very consistent across participants. The position at which participants interacted with the sensor, however, varied strongly, ranging from the outside to the inside of the arm. This supports the argument to prefer non-local controls like Pinstripe over capacitive buttons and similar controls. To work reliably for a large population, the sensor needs to be large enough to span the full circumference of the arm to avoid 'slipping off' the sensor patch.

Participants immediately understood the concept.

Ease of use: All participants felt at ease navigating the graphical menu and immediately understood how to activate menu items by grabbing the fold with the full hand and crumpling the textile. Most people preferred the visual feedback of the menu condition over the audio only feedback of the music player. A possible explanation is that users can see the direction they are navigating and the step size they are using. This problem of visibility might be of less importance if they navigate their own, well-known playlists.

Variations in gesture: During the study, we made an interesting observation: while none of the participants experienced problems to switch to the next track or turn the volume up, some struggled with the opposite direction. We took a closer look and found that these users performed the Pinstripe gesture in a way that resulted in asymmetric forward and backward finger motions. To perform the forward gesture, they held the thumb against the index finger and moved the fold by bending and stretching the index finger while holding the thumb steady. This is similar to the way the instructor performed the demonstration gesture (Figure 2.8 left). However, when going the opposite direction, they bent the thumb against the steady index or middle finger, which results in a limited movement due to the smaller angular range of the thumb joint. Users performing the gestures in this way usually felt that reaching the threshold for, e.g., skipping to the previous song required a larger movement of the thumb in contrast to the index finger although the threshold was equal in both cases. Some participants performed the gesture in an entirely different way with the thumb and fingers being parallel to the fold (Figure 2.8 right). Instead of bending the fingers, they formed a flat surface with their fingers and then rolled the fold by sliding with the thumb over that surface. These users perceived no difference in the amount of movement required to issue forward and backward commands, presumably because the thumb can be slid sideways in both directions equally well. However, they generally felt that the distance they had to move their thumb to reach the next piece or menu item was too large.

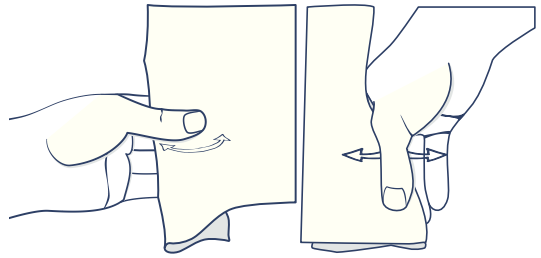


Figure 2.8: Different Pinstripe gestures (Adapted from [Karrer et al., 2011]).

The mapping of movement direction to what the users expected to match 'forward' was consistent with our predictions based on the first study in [Karrer et al., 2011]. In 79% of the cases, the system's behavior matched the user's expectations, while the rest found the mapping to be more natural the other way around. Thus, the mapping should be easily adaptable to the user's preference in final versions of the product.

The mapping of up/down matched the expectations of 79% of the participants.

All participants understood the domain scaling concept that the fold size influences the granularity of volume changes. However, some users encountered the problem that while rolling the fold through their fingers to change its position, they also changed its size. This made controlling the volume more difficult for these users.

Participants understood the mapping of fold size to change granularity.

Further experiments with different materials revealed that not only the conductive material has an impact on how good the fold can be rolled between the fingers, but that the type of base fabric also determines the amount of grip and thereby the amount of force one needs to apply to the Pinstripe sensor.

2.2.5 Embedded Prototypes

To be able to study this influence in more detail, we iterated on the prototypes to have a self-contained version that

we could easily use with different types of base fabric. At the same time we wanted to come closer to an end-user ready implementation and get rid of the complex and fragile wiring of our previous prototypes. Instead of the fabric with the patches, we used a knitted fabric that was manufactured to our specifications by the textile engineering department of our university[5] (Figure 2.6b). It is made of 3 mm wide stripes of conductive thread, separated by 4 mm of non-conductive material. To improve the wearability we reduced the rigid components required to a minimum and placed them on a 19×25 mm^2 large PCB [Thar, 2013]. To connect the PCB to the fabric we used a flexible PCB which we stapled to the stripes on one side and soldered to the rigid PCB on the other side. This can certainly be improved with industrial manufacturing, where the microcontroller would be soldered directly onto a flexible substrate and the connection would be glued or stitched [Linz et al., 2005].

To evaluate the impact of the base fabric, we built a self-contained prototype.

We also simplified the sensing algorithm and adapted the filtering part to better fit the limited floating-point capabilities of the microcontrollers. The two values we need to measure are fold size and fold position. In the first implementations we recorded the entire connection matrix to derive this information, which mostly consists of redundant information for the desired result. Basically, it is not relevant to know to which other stripe the currently measured one is connected, but only that there is a connection between two stripes. The modified sensing algorithm works as follows: Instead of a large matrix of connections, we only have a one-dimensional array the size of the number of stripes containing the connection information. The filtering was reduced to simple integer comparisons to be run smoothly on the microcontroller itself. After a debouncing of the stripes to reduce jitter in the array of connections, fold size and position are compared to specific threshold before triggering the according commands.

To run on a microcontroller, we simplified the sensing algorithm.

With this simple setup, we built five prototypes which only differed in the type of base FABRIC. For the simplest one (*nothing*), we just applied a transfer film for t-shirt prints on the back of the striped fabric. For the other prototypes, we used this transfer film to apply a sheet of silk, cotton,

We tested how well people can manipulate the fold depending on the type of base fabric.

[5] www.ita.rwth-aachen.de

input : stripes: An array containing the mapping from stripe numbers to I/O pins
output: connections[# of stripes]: An array containing the connection information of the stripes
set all pins as output
set all pins to LOW
for *(i < |Stripes|)* **do**
  set stripes [i] as input;
  enable PullUp for stripes [i]
  if *(stripe == LOW)* **then**
    connections [i] = 1;
  else
    connections [i] = 0;
  end
end
set stripes [i] as output;
disable PullUp for stripes [i]

Algorithm 2: The simplified Pinstripe sensing algorithm

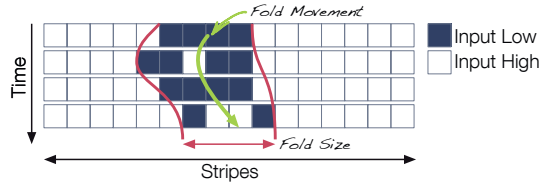


Figure 2.9: The results of the simplified sensing algorithm (Adapted from [Thar, 2013]).

polyester/cotton mix, or fly net to the back of the fabric. The materials range from low (*silk*) to grip (*fly net*) and from flexible (*nothing*) to stiff (*cotton*).

Evaluation of the Base Fabric

Participants had to try out the Pinstripe gesture on all fabrics and rate them on a five point Likert scale according to their appropriateness for this purpose. After that, they could choose their preferred fabric for the remaining target acquisition tasks. We compared different types of CON-

TROL mechanisms regarding their usability by implementing a continuous *volume* slider, a discrete *skip* slider, a combination of the two where the fold size is used to *switch* between the sliders, and an indexed slider where the user has to crumple the sensor to confirm the *select*ion All of these mechanisms were evaluated in a target acquisition task, where a certain value had to be set on a slider. After each condition participants had to fill out a questionnaire covering specific aspects of the usability of that condition: response time, erroneous movement in the wrong direction, no reaction, possibility of error correction, and everyday usability. All participants completed the tasks in the same order.

Results

A total of 16 users (2 female) participated in this experiment with an average age of 32 years (SD = 16.3). The ratings for the five FABRICS can be categorized in two groups (see Figure 2.10): *nothing* (IQR = 1.75), *polyester* (IQR = 1.75), *fly net* (IQR = 1) all have a median rating of 2, whereas *silk* (IQR = 0.75) and *cotton* (IQR = 2) only reach a median score of 4 on a 5-point Likert scale with 1 being the best. Half of the participants chose *nothing* to continue the experiment as they judged it to be the best fabric The remaining users chose *Polyester* (4), *Cotton* (2), and the *fly net* (2). The textile sensor needs to provide a good balance between grip and flexibility. Silk for example, is highly flexible, but the fine structure does not provide much grip, whereas the fly net is on the other side of the spectrum.

Silk is very flexible, but does not provide much grip.

The number of actuation errors, e.g., scrolling in the wrong direction, was highest in the *switch* condition (M=57.4, SD=67.5), but much lower in the remaining conditions: *volume*: M=8.3, SD=13.6, *skip*: M=1.4, SD=1.8, *select* : 4.9, SD=6.9. The difficulties in handling the *switch* control also showed in the participant's ratings. All but the *switch* condition received positive ratings (Figure 2.11). Grabbing different fold sizes showed to be more complicated than expected, which means it should not be used as a dedicated selection switch, but only as a scaling factor for the changes

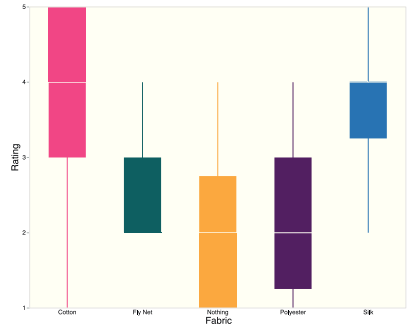


Figure 2.10: Median ratings of the different fabrics on a 5-point Likert scale (1 being the best) regarding their appropriateness for the Pinstripe fold-and-roll interaction (boxes denote quantiles, error bars the range).

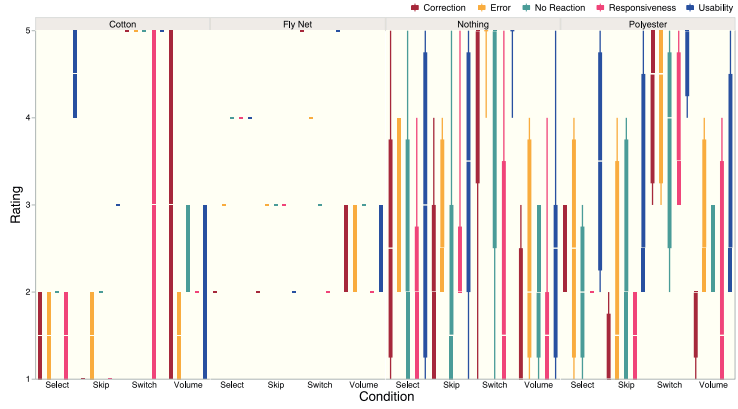


Figure 2.11: Median ratings of the different fabrics on a 5-point Likert scale (1 being the best) regarding the aspects of response time, errors, reaction time, error correction, and everyday usability.

The fold size should only be used to scale the input domain.

in value. Participants did not like confirming the selection with an additional crumple gesture at the end as they perceived this to be an unnecessary additional step. Both the continuous and discrete slider received similarly good ratings.

2.3 Intuitex

While the concept of Pinstripe is simple to implement and robust, it is also quite limited because it can only manipulate one dimension at a time and the fold has to be picked up parallel to the stripes to work well. One way to tackle the problem of limited dimensions would be to integrate several sensor areas for different purposes, e.g., the left arm controls volume, the right arm controls the playlist position. However, the fold orientation remains problematic: With the sensor integrated into the sleeve at upper arm level, the orientation is mostly determined by ergonomic factors, but at lower arm level there is no natural orientation of the fold. To provide a sensor that works independently of the fold orientation we designed a two-dimensional version of Pinstripe called Intuitex.

Intuitex uses the interaction of Pinstripe for 2D input.

Again, we wanted to leverage the textile's affordances of grasping and folding. Instead of parallel stripes, we use a hexagonal pattern of 30 hexagonal conductive pads, which are connected one by one to a microcontroller. To allow an easy movement in all directions, they are embroidered using a circular arrangement of stitches, preventing the filaments to snag on each other when the user moves the fold. The lines from the patches to the landing zone for the microcontroller are insulated by a non-conductive thread stitched over them. The user folds the sensor and moves the fold with her thumb on the surface defined by the remaining four fingers. The working principle is the same as in Pinstripe: when folding the textile, some of the conductive patches get connected. Observing these connections over time while moving the fold gives us a change in values of a 2D coordinate system. To provide an orientation independent two dimensional output, we first need to determine the axis of symmetry in the matrix, i.e., the direc-

Instead of stripes, we use conductive patches.

tion of the fold. Movement along this axis will be mapped to output on the X-axis while a perpendicular movement to the axis will result in changes of the Y-axis (Figure 2.12). To reduce sensor jitter, we filtered the connection matrix by taking the two last measurements into account with the following formula $M_t = (M_{t-2} \vee M_{t-1}) \wedge M_t$. We apply a

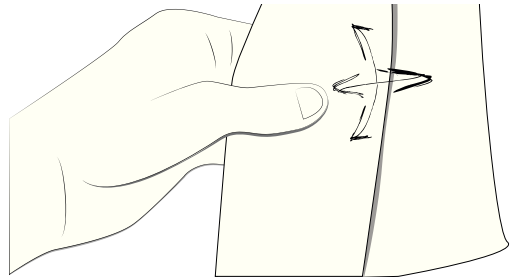


Figure 2.12: Input is generated by creating a fold and moving between the fingers along the two axes. Movement along the fold is mapped to Y-axis, while movement perpendicular to the fold is mapped to the X-axis (Taken from [Heller et al., 2015]).

principal component analysis (PCA) on the connection matrix to determine the fold angle. We then divide the connection matrix along the line of symmetry defined by the first PCA component (first eigenvector) followed by a calculation of the center of gravity on one half of the matrix. We then apply a simple low-pass filter ($\alpha = .3$) on these coordinates before they are communicated to the host.

Since every patch requires its own connection to the microcontroller, we had to think about how to easily connect a large number of conductive threads. Tying the threads to conductive pads like with the LilyPad Arduino was not an option as we wanted to be able switch between various embroidered patterns during our prototyping process. This possibility is also potentially relevant for the final product, as it allows to remove the electronics before washing and to use the electronics on different textiles. We developed a clipping mechanism that simply pushes the ends of the conductive threads against a contact area on the

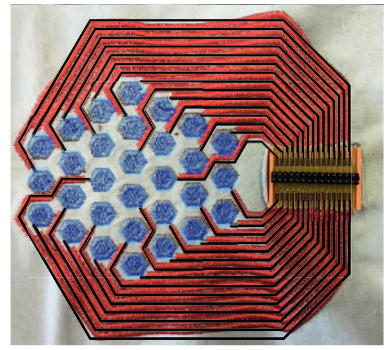


Figure 2.13: The Intuitex prototype with 30 conductive pads connected to the landing zone for the microcontroller on the right.

We developed a clipping mechanism to attach the electronics to the textile.

PCB containing the microcontroller. During the prototyping process, we encountered problems with unwanted connections, which we could track down to single fibers detaching from the thread and bridging the gap to the neighboring contact pad. We solved this problem by creating little bins in the plastic clip (Figure 2.14a) which fix and separate the ends of the conductive thread. The snaps at the edges of the plastic clip (Figure 2.14b) reach the PCB on top of the fabric through two holes in the fabric, and pull the PCB down against the fibers, ensuring a stable connection.

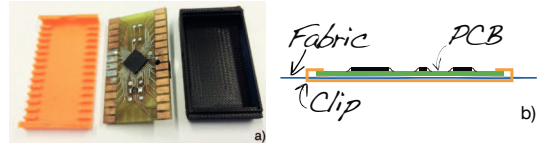


Figure 2.14: The orange clip provides bins for the endings of the conductive thread. The PCB just has simple contact areas on the bottom side and is pressed against the fabric by the orange plastic clip. The black part is the top case of the enclosure (Taken from [Heller et al., 2015]).

2.3.1 Limitations

In its current state, the prototype has only a very coarse resolution as the number of conducive pads is low due to the manufacturing process. When the conductive thread passes the needle during the stitching, single fibers detach from the thread and may create unwanted connections. This limits the minimal distance between two parallel lines of conductive thread to around 3 mm. These restrictions make it very hard to control something like a cursor with this sensor as only few discrete steps can be differentiated. However, controlling a cursor is not a setting suitable for wearable interfaces anyways as it requires high resolution visual feedback. Through its low physical resolution, the sensor basically discretizes the user's input, making it a potential controller for different types of commands. If the software interface is adapted to the wearable setting, actions such as navigating a playlist, changing the volume, or taking a call are feasible to perform. For binary decisions, we only need to differentiate between up/down or left/right, while we have to take the movement distance into account for continuous value input.

The resolution is limited by the manufacturing process.

2.4 Fabritouch

As smartphones are more and more replacing the dedicated portable MP3 player, and at the same time provide more functionalities and rich interaction through touchscreens, we explored the possibilities to extend wearable input from the one dimension that Pinstripe offers to two dimensional input. During our user tests with Pinstripe, we often observed that people first tried to interact with the textile interfaces as they know it from touchscreens, by tapping and swiping. While this contradicts our initial thoughts of leveraging the affordances of cloth, we wanted to investigate how to take advantage of this, now ubiquitous, interaction pattern. Furthermore, the touchpad allows a much higher resolution than our Intuitex approach.

The DIY community has prototyped different approaches

The DIY community has prototyped textile touchpads.

to textile touchpads[6], but it is unclear how well they perform in practice. A standard capacitive touchscreen can be re-calibrated to work reliably under an additional layer of fabric, which allows rich and precise input [Saponas et al., 2011], even up to single letters. Thomas et al. [2002] evaluated the placement of a regular PC touchpad at different body locations and with at different postures. They found that a placement of the touchpad on the upper thigh works best when sitting, kneeling, or standing, while obviously, it does not work in a prone position. Thus, while it is possible to use a standard capacitive screen through clothes, the rigid enclosure reduces the *wearability*. Simply removing the enclosure to make the sensor flexible will not work if the sensor is placed directly on the skin. Additional shielding or spacing layers are needed and the calibration has to be adapted further.

Capacitive input using textile sensors on the body is difficult.

2.4.1 Textile Touchpads

Because of its simplicity and robustness, as most textile touchpads we opted for a resistive implementation based on conductive fabric and piezoelectric foil. A *spacing mesh* separates a *piezoresistive foil* from a layer of *conductive fabric* to prevent any touch detection when no finger is placed on the surface (Figure 2.15). The piezoresistive foil's electrical resistance varies with the force applied at a touch point and its distance to the points of measurement. If we apply a reference voltage to the conductive fabric and press on the surface, we create an electrical connection between the fabric layer and the foil by bridging the gap created by the spacing mesh. With the measured relative voltages at the four corners of the foil we can calculate the position and pressure level of a single touch.

2.4.2 The Fabritouch Prototype

The placement of wearable controls on the body has received great attention and following the literature [Holleis

[6] instructables.com/id/EJKTF3WGV490JGK

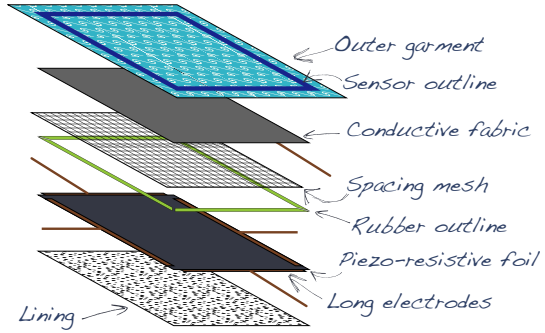


Figure 2.15: The layered architecture of our textile touchpad.

et al., 2008; Karrer et al., 2011; Thomas et al., 2002; Rekimoto, 2001; Wagner et al., 2013], we chose the upper thigh to place our Fabritouch pads because it offers a large interaction surface and can easily be reached in different body postures. To determine the appropriate size and position of the sensor, we let 26 participants perform a series of simple gestures (e.g., circles, lines, crosses) with baking flour on a piece of fabric attached to their upper thigh. After each participant, we took a picture of the fabric to document the coverage of these gestures. Superimposing these pictures showed that an 80×80 mm^2 sized interaction surface was suitable. It should be placed parallel to the thigh, centered 285 mm down from the waist and 10 mm towards the outside from the top of the thigh (Figure 2.16).

We determined optimal size and position for a wearable touchpad.

Based on these findings, we constructed a series of prototypes. Figure 2.17 shows the final version[7]. We used .1 mm thick Caplinq ESD protective sheet as piezoresistive foil. The conductive textile layer was made of Shieldex Med-Tex180 silver-plated nylon cloth. The spacing layer consists of tulle, a textile mesh, with a thickness of .45 mm and a hole diameter of around 2.1 mm. Placing the measuring electrodes at the corners of the piezoresistive foil has the disadvantage of having distorted measurements, mostly towards the borders of the layer, which requires a

[7]Build instructions at hci.rwth-aachen.de/fabritouch

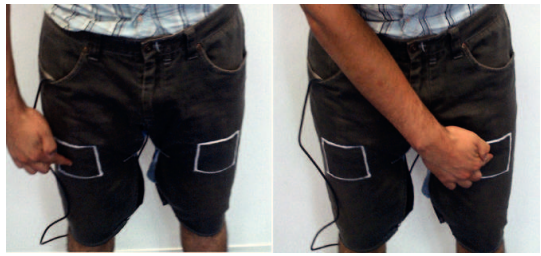


Figure 2.16: Two Fabritouch pads integrated into a pair of trousers. Users tried both *parallel* (left) and *crossed* touch gestures (right) (Taken from [Heller et al., 2014a]).

more complex calibration and reduces resolution. We used long strips of copper foil placed pairwise on top and bottom side of the piezoresistive layer as electrodes measuring along the two axes. We raised the border of the sensor surface by placing a rubber outline under the outer garment as we noticed that it is difficult to feel the borders of the sensing area. An Arduino board collects the measurements and communicates (x, y) coordinates at a resolution of 100×100 points (31.75 ppi) at 30.3 Hz to the attached computer. There, we use the 1€ filter [Casiez et al., 2012] to reduce sensor noise in software. The pressure signals were quantized into binary single-touch input.

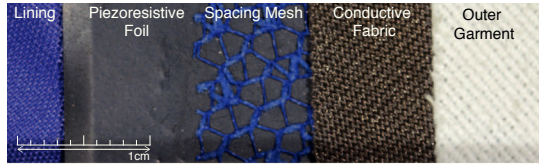


Figure 2.17: Fabritouch layer materials, placed side-by-side to show material structures (Taken from [Heller et al., 2014a]).

2.4.3 Evaluation: Support Surface Rigidity

While many sensors are demonstrated and tested on rigid surfaces such as tables, actually integrating them into cloth-

ing leaves them on top a flexible, nonplanar surface without firm support, and subject to body movements. We hypothesized that this would significantly impact touch input performance, and tested this in a study.

Procedure: Users used the Fabritouch pad to manipulate a cursor and acquire circular targets (70 px diameter) that randomly appeared in a 5×5 grid on a desktop computer screen. To manipulate the cursor, users depressed the pad to generate input signals, which were mapped absolutely to 800×800 px on-screen. To acquire the target, users had to stay engaged in the target area for at least 2 seconds (a visual countdown was provided). Lifting the finger reset this engagement.

We performed a within-subject study; the touchpad was placed either on a table or on the upper thigh (counterbalanced order, 20 repetitions per condition). For the thigh condition, the touchpad was mounted on a large piece of cloth firmly attached to the users' trousers. Prior to each condition, users familiarized themselves with the touchpad until they felt comfortable. The dependent variable was task completion time. We log-transformed the data and used mixed-model ANOVA with USER as a random effect.

The support surface rigidity has a significant impact on the usability of our touchpad.

Participants: We recruited 26 volunteers (8 female, age 18–34, M = 25) from our campus. All had a computer science background and reported high familiarity with laptop touchpads (Mdn = 5 out of 5-point Likert scale).

Results: Users performed *twice* as fast on the table (M = 5.90s) as on the thigh (12.00), $F_{1,897} = 296.64$, $p < .001$, Cohen's d = 1.01 (Large effect size). The lack of statistical significance of repetitions ($F_{19,897} = 1.11$, $p = .3302$) and interaction effect ($F_{19,897} = 0.82$, $p = .6873$) indicates no learning effect.

We also observed that all users applied more pressure in the thigh condition. Even so, they perceived this condition as less stable (P10: *"It felt like writing on a sheet of paper on your thigh"*, P4,7: *"You should really hold your breath"*). Both increased pressure and perceived instability could be a cause of the slower performance in this condition. Even though

the muscular nature of the upper thigh provides a rather firm base, finger pressure is still distributed over a larger area, reducing the sensitivity of the touchpad. These factors indicate that pointing input may not be suitable for fabric touchpads.

These results suggest stark differences of user behavior between the rigid support of the desk and the soft support of the thigh. Therefore, it seems crucial to assess and fine-tune wearable user interfaces with realistic sensor placement, on the body rather than conveniently on a lab desk.

2.4.4 Evaluation: Usage Posture

Users' posture influences their performance in wearable UIs [Thomas et al., 2002]. Additionally, the progress of each touch movement changes trackpad properties, such as its flatness, rigidity, or contact to the surface below. In this study, we investigated how these two factors influence gesturing performance. We chose horizontal and vertical swipe gestures for their simplicity and ubiquitous use in 2D touch UIs.

We evaluated our prototype while sitting, standing, and walking.

Procedure: In our within-subject study, users navigated a two-level hierarchical menu [Zhao et al., 2007] using Fabritouch integrated into a pair of trousers (Figure 2.16). Navigation on the top level was performed using horizontal swipes while the second level was navigated with vertical swipes. A rubber band ensured tight fitting of the trousers. The independent variables were POSTURE = {sitting, standing, walking} and swipe DIRECTION towards the user's {FEET, HEAD, non-dominant hand (NH), and dominant hand (DH)}. HEAD swipes mapped to moving the cursor downward (Figure 2.19), as recommended by [Thomas et al., 2002] and supported by our pilot study (6 users). Horizontal swipes were mapped like on a smartphone: Swiping towards the left moved the selection to the left.

Participants had to navigate a 2D menu.

As with a standard menu bar, the top level and the current submenu were always visible. A trial ended when the cursor reached the target item; the subsequent trial contin-

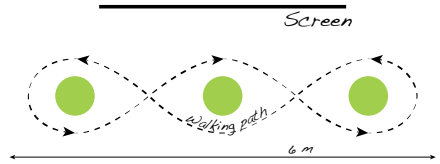


Figure 2.18: The path the users had to follow in the walking condition (Taken from [Heller et al., 2014a]).

ued without resetting the position of the selection. The sequence of menu items was predetermined to balance the number of swipes across all directions. Users acquired five targets for training and seven for testing per POSTURE, resulting in $M = 68.60$ swipes per POSTURE ($SD = 8.86$).

In the WALKING condition, where users had to walk around a predefined path (Figure 2.18) in the room, we projected the menu on a wall to ensure its visibility.

Data analysis: For each recognized swipe, we analyzed overall task completion TIME, DURATION of individual gestures, and the dimensions of the gesture bounding box: the LENGTH along the swipe direction and the DEVIATION orthogonal to the swipe direction. All variables were log-transformed before analysis with a mixed-model ANOVA with USER as a random effect, followed by a Tukey HSD for post-hoc tests. Descriptive statistics were calculated by inverse-transforming log statistics.

Participants: We recruited 17 volunteers (3 female, age 21–34, $M = 26$) from our campus. Six were ambidextrous[8], and two were left-handed. They all had a technical background and reported high familiarity with typical laptop touchpads ($Mdn = 5$ out of 5-point Likert scale).

Results and Discussion

Posture: There was a significant effect of POSTURE on TIME $F_{2,32} = 3.44, p = .0442$. Post-hoc testing indicates that

[8] They scored less than 4th decile in Edinburgh laterality

Effects		Dependent variables: Duration		Length		Deviation	
	df	*F*	*p*	*F*	*p*	*F*	*p*
Posture	2, 176	15.35	<.0001	2.86	.0602	0.32	.7251
Direction	3, 176	3.76	.0119	12.61	<.0001	14.37	<.0001
Posture * Direction	6, 176	0.43	.8597	0.83	.5502	1.86	.0904

Table 2.2: The effect of direction on the gesture is significant across the board while posture has significant effect only on the duration.

To be usable under all conditions, commands should be triggered by simple gestures.

only walking (M = 470s) took significantly longer than sitting (351). Standing (409) did not significantly differ from both. Gesture duration while walking (M = 1.42s, 95% CI [1.35,1.50]) was significantly shorter than sitting (1.70, [1.60, 1.81]) and standing (1.71, [1.61, 1.82]) (cf. Table 2.2). The longer TIME and the shorter DURATION suggest that gesturing while walking was more difficult than in other postures.

Gesture directions: DIRECTION has a significant effect (Table 2.2). NH swipes were slowest (1.73s [1.57, 1.92]) and were significantly different from FEET swipes (1.50 [1.41, 1.59]), which were fastest. Users were significantly less precise in performing horizontal swipes (DEVIATION M = 1.85mm [1.59, 2.16]) than vertical swipes (1.35 [1.24, 1.46]) (Figure 2.19). NH swipes were significantly shorter (LENGTH M = 5.03mm [4.79, 5.28]) than other directions (5.68 [5.46, 5.91]).

Horizontal swipes (NH, DH) were harder than vertical ones. One reason was that horizontal swipes generated more wrinkles in the fabric while vertical swipes (especially FEET) stretched the cloth. The upward movement from the outside of the thigh towards the center in NH accentuated this effect, producing shorter swipes. The non-significant interaction effect indicates that the movements during walking did not make any particular DIRECTION harder.

Gesture location: While users reported that the ridges allowed them to orient their finger (Mdn = 4 out of 5-point Likert scale), most gestures were performed in the middle

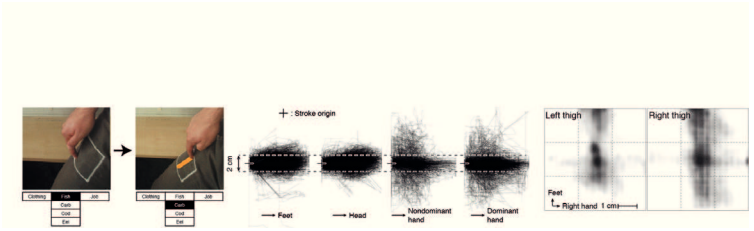


Figure 2.19: Left: Menu navigation mapping in Study 2. A swipe towards the user's head moves the menu selection downward. Middle: Gesture traces in different directions show high deviation of horizontal swipes towards the non-dominant hand. Right: Contour plots show the density of touch locations from Study 2. Users perform gestures mostly in the center third of the touchpad (Taken from [Heller et al., 2014a]).

third of the sensor (Figure 2.19). This indicates that users used the ridges to orient their finger initially but performed the swipes without relying on the ridges. Informal observations during our study and qualitative feedback indicated that users rarely looked at the touchpad during the test.

Handedness preference: Despite no explicit instructions, almost all users used the touchpad on the side of their dominant hand. Only P1, who was right-handed, used the left touchpad with his right hand to *"give it a try"* in the STANDING condition. His performance here did not differ from others'.

2.4.5 Design Implications

Use vertical swipe gestures instead of horizontal ones: Users perform vertical swipes faster and in a smaller bounding box than horizontal swipes, which should be considered in the gesture recognition. Horizontal swipes from the outside of the thigh towards the center result in dragging upwards which requires a constant complex adaptation of pressure and should therefore be avoided. Due to the high touch pressure required, we do not recommend using this touchpad type for pointing.

Vertical swipes are easier to perform on the thigh than horizontal ones.

Performing gestures on fabric touchpads while walking is harder: If activity detection is possible, e.g., via accelerometers, relaxing the gesture duration criteria of the gesture recognizer's tolerance during walking could reduce gesturing difficulties. Since the gesture duration is shorter, designers should avoid including both sliding and flicking in the gesture alphabet used while walking.

On-body and multi-posture tests are necessary: To cover the breadth of realistic user experiences, fabric touchpads need to be tested on-body in both static and dynamic postures. According to our study, we recommend testing with at least standing and walking postures.

2.5 Conclusion

In this chapter we investigated the extension of the haptic interaction bandwidth in audio playback interfaces using the example of wearable controls. While we did not extend the functionality of audio playback interfaces, we adapted the control mechanisms to leverage our manual skills to a larger extent. As we could not revert to metaphors whose origins evolved over centuries as in the case of drawing, we utilized the natural affordances of fabric to make clothing a ubiquitous interaction surface. Pinstripe detects the size and relative movement of a fold created by the user pinching into a piece of cloth. This manipulation can easily be mapped to changes of a continuous linear value, like volume, making it a natural interface for portable music players. The gestural input on the two dimensional textile touchpad Fabritouch accounts for the touchpad-style input users are acquainted with. It avoids the problem of involuntary activation we know from wearable buttons and, if of adequate size, provides a convenient interaction surface.

Since no evolved ancestor exists, we looked for other interaction surfaces.

Participants in our first study successfully used Pinstripe to change volume and select tracks from a playlist in a music player application. In a second study, they used Fabritouch to successfully select items from a two dimensional graphical menu. While it required some practice to confidently

manipulate these interfaces, this is also the case for memorizing the button arrangement on your music player.

The physical resolution of textile interfaces is comparatively low.

Yet, they leverage our fine manual motor skills.

Compared to the precision achieved by standard electronic circuits with sub-millimeter pitch, the possibilities of textile manufacturing are fairly limited. The flexibility of the support material and the real-world problems of filaments coming loose from the conductive thread make it difficult to create input devices with a comparable resolution to the one we know from physical computing. This leads to the paradox that we leverage our very fine manual motor skills using devices with a low physical resolution. However, the physical resolution of our wearable controls is still substantially higher than that of a binary button.

In this chapter we analyzed possibilities to increase the interaction bandwidth via haptics. In the following chapter we build on an existing system with high haptic interaction bandwidth to extend the visual modality.

Chapter 3

Visual Augmentation of a Digital Vinyl System

"...there are three levels of design: standard spec, military spec, and artist spec. Most significantly, I learned that the third, artist spec, was the hardest."

—Bill Buxton [Buxton, 1997]

3.1 Introduction

To explore the effects of increasing the interaction bandwidth of the visual modality, we will build on an audio playback controller that already takes rich haptic input: the turntable. From all the analog audio playback interfaces, the turntable is probably the one which best matches our

Publications: DiskPlay was published as a note and presented as interactivity installation at CHI '12 [Heller, Borchers, 2012]. The second iteration was published as short paper and demo at NIME '14 [Heller, Borchers, 2014b]. For all publications, the author of this thesis was the main author. Both Justus Lauten [Lauten, 2011] and Sebastian Burger [Burger, 2013] worked on this project as part of their Bachelor or Diploma thesis under the supervision of the author of this thesis.

current concepts of tangible interaction and direct manipulation. Although it seems old-fashioned and primitive, in contrast to many digital players the turntable provides immediate access to playback controls and detailed visual feedback. Its basic interface consists of a Play/Pause button, a possibility to switch to the according playback speed (33 or 45 RPM), and an optional pitch control that allows to adjust playback speed by around $\pm 10\%$. The difference to the transport controls on other players is that the navigation on the medium is performed by displacing the stylus on the record by moving the tonearm. While there is no precise timing information available for navigation, the different groove styles on the record clearly show where the track starts and ends as those in the lead-in and lead-out areas (before and after the actual track) are spaced much more loosely (Figure 3.1a). Additionally, grooves in louder parts of the track are larger and need more spacing than those in quiet parts [Schlager, 1994] (Figure 3.1b), making the song's structure visible on the record which can be used to navigate to certain parts of a song, e.g., a break, and allow similar navigation as with a waveform visualization on the computer. Playback progress is easily perceived by looking at the position of the stylus within the track.

Although a purely analog technology, the turntable is still common DJ equipment.

Traditional vinyl records contain visual cues to their content.

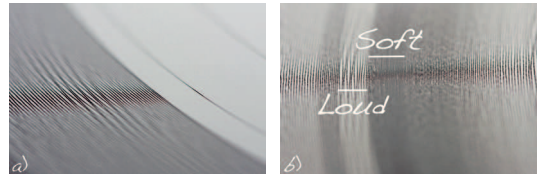


Figure 3.1: Physical structure of a vinyl record. **(a)** The lead-out at the inner end of the record looks different from the grooves inside the track. **(b)** Groove spacing varies depending on the volume of the audio signal (Taken from [Heller, Borchers, 2012]).

The intended way of playing back a record on a turntable actually does not include direct manipulation of the record itself. In fact, to achieve a lasting high audio quality one would be very careful to keep any form of dirt off the record. But the open design allowed different interactions

to emerge, like DJs stoping the record and spinning it back and forth by simply touching the record. Around the direct manipulation of stylus and record emerged the *scratching* technique, an own form of art which uses the turntable more as musical instrument than as mere playback tool [Hansen, 2010]. The creation of new sounds by quickly moving the vinyl back and forth, which results in playing a small audio sample over and over at different speeds and in different directions, requires a lot of practice to be performed well. The fact that DJs who put effort and time to perfect their skills did not want to switch to the less expressive CD player made this purely analog technology an irreplaceable performance tool [Lippit, 2006] and survive long into the digital era.

DJ value the turntable for its unique haptic nature.

3.1.1 From Analog to Digital

Some CD players, such as the Numark CDX[1], the Denon DNS5000[2], or the Technics SL-DZ1200, are equipped with a turntable-like control interface to mimic its handling, but their adoption in the community was poor. Since today's music production workflows are largely digital and digital music distribution channels are fast and cheap, the market for vinyl records is small. While you can easily carry your entire music library on a laptop, the physical records a DJ needs for a set become heavy and cumbersome quickly, and wear and tear of a record that is played extensively limits its lifespan [Hansen, Bresin, 2010].

Some CD players try to mimic the handling of a turntable.

When using traditional records increasingly felt out of date, digital vinyl systems (DVS) bridged the gap between the haptic feedback the DJs are accustomed to and the digital media on the computer. Today's most important competitors are Serato's Scratch Live[3] and Native Instruments' Traktor Scratch[4], which both work with a similar setup. Instead of a music track, the control vinyls for DVS contain an analog version of a digital timecode combined with

Digital vinyl systems lifted the turntable in the digital era.

[1] www.numark.com/cdx
[2] denondj.com/products/view/dn-s5000
[3] serato.com/scratchlive
[4] www.native-instruments.com/traktorscratch

Figure 3.2: *Traktor Scratch Pro* user interface. The left track is close to the end, indicated by the waveform overview flashing red.

a sine-wave signal [Wardle, 2007]. The sine wave allows very fast detection of changes in speed, as slowing down the record will result in a lower pitch and in a higher pitch when speeding up the record. The digital timecode is used to determine the absolute timing information on where the stylus currently is on the record. The timecode record is played on an unmodified turntable and the signal is routed to the computer trough an additional audio interface. There it is interpreted into parameters such as playback speed, direction, and absolute playback position and mapped to the MP3 playback. The result is then sent back to the DJ mixer through the audio interface (see Figure 3.3). This lets DJs build on their perfected manual skills with the usual equipment, while providing the advantages of digital media storage and playback, including independent control of pitch and tempo.

Timecode records contain an analog version of a digital timecode.

Timecode records do not provide visual cues related to the song that is played.

To be usable as generic controller for a large variety of songs, the control records contain around 10 to 17 minutes of timecode. However, neither the length of the timecode nor the physical features on the record that were visual cues on traditional records relate to the song loaded in the software (Figure 3.4). As visualization and control are separated in this kind of setup, this forces the DJ to look at the computer screen to find essential information such as the remaining time in the song while at the same time handling the turntable. While all information that was visible on a traditional record is still there, it is not co-located with

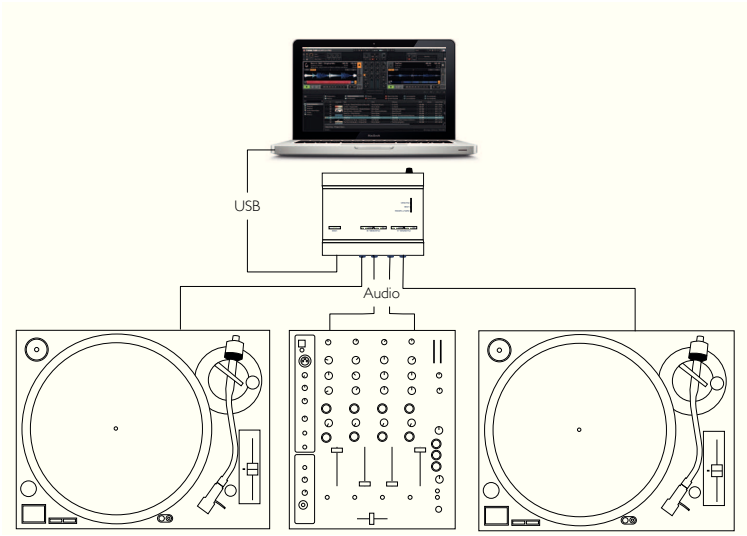


Figure 3.3: The DVS setup: the timecode signal is sent to the computer, analyzed, and mapped to MP3 playback. The result is sent back to the mixer.

the haptic input anymore. This reduction of visual output bandwidth complicates task such as navigation since orientation based on the groove pattern is only possible in the time domain. Searching for a certain point in a track thus degrades from placing the stylus close to the designated groove to a binary search with visual feedback on the distant screen. This leads to a phenomenon where the DJ seems to constantly stare at his laptop and loses the connection to the crowd and called the "Serato face" [*The Serato Face* 2013].

Haptic input and visual output are physically separated.

In this chapter, we will show how increasing the interaction bandwidth through visual augmentation of the turntable can bring back the missing features of traditional vinyl records to DVS setups. This recreates the embodied interaction unit that the turntable was in the analog era.

Figure 3.4: The physical features of a *Traktor Control Vinyl Mk II* timecode record. The timecode is textured in sections of 1 minute length to simplify the orientation, however, these features are not related to the track loaded in the software.

3.1.2 Terminology

DJ terminology and tasks have been elaborately explained in [Beamish, 2001], but we will briefly recap some of the terms. The DJ culture builds around two types of DJs: the scratch DJ and the beatmixing DJ. For both types, the basic setup consists of two turntables and a mixing console. A felt slipmat between platter and record lets the DJ manipulate the record independently of the rotating platter beneath (Figure 3.5). The scratch DJ uses short samples of a song to create a new one by quickly moving the record back and forth and switching between the two records. The beatmixing DJ plays a series of tracks, called a *set* or *mix*, thus creating a single longer, seamless new track. She starts by playing a record on the first turntable. While this *outgoing* track is playing, she puts another track on the second turntable and, using headphones, matches the tempo of this second, *incoming* track to the one of the first, but without playing it to the audience yet. To synchronize the speed, the DJ will first search for the first beat on the in-

coming track and, without stopping the turntable, halt the record with her fingers. She then waits until the according beat of the outgoing track is reached and releases the vinyl such that the tracks are now playing in parallel. By accelerating or slowing down the platter, the DJ keeps both tracks in sync while adjusting the playback speed with the turntable's pitch fader until both tracks play at the same tempo. At some point, usually close to the end of the outgoing track, the DJ will mix the incoming track into the outgoing one, such that the audience cannot determine where the outgoing track ends and the incoming one starts. After this transition, she switches the track on the first turntable to a new incoming track, and the process starts again, with the two turntables switching roles. This way, "each song will be mixed into the next to give the appearance of a seamless stream of music" [Beamish, 2001].

DJ mix several songs to give an appearance of a seamless stream of music.

3.2 Related Work

Numerous projects in research and industry have aimed to enhance turntable-based interaction in the DJ context.

TIMBAP [Pabst, Walk, 2007] focuses on turntable-based navigation of a media library. Using a top-mounted projector, the artwork of each piece is displayed on the record, and the DJ can navigate through his music library by either seeking linearly (i.e., spinning the record) or searching tracks by a tag cloud interface that is manipulated by displacing the stylus on the record. However, this system is used only for track selection. It does not support in-track navigation and does not bring back the individuality of the medium.

An artistic installation using timecode records is Vinyl+ [Bohatsch, 2010]: an image of colored bubbles and dots is projected onto the record. When the stylus passes a dot, a specific sound sample and a visual effect are triggered. Vinyl+ connects visual and auditory channels, but does not support navigation inside existing tracks, making it more of a musical instrument.

The *Lupa* hard- and software interfaces [Lippit, 2006] are designed to prohibit all physical and minimize visual interaction with the laptop during the performance. The user interface provides an at-a-glance overview and does not support presets or automation. This design promotes the liveness of a performance and creates an experience for the audience that is truly unique.

Several research projects augmented the turntable.

D'Groove [Beamish et al., 2004] is a force feedback-enabled turntable to explore new ways of manipulating music. The turntable has distinct marks for the four beats of a bar, and its rotation speed is coupled to the song tempo such that the beat marks form a spatial landmark while beatmatching two songs. A motorized slider indicates the progression of the track over time and allows to control the playback position. Among the force feedback modulations implemented are a *bump-for-beats* mode providing a physical sensation of each beat, and a *resistance* mode that makes it harder to move the record "when it is playing an area of high-energy music". The system conveys additional track information over the haptic channel, which supports local in-track navigation. However, D'Groove introduces new turntable hardware with new interaction techniques, and it does not provide an at-a-glance overview of the track structure.

A series of projects looked into the use of multitouch screens and interactive tabletops for their use as DJ controllers. Lopes et al. [2011] compared the mixing performance of a multitouch DJ interface running on a tabletop to traditional vinyl, DVS, and a standalone software. Although the participants showed great interest in the system, they took longer to complete the mixing task using the multitouch installation. While the multi-touch system suits the expectations of mix-DJs, especially scratch-DJs preferred the turntable and DVS, since they provide better haptic feedback and control.

Instead of having a virtual turntable rotate, the interface can also be a viewfinder moving over the waveform, also referred to as the "conveyor-belt" metaphor [Lopes et al., 2011]. Consisting of a large touchscreen, the *Attigo TT*[5] is

[5] http://www.attigo.co.uk

designed as an in-place substitute for the turntable. Similar to vinyl records, it lets you manipulate the song by touching the waveform, making it easy to shortly stop the track, scroll forward and backward, or scratch, but it lacks haptic feedback and requires learning a new set of gestures.

Touchscreens and tablets are now used for DJing as well.

Being portable and providing enough storage capacity for a large music collection, multi-touch tablets are an attractive platform for DJs. *Traktor for iOS*[6] does not mimic the traditional setup of two turntables and a mixer, but provides an interface adapted to small touch screens. It uses two conveyor belts to show the waveforms and provides a two-channel mixer with equalizer and effects section. The integration of loops and effects extends the DJ's performance from mere playback to live remixing of tracks. Fukuchi [2007] presented a similar multi-track mixing interface that allows rapid switching between tracks by just dragging from one track to the other, thus using the entire surface as crossfader.

Like all touch devices, it does not provide any haptic feedback and the space for artistic expression is fairly limited. In their analysis of scratching, Hansen and Bresin [2010] described that the crossfader can be opened in bursts as short as 10 ms, for which the predominant techniques require a physical control [Hansen, Bresin, 2006]. This need also explains the growing number of dedicated hardware DJ controllers available for iPad and iPhone.

Touchscreens lack haptic feedback.

To speed up navigation in the track, modern DJ CD players like the Pioneer CDJ-2000 or controllers like Native Instruments' S8[7] provide a touch-sensitive strip that is used as a slider to jump to a certain position in the track. The song is mapped to the length of the strip, meaning that if you press in the middle of the slider, playback jumps to the middle of the track.

[6] http://www.native-instruments.com

[7] native-instruments.com/de/products/traktor/dj-controllers/traktor-kontrol-s8

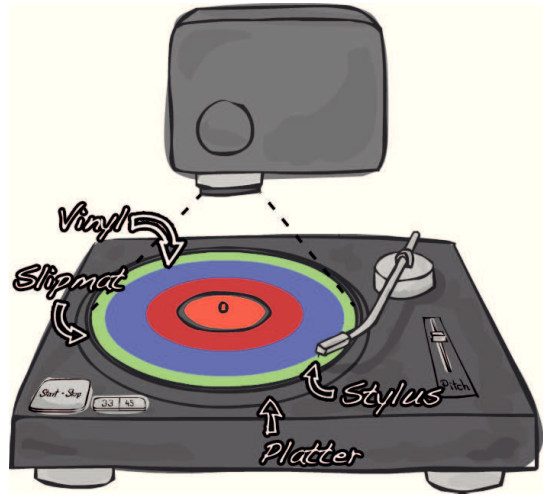


Figure 3.5: Using top projection, DiskPlay augments a white timecode vinyl disc with information about the structure of the current track, such as its starting point, length, and cue points. (Taken from [Heller, Borchers, 2012]).

3.3 The DiskPlay System

DiskPlay re-creates the visual cues of traditional vinyl on timecode records.

The focus of DiskPlay is to augment a DVS setup such that the information that was available on a classic vinyl becomes visible again. Current DVS implementations separate visualization (on the computer) and control (on the turntable), which requires the DJ to repeatedly glimpse at the computer display while he is mainly working with the turntable. For example, most DVS software implements some kind of visual alert to inform of the upcoming end of a track, e.g., by some flashing UI element (Figure 3.2), which is easily missed if the DJ is focussed on the turntables or requires permanent attention. DiskPlay integrates visualization and control on the turntable by augmenting the timecode record with important information (Figure 3.6).

Figure 3.6: DiskPlay displaying a track. (a) Remaining playback area (blue). (b) Part already played as progress indicator (green). (c) Unused timecode part (red). (d) Cue points (yellow dots) (Taken from [Heller, Borchers, 2012]).

DiskPlay visualizes track start and end, playback progress, and cue point position.

The information which was most obvious on traditional records, but which is not visible on the timecode record anymore, is track length. When loading a track onto a virtual deck in the software, DiskPlay colors the part of the timecode vinyl that covers the playback length of the track in blue while the remaining, unused part is colored red. This shows where the track begins and ends.

As playback progresses, the part of the track that has already been played is colored green to indicate the progression over time (Figure 3.6).

We took further design inspiration from watching several videos of the DMC World DJ Championships[8]. To quickly jump to specific points in a track, DJs use stickers that they place as markers on the record. If placed correctly, these stickers can also push back the stylus by one groove, thereby creating an infinite loop. While with traditional vinyl, one can leave the stickers on the record, this conflicts with the idea of the timecode record as generic controller. Therefore, we integrated the bookmarking idea of

[8] dmcdjchamps.com

cue points and visualize them on the record with yellow dots. To simplify navigation to a cue point, an orbit, a concentric black circle is drawn with the dot's radial distance to the center as radius. This helps place the stylus in the correct groove while the record is spinning.

To study our interaction design, we built a prototype around a standard DJ turntable using a white timecode record and a projector above it (Figure 3.7). We extended the open-source DJ software Mixxx[9] with an on-record display. Mixxx [Andersen, 2003] is a software framework to explore new interaction techniques with regard to DJ applications. Its flexible software design makes it easy to integrate modules for different input and output modalities. We just added an additional full-screen output window that was rendered on the turntable by the projector. The entire timecode-processing was done by Mixxx and we just retrieved timing and track information to adjust our visualization. The user interface of Mixxx is common among popular DVS software, providing our users with a known environment. One of the turntables was equipped with DiskPlay, the other one with the current standard tool set. Tasks of mixing towards and away from the DiskPlay turntable stressed different aspects of the visualization. While actively mixing with DiskPlay, track start and cue points are more important, whereas the playback position and track length are needed while handling the other turntable.

We integrated this visualization into Mixxx.

3.3.1 Evaluation

Since DiskPlay is a tool to support artistic expression, we conducted an observational study with four professional DJs, referred to as DJ1–DJ4, to gather qualitative feedback. All DJs had between 5 and 20 (average 12) years of experience, and between 0.5 and 5 (average 2.5) years of experience with DVS. Our setup consisted of two Technics SL–1200MK5 series turntables (one equipped with DiskPlay) and a standard Gemini BPM–1000 mixing console. After a brief explanation of the system and importing each DJ's

[9] www.mixxx.org

personal music library into Mixxx, each participant mixed for 25 minutes with traditional timecode records as control condition. We then asked participants to take a short break, and enabled the DiskPlay visualization. Starting with the augmented turntable, the DJs continued mixing in the experimental condition for another 30 minutes. Participants were encouraged to think aloud and mention anything interesting or intriguing. We then let the DJs perform a seek task to measure the time they took to find a certain point they selected from one of their records. We repeated the task three times, with the order of the conditions *DiskPlay*, *DVS*, and *By Ear* being randomized. After the mixing session, we asked participants about anything we had noticed, along with some general feedback questions about the systems. In the following, we present our observations and insights from those semi-structured interviews.

We evaluated DiskPlay with four professional DJs.

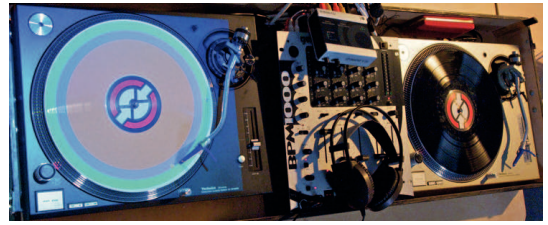


Figure 3.7: The DiskPlay test setup with the augmented turntable on the left and the traditional timecode vinyl on the right. The projector (not in the picture) is above the left turntable (Taken from [Heller, Borchers, 2012]).

The number of attention switches between the computer screen and the turntable largely differed between the participants and ranged from heavy usage to using the display only for selecting a new track. The usage of the computer screen was not related to the amount of time that the DJs had been working with vinyl records. DJ1, who had worked with turntables for eight years, but of which only half a year with timecode vinyls, made heavy use of the software and its visual aids whereas DJ3, who had five years of experience with turntables and three of them with DVSs barely used the computer display at all.

Without DiskPlay, all participants used the computer display to orient themselves in the track structure and to see how long a track still had to play. This changed with DiskPlay. DJ2 only used the computer to load the next track, then solely worked with the turntables and the mixer. DJ1 and DJ4 used the cue point functionality to mark the beginning of the part of the song they wanted to use. Some tracks contain long intros that DJs often skip, at least while beatmatching the tracks, as these intros mostly contain only light rhythmic cues. DJ1 also used the cue point visualization for coarse in-track navigation, but then reverted back to the computer display. He explained that he could not hit the exact groove of the cue point and did not know if he had to spin the record one, two, or three times to reach the cue point. DJ3 looked at the computer very often, with and without DiskPlay. When asked why, he explained *"I often look to the display, no matter if I want to gain information from it or not. It's a habit"*. The most important aspect, the visualization of start and end of a track, was very well received. As DJ1 stated, *"the most embarrassing thing that can happen to a DJ is that the song is over without him noticing it and therefore has no time to create a smooth transition by beatmatching"*.

Participants received the system very well.

This happened to one of our participants during the accommodation phase of the test. To prevent such mistakes, DJ3 asked for a functionality that alerts him of the upcoming end of the track *"something flashing would be nice"*. Three of our participants suggested adding an absolute time display onto the turntable, and two asked for a BPM indicator. All participants liked the simplicity and intuitiveness of the augmentation. DJ4, who was more into scratching than the other participants, mentioned the cue point functionality as a good replacement for the labeling technique used by scratch DJs.

Taking into account the small number of participants, the results from the seek task only reveal a general trend. The time from lifting the tonearm until reaching the designated point in the track was, not surprisingly, longest when performed *By Ear* (M=19.3s, SD=9.4), even though the test was performed with records from the participant's personal collection. Using the *DVS* only was slightly faster (M=11s, SD=4.7) then with *DiskPlay* (M=13.3, SD=10.3). A repeated measures ANOVA on log-transformed task completion

time with user as random factor found no significant effect of CONDITION on task completion time ($F(2,24)=3.217$, $p=0.0578$). The slightly higher seek times for DiskPlay can be explained by the low resolution of the projection, which makes it difficult to hit the exact groove, compared to the high resolution of the various information displays on the computer screen, i.e., waveform, track overview, cue points, and precise timing information.

3.4 Integration Into a Commercial DVS

Although the development of the first implementation was easy thanks to the open Mixxx platform, we noticed two drawbacks of this implementation. First, the tracking algorithms used to decode the timecode in commercial DVS software work much better, especially at low rotational speed of the record. Second, the participants of our study had to adapt to a new user interface, which, although very similar to the commercial ones, might influence their perception of the system.

To avoid switching host software, we integrated DiskPlay into a commercial DVS.

We took these issues as starting point for our next iteration which should not only be an implementation of the existing features in a new host software, but extend the visualization with details that participants of our first study requested.

Waveform Display We added a semitransparent white waveform laying on the top layer and positioned at the half radial distance from the outside of the record to the label (Figure 3.8). The waveform graphically represents the audio content of one revolution of the timecode record and is presented using a circular conveyor belt metaphor. The visualized section is synchronized with the playback position which makes the waveform appear to stick to the record, meaning that a peak on the waveform corresponds to a peak in the audio signal when it passes the stylus. To have a maximum of the waveform visible on the record, it appears/disappears opposite of the stylus.

The second DiskPlay iteration visualizes the track waveform.

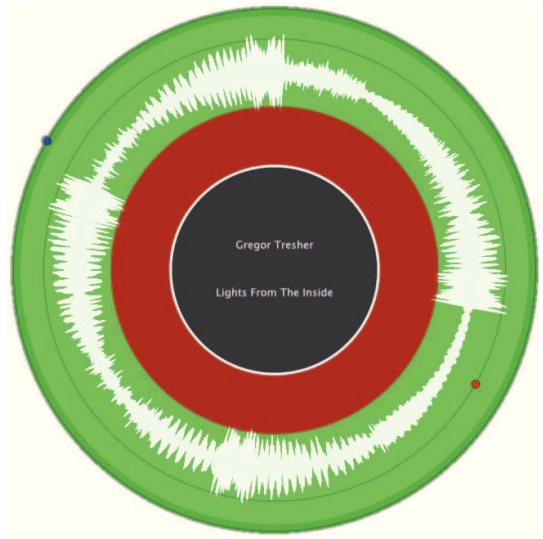


Figure 3.8: Visualization in the second iteration of DiskPlay. In addition to the features of the first version, we added a waveform display for precise navigation.

We integrated a visualization supporting navigation to a specific cue point.

Cue points As in the first iteration, the cue points are visualized as small dots, color coded as in the DVS software. Again, a black concentric circle, the orbit, serves as hint where to place the stylus to quickly navigate to that cue point. To compensate for the problem that the resolution of the projector is too low to render a line that matches a single groove on the record and thus it is hard to hit the cue point exactly, we added an animation that helps decide whether you have to rotate the record clock- or counter-clockwise to reach that cue point. When the playback is closer than 8 s to the cue point, a rectangle and a bar appear next to the stylus (Figure 3.9a). The horizontal bar indicates the time to the cue point in both directions and the direction in which to rotate. The bar length decreases when approaching the cue point (Figure 3.9b) and increases after passing the cue point (Figure 3.9d). When the cue point is hit, the

outline of the rectangle is thickened as visual feedback (see Figure 3.9c).

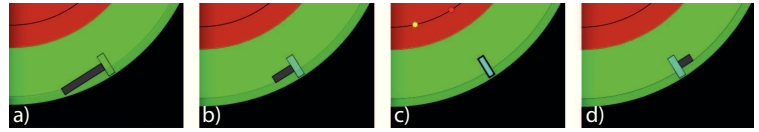


Figure 3.9: Detailed view of the visualization passing a cue point. The horizontal bar shows the time to the cue point (a) and the direction in which to rotate to reach it (b,d). The rectangle's outline is stroked when the cue point is reached (c). (Taken from [Burger, 2013])

End-of-Track warning As requested by some of the participants in the first study, the record flashes red when the track is closer than 30 s to the end, similar to the visualizations in the DVS.

3.4.1 Technical Setup

The goal of the second iteration was to integrate the DiskPlay visualization into a commercial DVS software to make use of the advanced tracking algorithms. Unfortunately, neither Traktor Scratch nor Scratch Live provide an SDK for such extensions. Traktor Scratch provides many MIDI input and output capabilities, but transforming them into information that we could use is tedious and error prone. Scratch Live can be extended with the Serato Video[10] plugin, initially designed to enable VJs to use turntables as controllers for their artistic performance. As such, it supports standard video file formats, with the speciality that the Mac version also supports Quartz Composer[11] patches. Quartz Composer (QC) is a visual programming language designed to quickly create animations such as iTunes visualizations or screen savers. These patches are very popular among the VJ community to create interactive visualizations, which is why Serato Video

We hooked into a VJing plugin to get the timing information.

[10] serato.com/video
[11] developer.apple.com

provides QC with detailed track and precise timing information. The visualization we wanted to create, however, was too complex to be realized with a QC patch, so we decided to stream the data to an external application using a UDP network connection. This separate application gathers all required information and renders the visualization on the disk. The information about the cue points, unfortunately, is not available through this path so we use on-screen OCR on the Serato Scratch Live interface to collect the according timestamps. A detailed description of the OCR algorithm can be found in [Burger, 2013].

3.4.2 Evaluation: Acceptance

We conducted an online survey to gauge the acceptance of a tool like DiskPlay in the DJ community. Especially in the slightly competitive scratch DJ community, the use of such a setup might be considered cheating. We compiled a package with the software and detailed explanations how to set up the system. We published the link to the survey and the software in a series of popular DJ forums and got 20 valid responses in total. The respondents had an average experience as DJ of 9 years (SD=6) and on average 4 years (SD=3) of experience with digital vinyl systems. Most used Scratch Live or Traktor Scratch, only one worked exclusively with a different software. Navigation within a track with timecode records was not perceived as being more complicated than with traditional vinyl (Mdn=3 on a 5-point Likert scale, IQR=2), but 75% of the respondents felt bothered by having to look back and forth between computer screen and turntable (Mdn=4, IQR=1, 5 being "strongly agree"). 70% of the respondents regularly used cue points to find certain positions in a track, and mostly 1-2 cue points are set per track.

Acceptance within the DJ community is high.

When asked if systems like this should not be used, the participants strongly disagreed (Mdn=1, IQR=2), which suggests a high acceptance. Most of the participants would feel comfortable to perform with a system like DiskPlay (Mdn=4, IQR=2, 90% approval). One of the major concerns that the respondents mentioned is the installation of one or

two projectors above the turntables. This is of course a limitation of the current prototype and the hardware would need to be ruggedized to be transportable to a nightclub, but for our research purposes it is a feasible solution. The system could also be part of the fixed installation in the DJ booth of a club, like the remaining equipment already is.

3.4.3 Evaluation: Mixing Task

To evaluate the use of our system in practice, we conducted mixing sessions with four professional DJs with two to 25 years of experience. We let the DJs perform a mixing task between different tracks with traditional vinyl, using Scratch Live, and with our system. We set up two Technics SL-1200 MK5 series turntables and a standard DJ mixer, along with Serato Scratch Live running on a laptop next to the turntables. The projector was mounted above the left turntable only. To ensure this would not affect observations, we asked the participants if they noticed any preference in the direction of mixing, but none reported such observations. The sessions were recorded by two cameras, one capturing the two turntables and the mixer, the other one was mounted above the computer screen to see where the DJ was looking.

All DJs took advantage of DiskPlay's waveform visualization. They used it to cue up beats and especially to find the start of the song. While initially, participants relied on the waveform display of the DVS software on the computer screen, this changed once the participants familiarized themselves with DiskPlay and started trusting the waveform as a high detail navigational tool. DJ3 and DJ4 additionally used it to pinpoint the exact beat when navigating to a cue point. Although one of the participants stated to not like to use visual aids, he started using the waveform actively after a short while.

In our lab study, no difference in attention to the laptop could be observed.

We could not verify our expectation that visualizing track information counters the "Serato face" problem. The number of attention switches between turntable and computer screen did not differ in the different conditions, however,

they vary greatly between participants. One DJ even stared at the computer screen when he was mixing with traditional vinyl, and the DVS did not show anything meaningful. The participants described this as a habit, which is consistent with previous evaluations [Heller, Borchers, 2012].

To achieve their sophisticated skills, DJs often restrict their equipment to a small set of hardware that they know really well. A new component probably needs some time to be fully integrated into the performance, which is why we suggest a long-term study to evaluate the changes in behavior.

We observed that two of the DJs used the headphones only for a last check or not at all, mixing purely with visual feedback from the DVS. Our system is very well suited for this kind of mixing, since finding a beat or a cue point is supported visually directly on the turntable.Overall, DiskPlay is a tool to provide an overview of the track structure, and as such, does not speed up the time to beatmatch two songs. To support this aspect, the visualization would need to have a BPM indicator, similar to the UI of the DVS, but this would add features that are not part of the traditional vinyl record.

3.4.4 Feedback

In both the online survey and after the mixing sessions, participants were asked for feedback about the system and how to improve it. The overall feedback was very positive to enthusiastic. During the survey we got responses such as *"everything...perfect idea and this would help DJs a lot"* or *"the idea is really top and thought through! thumbs up!"*. In the interview after the mixing sessions, three participants mentioned being bothered by the focus switches when using a DVS. One said *"It's about time that someone does something about this. It has been bothering me since I bought my DVS"*, and added that he liked the idea of having *"the visualization right where he is working"*. DJ1 caught himself looking at the computer screen even if he did not have to, which he commented with "Although I could easily use the wave-

form on the record, I am still looking at the waveform of Scratch Live". On the same issue, DJ2 mentioned "I think it's a habit that I continue to look at the screen".

3.5 Changing the Spectator Experience

One very interesting aspect was brought up by a participant of the first study: *"Most people in the audience don't know what the DJ is actually doing during his performance. It would be nice if the visualization could give the people an understanding of the DJ's job"*. Electronic music performances in general struggle with the fact that small changes on the tiny knobs of a filter bank can have a huge impact on the sound, but that the audience does not necessarily perceive a physical motion of the artist as correlated to a change in sound [Bell, 2010]. Especially the use of the laptop-turntable combination seems to be hard to understand for the audience, as the DJ could just be playing some prepared mix and working on something else. One of the participants mentioned that a spectator had once asked him if he was checking his email while playing a set. This lack of communication is crucial, because "watching the motions of the DJ during the performance can be almost as exciting as listening to the music being played" [Beamish, 2001]. As described in [Hook et al., 2011], the audience does not need to actually understand what is happening on stage, but they should be "seeing you on stage performing, and get a sense that something special is happening".

Electronic music performances are intransparent to the audience.

Spectator experience can be classified along two axes [Reeves et al., 2005] that describe how manipulation of a system in public and its effects both range from *hidden* to *amplified*. The activity in the DJ booth can be classified as *partially hidden*, as at least some of the manipulation is visible. With DiskPlay, this is leveraged to *visible* since it becomes easier to get a grip on what the DJ actually does. With an additional mirroring display, "DiskPlay could become an explicit part of the DJ's performance" and its visual appeal, as suggested during the interviews of our

DiskPlay makes the DJ's actions visible.

first study. Tools like the Emulator Elite[12] or the similar Waves VJ interface [Hook, Olivier, 2010] make the performance transparent by showing the artist's workbench to the crowd. With the same goal Rouages [Berthaut et al., 2013] abstracts the internals of electronic music setups and provides compelling visualizations. In contrast to the two former interfaces, this visualization does not represent the actual technology, but the visualizations can be simplified, abstract, or even artistic. Turntables are essentially music playback devices, so understanding their working principle is very easy. However, understanding what a DJ does with it is difficult, since a lot of the magic happens, unheard by the audience, in the headphones of the DJ. The Cubic Crossfader, a bluetooth enabled tangible control part of the ColorDex system [Villar et al., 2007], allows the DJ to move around and thereby get more engaged with the crowd. However, its gesture-based interaction design does not support the visibility of the DJs work. A visualization like the one presented in this chapter, again, makes the performance more transparent and might support the interaction between crowd and artist.

3.6 Future Directions

With the integration of sample players in the DVS software (called Remix Decks in Traktor or SP-6 Sample Player in Scratch Live), which can be triggered by standard MIDI controllers, the barriers between the performance as a DJ and as an electronic music live act blur. A controller like the Novation Dicer[13], today, is a common add-on to the traditional setup of two turntables and a mixer. The Reloop RP-8000[14] turntable even has the controller already integrated. This evolution represent a shift, moving the turntable from a pure playback device more towards being a controller. With the adoption of new technology, we imagine the display becoming an integral part of the turntable, making the additional top-projector obsolete and having a single, ro-

We could integrate the display directly into the turntable.

[12] smithsonmartin.com
[13] novationmusic.de
[14] reloop.com

Figure 3.10: Left: The Novation Dicer MIDI controller for DVS (Image curtesy of Focusrite). b) The Reloop RP-8000 turntable with integrated buttons to control the host DVS software (Image by Reloop)

bust piece of hardware. This would also simplify the setup, which participants of our second study feared to be complicated and, therefore, would prefer the system to be preinstalled in a club.

The visualization we presented in this paper can also be transferred to CD players or DJ controllers. These MIDI controllers, potentially with motorized platters like Numark's NS7II[15], essentially have the same problem of separating visualization and control. Augmenting these with an additional display in the jog wheels (similar to Pioneer's CDJs[16]) would make these even more powerful. The handling would still be different than a real turntable, but could fit the personal preference of some DJs.

[15] numark.com
[16] pioneerdj.com

3.7 Conclusions

We visualized cues that were present on traditional records, but were lost during digitalization.

We increased the visual interaction bandwidth to have haptic input and visual output co-located in a single device.

To study the extension of the visual output bandwidth, we built upon a system which already has a high haptic input bandwidth: the DJ turntable. Digital vinyl systems took the effort to lift the analog turntable into the digital age because of its unreached haptic input capabilities. While having the advantage of running on an unmodified turntable, these systems reduce the visual output bandwidth of the medium itself. Visual cues present on traditional vinyl records that helped navigate within a song were lost, as timecode records have become generic controllers for nearly random MP3 files. The software interface on screen shows all information visible on the traditional record and even more, but the visualization is spatially separated from the haptic interaction. This means that the DJ has to split attention between the turntable and the computer. To create an embodied interface, we increased the visual interaction bandwidth of the combined turntable and DVS record, to show the required information at the actual location of attention and interaction.

We published the software online along with a questionnaire regarding the acceptance of such a system within the DJ community and conducted two lab-studies with professional DJs. While we achieved our initial goal to merge haptic manipulation and visual output back into one device, we could not observe a significant decrease of glances at the computer screen. This is probably due to our lab setup in a quiet room where there was not much to see, which is quite different from the environment of a DJ booth. The overall feedback of our test DJs is very positive and in the same line with the results of our acceptance study. From feedback and observations during our studies we discovered that such an additional visualization potentially improves the spectator experience by reaching a higher transparency of what the DJ actually does behind the decks.

After having covered the visual and haptic modalities, the next chapter will handle the auditory channel and how we can use modern rendering technology to create environments with variable spatial arrangement of sound sources.

Chapter 4

Audio Augmented Environments

4.1 Introduction

The last of the three modalities we consider in this thesis is the auditory one. While we increased the haptic and visual bandwidth to enhance the interaction with audio, in this chapter we will rethink the presentation of audio to leverage our spatial hearing capabilities. We increase the information bandwidth in the audio output stream perceived by the user by adding spatial components to the recorded audio sources. At the same time, we increase the haptic input bandwidth to be able to control the playback parameters of this augmented auditory display in a natural way.

In everyday life, we are surrounded by rich soundscapes

Publications: The work in this chapter was published as Work-in-Progress at CHI'09 [Heller et al., 2009] and at MobileHCI'14 [Heller, Borchers, 2014a], as full-paper at CHI'14 [Heller et al., 2014b] and as Note at CHI'15 [Heller, Borchers, 2015] and MobileHCI'16 [Heller et al., 2016]. For all these publications the author of this thesis was the main author and performed the data analysis. Thomas Knott [Knott, 2009], Aaron Krämer [Krämer, 2014], and Jayan Jevanesan [Jevanesan, 2015] worked on this project as part of their Diploma or Bachelor theses under the supervision of the author of this thesis.

with sounds coming from different sources at different positions. Our ears and auditory perception use the spatial nature of sound to help us locate these sources and to discriminate between them, for example to focus on a conversation in a noisy environment (known as the cocktail party effect) [Arons, 1992]. Compared to this rich auditory environment, our everyday listening experience of recorded music is a purely passive activity. In common stereo recordings the position of the sound sources is encoded by different volume levels for the left and the right ear. When listening to these recordings via headphones, the sources are perceived as to be in the head, and not around them, which is called a lack of externalization. No sensor input is used to make the audio output react to actions of the listener. Using special recording equipment such as a dummy head which mimics the human body and its characteristics influencing sound perception, we can record audio in a way that during playback, the sources are perceived at their relative position to the dummy head. This so called binaural recording creates the impression that the listener is sitting at the same position as the dummy head and provides a static image of the spatial layout of the sources. Even though this technique only uses two separate audio channels, the placement of the microphones in the modeled ears of the dummy head results in the encoding of the spatial impression into the recording [Vorländer, 2007]. The parameters of the listener's head and torso that create this spatial impression are summarized in the head related transfer function (HRTF), which is individual to every human. If we apply this individual HRTF to a piece of recorded audio, we create the impression that the sound emanates from a location in the physical space. While some years ago, applying an HRTF required specialized DSPs, the processing power of today's smartphones is sufficient to process several sources in parallel [Sander et al., 2012]. In combination with motion tracking of the listener's head and body, this technology is used to create audio augmented reality experiences in which virtual sources are perceived to be at fixed points in the physical space. We use this technology to create an engaging experience and take advantage of our natural spatial perception of sound.

When listening to common stereo recordings, the relative position of the instruments is fixed.

When listening through headphones, HRTFs can create the impression of sound emanating from a source in the physical space.

4.2 Foundations of Spatial Hearing

Before we dive into the interaction with audio augmented reality systems, we will first review the foundations of spatial hearing and the methods available to create a spatial impression of a prerecorded sound. While the basic cues for sound localization in the horizontal plane have been discovered over a hundred years ago [Rayleigh O.M. Pres. R.S., 1907], the details of a perfect spatialization have only been discovered around 1982 [Shaw, 1982].

On the way from their source to the ears of the listener, the sound waves are subject to reflection and diffraction, not only by the space they are moving in, but also on the head and torso of the listener. In this thesis, we will restrict ourselves to the factors related to the listener, and leave the influence of the room aside, although room acoustics are also part of the human sound localization process [Kohlrausch et al., 2013]. The distortion of a sound wave on the body depends on the direction the sound comes from as it reaches the ears at different times and amplitudes. For example, if a sound comes from the left, the waves will reach the left ear before the right ear (interaural time difference (ITD)) since they have to travel a longer path, and they will be dampened resulting in a smaller amplitude (interaural level difference (ILD)) and shifted frequency distribution on the right ear.

Several physical factors account for our spatial hearing.

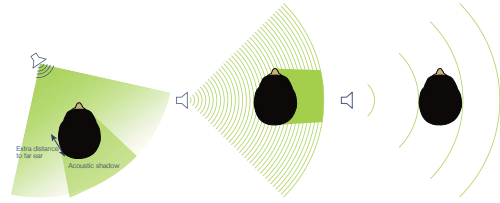


Figure 4.1: Interaural Time and Level Difference.

Localization in the horizontal plane is mostly performed with these two cues, with the level difference being the more important one since time differences can result in ambiguities in phase shifting for frequencies of approximately

The interaural level difference is the main cue for horizontal localization.

1.3 kHz and higher [Middlebrooks, Green, 1991]. The simplest form of spatial audio rendering is thus plain stereo panning, which is already sufficient to give a sense of direction to a virtual sound source. For the localization of sounds in the median plane other features have to be taken into account as changes in elevation of a source do not result in significant changes of phase or amplitude. These monoaural features are mostly distortions created by the pinna (the outer ear) and are thus not easily replicable since this information is individual to every human.

4.2.1 Head Related Transfer Functions

The head-related transfer function (HRTF) is a description of monoaural and binaural cues encoded in the audio signal when it reaches the eardrum. The technical background on HRTFs is explained in detail in, e.g., [Blauert, 1996; Vorländer, 2007]. Basically, an HRTF database consists of a series of recordings of sound arriving at the eardrum from different defined positions. Usually, a loudspeaker is fixed on a boom and moved around the head in discrete steps (e.g., 1°, 5°, 15°) playing small bursts of noise-like signals. The microphones are either placed at the position of the eardrums in a dummy head for a generalized HRTF [Vorländer, 2007], or small microphones are placed into the ear canal of a human listener to record her individual HRTF [Wightman, Kistler, 1989]. Eventually, the HRTF database contains the recorded distortion of a sound wave for a number of azimuth and possibly elevation angles. Applying this same distortion to a prerecorded single-channel audio signal and presenting it over headphones results in the sound being perceived as coming from the direction the HRTF was recorded for. In contrast to simple stereo panning, this technology has the advantage to result in a better externalization, i.e., the sound is perceived to come from a location outside the head.

Localization accuracy with one's individual HRTF is as good as in real spatial hearing.

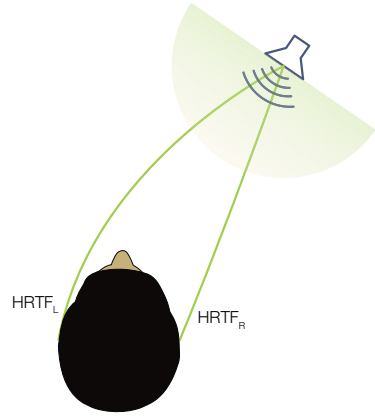


Figure 4.2: The head-related transfer function (HRTF) is a description of monoaural and binaural cues encoded in the audio signal when it reaches the eardrum.

Localization Accuracy

The localization performance with an individual HRTF is as good as in real world spatial hearing [Wenzel et al., 1993], but it only considers listener-related factors. Localizing sounds in an anechoic room feels as unnatural as a simulation without room acoustics which should be included for a perfect experience. The just noticeable difference between two signals is smallest in frontal direction. As an approximation, human sound localization accuracy is in the order of 1°for sources directly in front, around 10°for signals coming from left or right, and 5°for sources in the back. [Vorländer, 2007].

Measuring individual HRTFs is tedious, that's why we commonly use generalized HRTFs.

If a visual stimuli is present along with the auditory one, our perception tends to fuse these and the perceived location of the sound source snaps to the visual stimuli, the so called ventriloquist effect. Under optimal rendering conditions, this is true for a localization blur of about 10°, but can be present for up to 60°offset with very simple rendering algorithms [Alais, Burr, 2004].

4.2.2 Audio Augmented Reality

Audio augmented reality (AAR), similar to its more prominent visual counterpart [Mackay, 1998], lets the user experience a tightly coupled integration of virtual elements embedded in the real world. AAR uses spatial audio rendering and combines it with tracking of the user's motion to create the impression that virtual sound sources are at a fixed position in the physical space. Applications of this technology include auditory navigation systems [Holland et al., 2002; Stahl, 2007], interactive experiences [Terrenghi, Zimmermann, 2004; Vazquez-Alvarez et al., 2012; Heller, 2014], auditory menus [Marentakis, Brewster, 2006; Kajastila, Lokki, 2010], and games [Paterson et al., 2010].

Spatial audio rendering has been evaluated as navigation system.

The different applications assign different priorities to certain aspects of the simulation. For an outdoor navigation tool, realism is not a primary concern, instead it is sufficient to give the user a sense of orientation. AudioGPS [Holland et al., 2002], for example, only uses simple stereo panning and a harpsichord timbre for sources in front of the user, and a trombone timbre for sources in the back. Distance is mapped using a Geiger-counter metaphor, with only few sounds emitted when the user is far from a waypoint and many, more frequent sounds emitted when the user is close to a waypoint. The transition to interactive experiences is quite fluent, as a navigation system can create a more engaging soundscape to give the user a sense of place. The Roaring Navigator [Stahl, 2007], an AAR guide for a zoo, still uses stereo panning for orientation and source volume as distance cue, but instead of an unrelated beacon sound, it uses sounds of the animals in their respective enclosure. A Sound Garden [Vazquez-Alvarez et al., 2012] is a more exploratory experience, where the points of interest not only have a beacon sound attached, but also provide some additional information when the listener comes close. While successful navigation is possible with simple algorithms and coarse tracking [Mariette, 2010], simple algorithms require a considerable angular distance between two possible sound sources to be matched correctly [Heller, Borchers, 2015]. Similar to the sound garden, audio augmented reality can be used to create an immersive experi-

ence in museums. However, if we want to attach the virtual sound sources to smaller physical objects, like paintings, both sensing and rendering have to provide a much higher accuracy to be able to differentiate between the sources reliably. First, we need a rendering with a higher horizontal resolution, and since better algorithms are subject to larger interference through tracking latency [Mariette, 2010], we also need a fast orientation and position tracking. With the according technology, AAR can make paintings tell the visitor their own story [Terrenghi, Zimmermann, 2004].

Spatial audio rendering can be used to create engaging experiences.

Despite its unique features, audio augmented reality is still far away from broad dissemination, which is mostly due to the required sensors. To create a realistic experience, sensing and rendering should of course be state of the art, which is feasible in a lab setting using optical tracking systems and complex rendering algorithms. If we want to make AAR accessible to a broader public, we have to rely on sensors and algorithms that are easily available. The question that we want to answer in this chapter is, if this reduction in tracking and rendering quality actually affects the users' experience, or if most of it remains unnoticed.

The hardware required for a spatial audio experience hinder a broader dissemination of the technology.

In the remainder of this chapter, we will summarize related work on audio augmented reality and then focus on sensing and rendering parameters and their influence on the interaction with virtual audio spaces.

4.3 Related Work

Interacting with virtual audio spaces is actually quite complex as it is influenced by many sources of imprecision which might result in a suboptimal experience.

Loomis [Loomis et al., 1990] and Mariette [Mariette, 2010] analyzed the paths and head orientation of people walking towards virtual sound sources. Results show that in the case of a large space, users do an initial, large head turn to get an estimate of the direction, followed by smaller movements to stay on the path towards the source.

Speech sounds are harder to localize than non-speech beacons.

The type of beacon sound has an impact on the localization performance in AAR. Speech sounds are harder to localize than non-speech sounds [Walker, Lindsay, 2006], which is unfortunate for applications such as museums. While most lab-experiments are conducted with white noise as beacons sound, this is not feasible for a real-world application. Alternatives like a single pure tone or a sonar ping sound seem more appropriate and experiments show that these are preferred over speech beacons [Tran et al., 2000]. The ping sound has the disadvantage that in only allows localization in bursts, which results in a longer task completion time.

It is possible to interact with the virtual audio space using a pointing device.

Marentakis et al. [2006] evaluated pointing as an interaction technique in virtual audio spaces. While walking, participants experienced a sound coming from somewhere around them and had to point to that source. Whenever the heading of the pointing gesture was within a certain angle from the actual position of the source, a feedback sound was presented to facilitate the task. Results show that this technique is feasible to interact with virtual audio spaces, e.g., auditory menus where the items are arranged spatially around the head [Kajastila, Lokki, 2010]. While the interaction proposed in section 4.6 is similar, we do not focus on selecting a certain sound source, but want to create an auditory image of the source positions for navigation.

Corona re-creates a medieval coronation ceremony.

Our inspiration: the Corona audio space Our test case is an audio augmented reality experience deployed in the Coronation Hall (Figure 4.3) of the historic city hall in Aachen, Germany. This room was the location of coronation feasts for important emperors in medieval Europe, including Charlemagne. Of these festivities no apparent visual traces remain except for a series of coats of arms engraved in the pavement. To bring back the atmosphere of such festivities, we created an audio space that depicts the well-documented coronation feast of Charles V. from 1520. Virtual characters discuss different aspects of the ceremony, providing the visitor with insights in a more personal manner: Maids discuss the order of the dishes, characters at the window describe the festivities for the common people they are watching, and clerics and the king discuss the

Figure 4.3: The historic Coronation Hall where the Corona virtual audio space is deployed (Taken from [Heller et al., 2014b]).

perils of the Black Death. Since the sound sources are not connected to concrete physical objects but to meaningful locations, we consider this an augmented environment. The CORONA audio space combines the atmosphere of a medieval coronation feast with educational content into an experience of serendipitous discovery.

4.4 Orientation Measurement in Audio Augmented Reality

In this first section we will analyze how users orient towards sound sources with natural spatial hearing and with headphones using a spatial audio rendering. Since it is possible to render high quality spatial audio on a modern smartphone [Sander et al., 2012], our focus lies more on reducing the sensor hardware complexity. The goal of this first experiment was to determine if there is a significant difference between the orientation of the head, body, and device when moving towards a sound source. This should give us an indicator where to place the compass

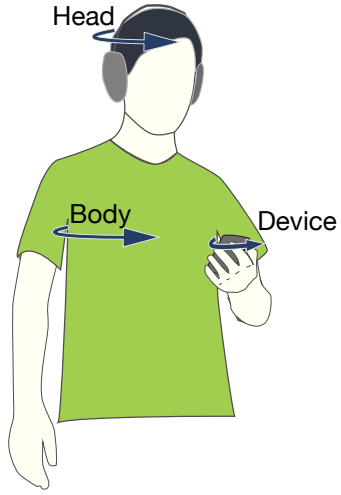


Figure 4.4: The three different reference systems for orientation measurement we compared in this experiment (Adapted from [Heller et al., 2014b]).

Can we use the smartphone compass to measure orientation?

when designing a AAR application. Since nearly every current smartphone has an integrated digital compass, using device orientation is very easy. If the display is not used, measuring body orientation can be achieved by attaching the smartphone such that it rotates with the torso (e.g., through a lanyard), whereas using head orientation requires additional hardware like, for example, the Intelligent Headset[1].

Technical Setup

To compare the behavior of real spatial hearing with the orientation in a virtual audio space, we created the following experimental setup. We placed 24 Wavemaster Mobi loudspeakers at 15° intervals in a circle with 4 m diameter. The loudspeakers are designed to stand upwards and emit

[1] intelligentheadset.com

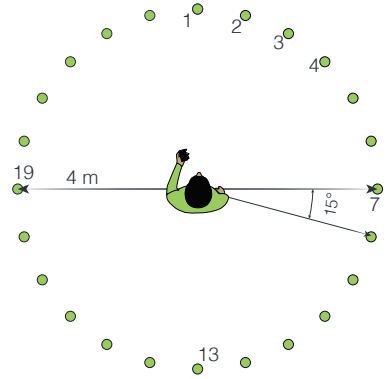


Figure 4.5: The experimental setup with 24 sound sources placed at 15°intervals (Taken from [Heller et al., 2014b]).

sound in an omnidirectional pattern, which is how sound sources are modeled in our virtual audio space. We placed the loudspeakers at roughly head height (140 cm) above ground to reduce the impact of elevation angle on the localization. The audio output of the smartphone was connected to the loudspeakers via a cable hanging from the ceiling. The cable was attached to the user's waist to avoid pulling forces on the device and to keep users from stumbling over it.

To quantify the influence of smartphone-based spatial audio rendering, we created a virtual audio space that represented this same setup. Positional tracking was performed with a Vicon optical tracking system with an update rate of 100 Hz. Spatial rendering was done on an Apple iPhone 4S running iOS 5.1.1 using the OpenAL library and presented using AKG K-512 headphones. The headphones fit firmly and have a supple cable so as to reduce the impact on the amount of head turning. Since we needed the optical tracking markers for the head in both conditions, participants had to wear a headband during the loudspeaker trials, which balances this influencing factor. While state of the art spatial audio rendering technology achieves aston-

We used a smartphone as rendering platform.

ishing results[2], the auralization results of this framework are less realistic. We decided to use this one, as it is a representative for a variety of spatial audio rendering frameworks available for mobile phones, comparable to, e.g., the AM3D Framework[3] used in [Marentakis, Brewster, 2006] or the Java Advanced Multimedia Supplements used by Vazquez-Alvarez et al. [2012]. We used the OpenAL ex-

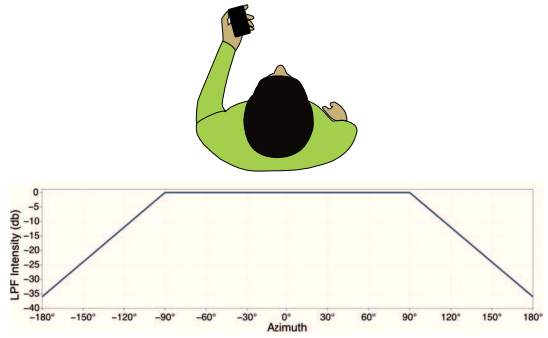


Figure 4.6: Our improvement of the available spatial audio rendering. To reduce front/back confusion, a low pass filter is applied to sources in the back of the listener. The intensity of the filter is interpolated linearly from 0 dB at 90° to -36 dB at 180° (Taken from [Heller et al., 2014b]).

We applied a low-pass filter to create a muffled impression for sources in the back of the listener.

tension `ALC_EXT_MAC_OSX` which provides a better spatialization based on a spherical head model and including the following filter factors: interaural level difference, interaural time difference, head filtering, and frequency dependent distance filtering. To improve the perception of sources that are behind the head we used the `ALC_EXT_ASA` extension that enables additional effects, such as reverb, obstruction and occlusion. As the rendering suffers from front-back confusion, which is a common problem in spatial audio rendering [Middlebrooks, Green, 1991], we added a low-pass filter that muffles the sounds that are behind the listener.

The low-pass filter intensity is interpolated linearly be-

[2] Virtual Barber Shop: `http://youtu.be/IUDTlvagjJA`
[3] www.am3d.com

tween 0 dB and 36 dB for sources with an azimuth angle between 90° and 180° (Figure 4.6). For the reverb, we used the medium room preset which best matched our impression of the physical room's characteristics. We also tuned the audio rendering parameters to make the scene sound as similar as possible in both conditions.

No delay of location or orientation measurement was perceived. With a specified latency of 2.5 ms of the Vicon Tracker and an average round trip time for the WiFi connection of 4.7 ms, we are below the limit of 376 ms total system latency defined in [Mariette, 2010] and the 80 ms head tracker latency defined in [Brungart et al., 2005].

4.4.1 Conditions & Methodology

Since related research indicates that there are performance differences in the localization of different source sound types [Tran et al., 2000; Walker, Lindsay, 2006], we decided to use a non-speech sound and a speech sample. As a non-speech sound, we chose a drum sample that covers a large frequency range (Figure 4.7) and is repeated every second. The repetition rate was chosen based on the recommendations in [Tran et al., 2000]. We favored the drum sound over artificial sounds, e.g., noise or a square waveform, because it fits the mental model of a sound emerging from a precise single location. The speech sample is a continuous monologue of a male voice, which is close to our use case. Together with the two presentation modalities headphone and speaker, we have four conditions that were balanced across all participants. Our 24 participants, 3 female, age 19-53 (average 26), mostly had no prior experience with spatial audio and did not report any known hearing problems. The position of the sound source was randomized, such that each participant had to navigate to each of the 24 sources under every condition.

We used a drum sound and a speech sample as beacon sounds.

We recorded the position of the head as well as head, body, and device orientation. Due to a technical issue, headphone measurements for source no. 7 were discarded, leaving 23 sources for the evaluation. We did not measure task com-

We logged head position and head, body, and device orientation.

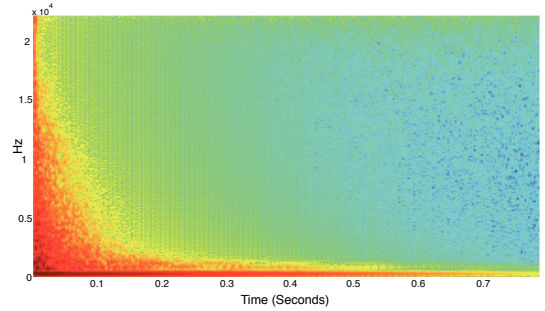


Figure 4.7: The frequency spectrum of the non-speech beacon sound. We used a drum sample with a broad frequency range as it fits the mental model of a sound that emanates from one specific location (Taken from [Heller et al., 2014b]).

pletion time, since this is highly dependent on the source position (you have to turn around to reach a source in your back), and it is dependent on the type of beacon sound as a pulsed signal like the drum only allows localization in bursts in contrast to a continuous signal.

To be close to our designated use-case in the CORONA audio space, which might be similar to implementations that use the display to provide additional information, participants had to hold a smartphone in their hand. Participants were instructed to start the task using a button on the smartphone, go to the sound source currently playing until it was directly in front of them, and end the task using the stop button on the device. Participants practiced all conditions in a 12 trial training session before the actual experiment.

4.4.2 Results

By looking at the recorded paths of the participants, we can already see that the rendering has an effect, as the paths in the speaker condition are much smoother and lead to-

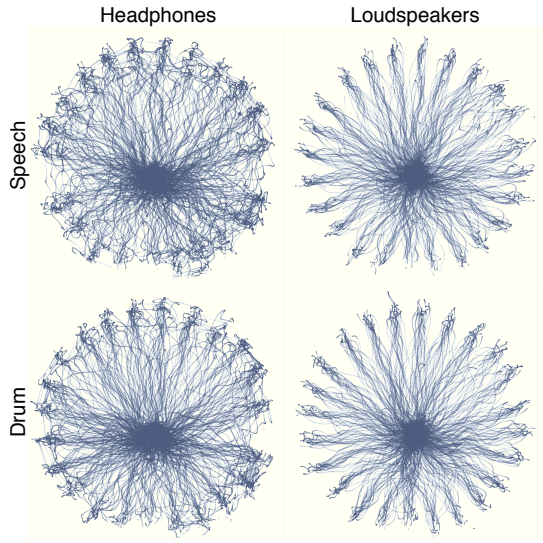


Figure 4.8: Paths on the way from the start in the center to the sources on the periphery of the circle. We see that the paths in the speaker condition are headed more directly to the source than with virtual audio rendering (Taken from [Heller et al., 2014b]).

wards the target more directly (Figure 4.8). From the three orientation measurements head, body, and device, we calculated their relative angles. Following the definition in [Mariette, 2010], we define head-yaw (θ_h) as the relative angle of the head to the body, device-yaw (θ_d) as the relative angle between device and body, and head-device-yaw (θ_{hd}) as the relative angle between head and device. We transformed the values from their reported range of 0° to 360° to $[-180, ..., 180]^\circ$, with 0° being the direction of the user's torso. We subtracted the initial difference between head, body, and device measurement at the beginning of each trial, since this difference is caused by the placement of the tracking markers.

Most of the time, body and head are aligned, as the means

for θ_H are close to 0 for both conditions (Headphones: M = -1.57°, SD = 15.83, Speaker: M = -2.24°, SD = 19.98). The kurtosis[4] for the headphone condition is a bit higher (Headphones: Kurtosis = 3.08, Loudspeaker: Kurtosis = 2.21), which indicates that the participants turned their head less in the headphone condition. Although the headphone we used has a comparably long and flexible cable, we cannot totally exclude this as an influencing factor on the amount and range of head rotations. Overall, body and device are aligned most of the time, since we have $M = -0.18°, SD = 8.62°$, Kurtosis = 85.7 with headphones and $M = -0.35°, SD = 13.05°$, Kurtosis = 5.92 for loudspeakers.

Most of the time, body, head, and device are aligned.

Since positive and negative angles cancel each other out when calculating the arithmetic mean, we calculated the root mean square (RMS) head-yaw and device-yaw deviation ($\theta_{h(\mathrm{RMS})}$ and $\theta_{d(\mathrm{RMS})}$), which gives us the average amount of head and device turns. After performing a log-transform on the RMS data, a repeated measures ANOVA revealed a major effect of the used rendering (headphones or speaker) on $\theta_{h(\mathrm{RMS})}$ ($F(1, 2083) = 132.76, p < .0001$). However, if we take a closer look at the data, we see that the RMS means only differ by about 4° (Headphones: $M_{\mathrm{RMS}} = 13.86°, SD = 8.05$, Loudspeakers: $M_{\mathrm{RMS}} = 17.75°, SD = 9.93$), which places it in the order of the just noticeable difference of the rendering. This slight difference is also noticeable in the head-yaw (θ_h) distribution. The RMS means for the angle between head and device ($\theta_{hd(\mathrm{RMS})}$) are in the same range as for those between head and body (Headphones: $M_{\mathrm{RMS}} = 15.06°, SD = 9.19$, Loudspeakers: $M_{\mathrm{RMS}} = 19.73°, SD = 11.84$), which shows that body and device orientation can be considered equal in this setting.

Body and device orientation can be assumed equal.

Not surprisingly, the source position also has a major effect on $\theta_{h(\mathrm{RMS})}$ ($F(23, 2083) = 22.63, p < .0001$). When orienting towards a source in your back, the amount of head turns will of course be larger. If we look at the values for the individual sources however, we cannot attribute this effect to sources in a specific location.

[4] *Kurtosis* is a statistical measure that describes the distribution of data around the mean. A positive or high kurtosis characterizes a sharp, peaked distribution.

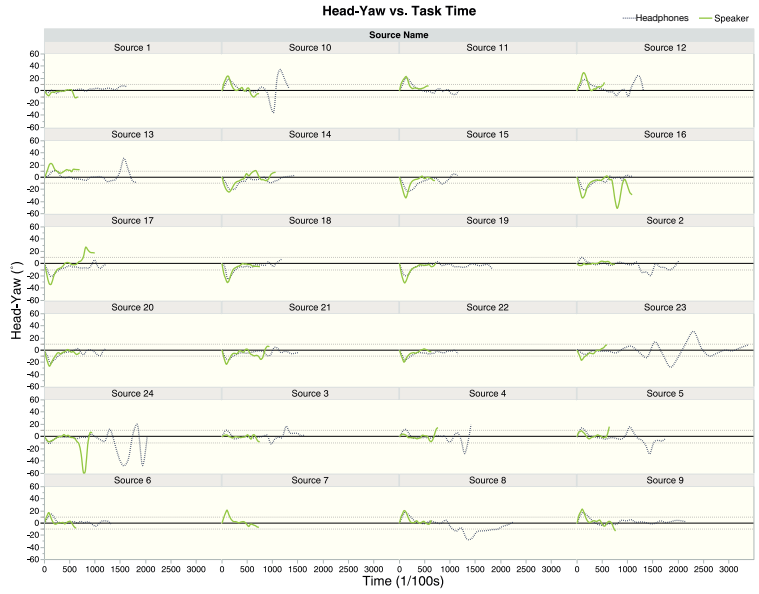


Figure 4.9: Mean head-yaw per source. Sources are numbered from 1 to 24 clockwise, starting at 12 o'clock. The large deviations at the end come from two users that took exceptionally long and turned their head extensively at the end of the task (Adapted from [Heller et al., 2014b]).

4.4.3 Discussion

The head-yaw tracks over time look very similar for both rendering conditions (Figure 4.9). The high fluctuations at the end of the tasks are caused by the fact that two participants took exceptionally long and turned their head extensively to discriminate between two possible candidates. The overall observation is that after a larger initial head-turn to get an orientation, the head-yaw stays within a 10°angle to both sides. If we just look at $\theta_{h(\mathrm{RMS})}$, the mean value of 14°is not extensively large, taking into account that our rendering has a just noticeable difference of about 4°. Similar to Fitts' law tasks, we have a large movement at the

Using device orientation minimizes the large initial head-turn, resulting in a less natural interaction.

beginning which is then slowed down to achieve a precise homing. By using body or device orientation, we risk loosing the large head-turns we see at the beginning of each trial. This might lead to a seriously degraded sense of presence in the virtual space as these rotations are necessary to get the initial orientation. The mean duration of the peaks exceeding 15°occurring in the first 4 s of each trial (cf. Figure 4.10) is 590 ms in the headphone conditions (SD = 750). This could be considered as an additional head-tracker latency, as the body follows the head with some delay. These 600 ms are too large to stay unnoticed by the user, but completing navigational tasks is still possible [Mariette, 2010]. Depending on the rendering resolution, technical setup, and designated use case this might be tolerable.

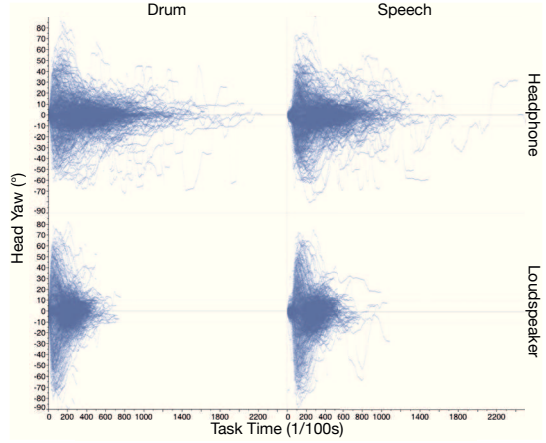


Figure 4.10: Head-yaw over task time of all users in the four different conditions. The initial head turn is present in both conditions (Taken from [Heller et al., 2014b]).

Do users notice which compass is used?

In our setting the device did not show any relevant information, but was used to start and end the trial. Nevertheless, all users held it in their hand in front of the body. If the smartphone is used to display some additional information, holding it in this position is further encouraged, which means that body and device will be aligned most of the time. Using the device orientation could also allow for

different interactions with the audio space, such as using it as a virtual directional microphone (see section 4.6).

4.4.4 Orientation Measurement and Perceived Presence

Our first experiment showed that to locate and move towards sound sources in the near field, the initial head turn is a natural behavior, which is not severely influenced by our rendering algorithm. This initial head turn, however, makes it difficult to use other sources than head tracking for orientation measurement. The tracking speed and accuracy necessary to measure and analyze this behavior is only achievable in a lab setting. As many of the existing implementations are deployed in much larger areas [Stahl, 2007; Vazquez-Alvarez et al., 2012; Lyons et al., 2000; McGookin, Priego, 2009], the question remains if the use of a different device orientation in a larger setting has an influence on the perceived presence and navigation performance. As indicated in [Vazquez-Alvarez et al., 2012], users might move slower, enjoy the experience, and pay less attention to the realism of the installation. To draw the right conclusions for practical installations from the results of the first study, we conducted a second experiment in a real-world setting.

We evaluated the impact of the compass position used.

Optical tracking is unfeasible for such a scenario, as it requires some kind of marker to be placed on the headphones and a considerable amount of cameras to cover large areas. GPS and magnetometers are less precise and may introduce higher latency and larger error to the measurements fed into the rendering algorithm, which might have an influence on the perceived presence.

To account for these different types of installations, we conducted a second experiment using sensors appropriate for larger implementations.

Technical Setup

Since optical tracking is not feasible for large areas such as the Coronation Hall (45×20 m), we used a Ubisense RTLS[5] with an accuracy of 15 cm in the center of the covered area, 50 cm at the outer borders, and a refresh rate of approximately 10 Hz. The location measurement has a specified latency of 234 ms and the WiFi connection used to transmit the location data to the smartphone adds an additional average latency of 42 ms, which results in a total location update latency of around 276 ms. The orientation measurement was performed using a tilt-compensated compass HMC6343 with a refresh rate of 10 Hz. Audio rendering was performed with the same engine as in the previous experiment.

4.4.5 Conditions & Methodology

We varied the placement of the compass, which was attached either to the middle of the headstrap of the headphones, to the left shoulder, or to the smartphone. To create comparable measurements, we used the same chip for all three measurements even though the smartphone had a built-in compass. Participants were not told which sensor placement was actively used in the respective trial.

We let participants navigate through a virtual audio space to different sound sources.

We created a series of six audio samples enumerating specific classes of objects, i.e., colors, first names, drinks, fruits, animals and cities, using a text-to-speech system. The participants were instructed to walk to the source shown on the smartphone display using text and an image. Upon successful arrival at a source, i.e., entering the capture radius [Mariette, 2010; Walker, Lindsay, 2006] of 2 m, a short sound sample notified the user. We created three distinct paths that were randomly assigned to the conditions. We randomized the sound sample played at a specific source position on that path and balanced the order of the compass placement across participants. The measurements recorded during the experiment include the path taken, the orien-

[5] www.ubisense.net

tation of the compass, and the orientation of the smartphone compass. After each trial, we asked the participants to fill out the presence questionnaire proposed in [Witmer, Singer, 1998], omitting the questions only related to vision or touch (Figure 4.11). To avoid that participants start to pay attention to specific details asked for in the questionnaire after the first trial, and thus making the results incomparable, they had to read over it before the first trial. Participants had to walk around through the audio space to get acquainted to it before the first trial.

4.4.6 Results

We collected data from 9 users, 2 female, age 20-25 (average 24), who all successfully completed the tasks. All questions were answered on a 7 point Likert scale, with 1 being the lowest and 7 the highest score. An analysis of the questionnaires showed no substantial difference between the three different compass placements. Head tracking receives the best overall scores (Mdn = 5.3, SD = 0.8), but the difference to device tracking is very small (M = 4.9, SD = 0.5) (cf. Figure 4.11). Since the perceived presence questionnaire [Witmer, Singer, 1998] is quite long we will only report the most interesting results. For the question *How natural did your interaction with Corona seem?*, *head* and *body* compass got the best results with a median score of 6 (*head*: IQR = 1.75, *device*: IQR = 1.0), followed by *body* tracking (Mdn = 5.0, IQR = 2). A pairwise Tukey-HSD test showed no significant difference with the smallest $p = 0.27$. The responsiveness of the environment was rated on a similarly high level: *head*: Mdn = 6, IQR = 0.75; *device*: Mdn = 6, IQR = 1; *body*: Mdn = 5, IQR = 1.5. The stability of the sources in space was perceived better in the *head* (Mdn = 6, IQR = 2) and *device* (Mdn = 6, IQR = 1.5) conditions than with *body* orientation (Mdn = 5, IQR = 3). For this question, the difference between *head* and *body* tracking is marginally significant with $p = 0.06$. The participants adjusted quickly to the virtual environment experience, again, with a slight but not significant advantage for *head* tracking. Some users mentioned that, although they were able to complete the task, they felt confused by the *body* tracking. The ratings indicate that the

No substantial impact of compass placement on perceived presence was found.

perception of the virtual audio space is not heavily affected by the different orientation measurements. This supports our hypothesis that head tracking is best, but device tracking sufficient for certain applications.

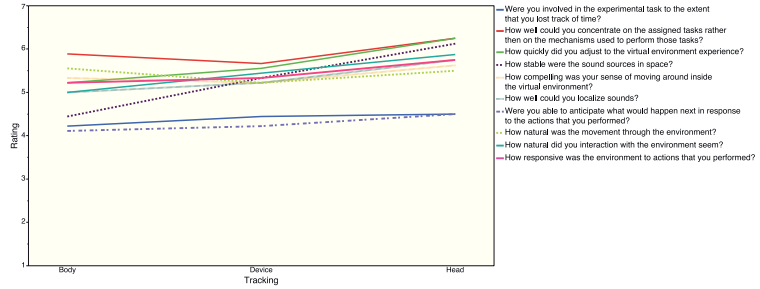


Figure 4.11: Mean ratings for the different compass placements on the presence questionnaire. Overall ratings are fairly high with a slight, although not significant advantage for the head tracking (Taken from [Heller et al., 2014b]).

From the log files, we calculated the relative angle between head and device, which we know from the first experiment to be a good approximation for body orientation. Since the hardware changed, the results are not directly comparable to those gained in the first experiment. As the compass chip uses accelerometer data to compensate tilt, the stability of the reading is reduced while walking. The average θ_H (M = -8.4°, SD = 33.0) and $\theta_{H(\mathrm{RMS})} = 34.1°$ are around double the results from the first experiment, which can be partly explained by the high fluctuation while walking. Future developments should take different filtering approaches into consideration and measure their influence on the overall latency. The task completion times for the three conditions showed a distribution similar to the ratings from the questionnaire. *Head* tracking was fastest with M = 192 s, SD = 63, followed by *device* tracking with M = 198 s, SD = 62, whereas *body* tracking was considerably slower with M = 245 s, SD = 106.

4.4.7 Discussion

The human brain is really good at covering up errors in the audio simulation. When physical artifacts are augmented with virtual sound, we can observe the "ventriloquist effect" [Alais, Burr, 2004]: smaller errors in the combination of tracking and rendering are simply ignored, and the sound source "snaps" to the object. When no physical anchor is present, the source position only needs to be perceived as stable, as exact positioning is not required. Even a total failure in the tracking system can be interpreted into something meaningful. In one case we encountered problems with the transmission of location data from the server to the client, so only orientation was updated. The user of the affected device commented this with *"That was amazing! After some time, the voices started walking with me!"*

The highest realism is achieved using head tracking.

Device tracking is a feasible alternative and requires less hardware.

As a conclusion from these experiments, we recommend to use head tracking if realism and navigation close to the virtual sound sources is important, e.g., when exploring small artifacts in a museum. In our experiment, the differences in the ratings between the three orientation measurements are very small and not statistically significant. Considering the small number of participants, this is not surprising, but we expect this not to change with a larger sample. If the focus of the implementation is rather on serendipitous discovery, the sources are further apart, or the use of additional hardware poses a problem, using the available sensors of a smartphone may be sufficient. Even in the case of nearby sources as in the first experiment, we think it is more a matter of communicating the functionality (see section 4.6).

Looking at the use of audio augmented reality as a navigational tool, the dimensions of the audio space increase dramatically, e.g., for city wide navigation. In such a scenario, the sources would probably become larger, blurring the error of the measurement.

4.5 Impact of Elevation on Audio Augmented Reality

Head orientation is measured in all three angles anyways.

While in this previous section we investigated if there is an opportunity to simplify the orientation sensing, we are now going to take a look at audio rendering complexity. Most of the existing AAR systems only simulate sources in the horizontal plane, i.e., they are perceived as to be at the level of the listener's ears. This is mostly based on two reasons: inertial measurement units (IMU) have significantly improved over the past 10 years, starting from simple two-dimensional compass ICs (e.g. the Honeywell HMC-6352) which had to be kept level to provide stable output, to modern ICs (e.g., the InvenSene MPU-9250[6]) which measure 9 degrees of freedom (DOF) and output a tilt-compensated and filtered heading.

Figure 4.12: Tilting your head changes the relative elevation of the virtual sound sources. Using an HRTF-based algorithm allows to integrate richer sensor data including head pitch and roll into the rendering, which we expected to increase localization accuracy.

HRTF-based rendering is capable of simulating source elevation.

The second aspect is that only HRTF-based rendering algorithms are capable of simulating source elevation, which are more complex to implement and require more resources. In the previous experiments, we used a medium-level rendering algorithm that simulates more factors than simple stereo panning, but not using HRTFs, and thus

[6] invensense.com

not capable of simulating source elevation. Since modern smartphones provide enough computational power to render several sources in parallel with a high fidelity [Sander et al., 2012] and the IMUs provide all required information, we want to evaluate the use of simulating source elevation. This can improve the experience with AAR systems in two ways:

1. including more DOFs into the rendering might increase the spatial resolution and allow a better discrimination between proximate sources.
2. it increases the realism for setups where the physical counterparts to virtual sources are placed at different heights.

4.5.1 Experiment

To measure the effect of rendering fidelity and simulated source elevation on the ability to localize virtual sound sources, we conducted the following experiment: We first measured the localization performance for sources that are all at equal height, approximately at the level of the user's ears. As tilting your head has an impact on the relative elevation of the source (cf. Figure 4.12), we ran the experiment with and without simulating the elevation (*flat* vs. *elevation*). We placed 17 cardboard tubes of 140 cm height in the range of $[-40°, 40°]$ at $5°$ intervals and at two meter distance from the listener (cf. Figure 4.13). To test the different angular distances between two candidate sources, we marked some of the tubes as active by placing a physical marker on top of it. We tested $5°$, $10°$, $15°$, and $20°$ spacings, by marking all, every other, every third, or every fourth tube as active respectively. Participants were instructed to look at the source directly in front of them before every trial to allow the experimenter to check the compass calibration and recalibrate if necessary. A sound was then played at the position of one of the markers, and participants had to locate the source and say its number aloud. Participants were encouraged to perform some trials before the experiment to become acquainted with the system. We only played

We only tested sources in front of the listener.

We placed sources at varying horizontal distance.

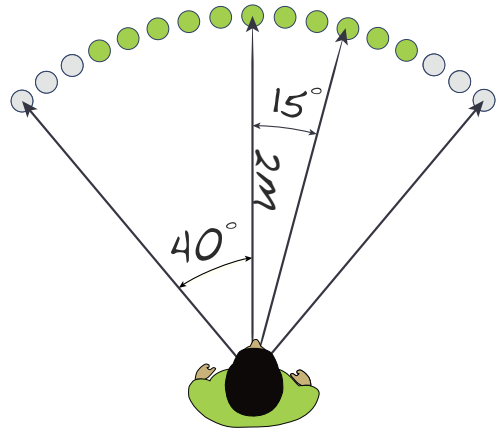


Figure 4.13: We placed 17 cardboard tubes at 2 m distance to the listener and tested sources spacings of 5°, 10°, 15°, and 20°. No sound sources were rendered at the outer six sound source positions (grey) to not artificially limit the choice of possible candidates towards the border. The active sound sources (green) were thus placed within a range of $[-\mathbf{25}°, \mathbf{25}°]$

We placed the sources at two different heights and 5° horizontal spacing.

sounds in the range of $[-25°, 25°]$ since head movement was not restricted and participants would turn their head towards the sound anyway. This had the advantage that the choices of candidate sources was not artificially limited towards the outermost sources. In the third condition (*height*) we placed virtual sound sources, together with their physical markers, at two different heights (140 cm or 70 cm), again at 5° intervals, to see if the use of different heights in a display of virtual sound sources has an influence on recognition accuracy. We only tested the smallest angle in this condition to make sure that we can see some effect, as we expected larger angles to be discernible successfully anyways. This resulted in a total of 90 trials per participant.

The sound sample we used was a continuous monologue of a male voice, which is closer to a real use-case than the white noise samples used in technical measurements

[Wenzel et al., 1991]. The order of conditions was counterbalanced, and locations were randomized using latin squares. We recorded task completion time, head orientation in three degrees of freedom, accuracy, and evaluated the perceived presence in the virtual environment using a questionnaire [Witmer, Singer, 1998].

We used a continuous male monologue as sound sample.

4.5.2 Technical Setup

We used the KLANG:kern[7] spatial audio rendering platform running on an Apple iPad Air 2, and tracked head orientation using the Jabra Intelligent Headset[8] The rendering uses a generalized HRTF which has a resolution of 1° in horizontal and 5° in vertical direction. We measured a minimum audible angle of around 6° in horizontal and 16° in vertical direction. The headset reports changes in head orientation at a rate of around 40 Hz and has a specified latency of around 100 ms, which is noticeable [Brungart et al., 2005] but well below the limits of 372 ms defined in [Mariette, 2010]. While sensor data was transmitted via Bluetooth, we used a wired connection for the audio output to minimize latency.

4.5.3 Results

A total of 22 users participated in the study (3 female, average age 28 years, $SD = 5$). None reported having a known problem with spatial hearing. 50% of the participants reported having prior experience with audio augmented reality.

In a museum setting, the most relevant result for real-world performance is the number of successfully recognized sources. Compared to the recognition rates reported for other systems [Heller, Borchers, 2015], the HRTF-based rendering we used in this experiment achieves similar rates

[7] klang.com
[8] intelligentheadset.com

The HRTF-based rendering algorithm doubled the horizontal resolution compared to our previous rendering algorithm. Recognition rate did not increase with sources being placed at different heights.

at an angular distance reduced by nearly 50% (normalized to 1m distance to source). While some participants were able to achieve an accuracy of 70% and higher for a source spacing of 5°, this spacing is not recommended for practical use without further guidance (*elevation*: M = 29%, SD = 45%, *flat*: M = 30%, SD = 46%). Even when sources are at different heights, recognition rates do not increase significantly (*height*: M = 33%, SD = 47%, $F_{2,657} = 0.37, p = .69$). The recognition rates vary largely at this spacing, which can be explained by the fact that it is in the order of the minimum audible angle of the rendering algorithm. If we increase the angular distance of source candidates (cf. Table 4.1), we can see that the recognition rate also increases, but the differences between *flat* and *elevation* remain marginal. Overall, the recognition rate is significantly higher for participants with prior experience ($F_{1,1743} = 8.02, p < .005$), which indicates that after some time, users accommodate to the auditory experience. For example, at 10° spacing, the average recognition rate jumps from 56% (SD = 49%) to 70% (SD = 46%) with prior experience. A post-hoc t-test showed this difference to be significant ($p < .005$).

Participants took much longer to localize the sources on two different levels in the *height* condition (M = 12.64 s, SD = 8.5) compared to the other two conditions with a source spacing of 5°, with *elevation* and *flat* being quite similar (M = 9.65 s, SD = 5.31 vs. M = 9.02 s, SD = 3.69, $t_{657} = -1.08, p = 0.2795$). A post-hoc Student's t-test with Bonferroni correction revealed the differences between *height* and the other conditions to be significant (vs. *flat*: $t_{657} = 6.16, p < .0001$). Again, prior experience has a significant impact. The task completion time in the *elevation* condition is significantly shorter for participants with prior experience (M = 7.05 s, SD = 3.9 vs. M = 8.2 s, SD = 8.2), which indicates that after an accommodation phase, localization performance increases [Majdak et al., 2013].

We calculated the root mean squares (RMS) of all three head orientation angles as an indicator of how much participants moved their head along the respective axes (cf. Figure 4.14). First we compare the differences across all three conditions in the RMS angles for the 5° source spacing. The amount

participants turned their head left and right is very similar in the *height* (M = 18.96°, SD = 5.04) and *elevation* (M = 18.54°, SD = 2.64) condition, and only slightly higher in the *flat* condition (M = 20.26°, SD = 6.38). A repeated measures ANOVA with user as random factor showed a significant effect of the condition on Roll ($F_{2,42} = 6.4287, p = .0037$) and Pitch ($F_{2,42} = 4.2739, p < .05$). Post-hoc t-tests with Bonferroni correction show that participants rolled their head significantly more in the *height* condition than in the other two ($p < .01$). The RMS pitch angles are only significantly different between the *height* and *flat* condition, which shows that, although not really noticeable, participants nodded while localizing the sources if all three angles were included in the rendering. For the other spacings, the RMS angles do not not differ significantly between the three conditions.

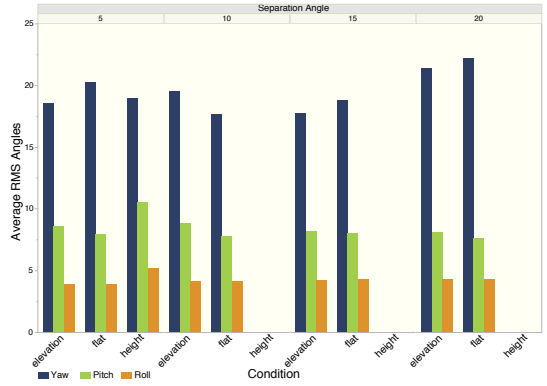


Figure 4.14: The average RMS angles for yaw (left-right rotation), pitch (looking up or down), and roll (tilting head sideways) by condition and source separation angle. The *height* condition was only tested at 5° intervals. RMS Yaw angles are largest as the task was to localize sources in horizontal direction. While the RMS pitch angles are slightly higher in the *height* condition, the difference to the *elevation* condition is not significant.

The median ratings given on a five point Likert scale (1 being the best) only differed marginally. None of the differ-

	Condition	Angle 5°	10°	15°	20°
Recognition rate	*elevation*	29 %	62 %	76 %	86 %
	flat	30 %	64 %	80 %	85 %
	height	33 %			
Task completion time	*elevation*	9.65 s	9.18 s	6.24 s	5.43 s
	flat	9.02 s	7.21 s	5.62 s	5.35 s
	height	12.64 s			

Table 4.1: Percentage of correctly identified sources and task completion time by angular distance. Including all 3 degrees of freedom of the head into the rendering does not significantly increase the ability to localize the origin of sounds.

Perceived presence is similar across conditions.

ences in the ratings was statistically significant according to a Wilcoxon signed rank test. Participants felt equally able to localize sounds in both conditions (Mdn = 3, IQR = 2). When asked how consistent the experience with the virtual environment seemed with the real world, participants gave slightly better ratings for the *elevation* condition (Mdn = 2, IQR = 2 vs. *flat*: Mdn = 3, IQR = 2.5). The ratings for the perceived naturalness of the virtual experience were the same for both conditions (Mdn = 2, IQR = 2). The participants rated the system as very responsive (Mdn = 1, IQR = 1) and that they could adapt to it quickly (*elevation*: Mdn = 1, IQR = 1.5; *flat*: Mdn = 1, IQR = 2).

4.5.4 Discussion

The use of a rendering algorithm based on generalized HRTFs creates a more realistic impression than any simpler rendering algorithm. Our results show that the increase in horizontal localization accuracy is drastic. In a realistic setting where the audio sample may contain additional information to which physical location it belongs, the recognition rates may rise even for the smaller source separation angles.

Contrary to what we expected, using the elevation rendering of the HRTF-based algorithm to place the sources at different heights did not increase horizontal resolution. While the euclidean distance between the sources increases when

placing sources at different elevations, the impact of the vertical difference is minimized by the generalized HRTF. We know that the resolution of human sound localization is lower in the vertical than in the horizontal plane and that the generalization of HRTFs mostly affects the features used to the determine the elevation of a source [Zotkin et al., 2004]. The inclusion of head roll and pitch into the rendering algorithm did not further improve horizontal resolution. While users moved their head more in the *elevation* condition, which indicates that it is noticeable, they also often asked if there was any difference between the *elevation* and *flat* conditions, indicating that the difference was barely noticeable consciously.

Perception of elevation is difficult with generalized HRTFs.

4.6 Smartphones as Platform for Audio Augmented Reality

In the previous two sections we evaluated where simplifications in a MAARS setup are potentially possible, either on the sensor side, or in the rendering. In this section, we will now verify our assumptions by testing the user's ability to localize proximate sound sources with a MAARS implementation that runs on a standard smartphone only. As stated previously, with location and orientation sensing, and sufficient processing power, modern smartphones integrate all components required to build a MAARS. However, using the device compass as opposed to tracking the head orientation decreases the realism of the simulation. By providing a well-defined mental model, we avoid disappointing users through a possible lack of realism as they discover that turning their head does not influence the audio output. We propose AudioScope, a metaphor that turns your smartphone into a virtual directional microphone. The user probes the audio space by simply pointing the device in different directions. If the sound source is to the left of the device, the sound on the left audio channel is louder and vice versa.

Can we use the smartphone as only hardware resource for MAARS?

We communicate the metaphor of a virtual directional microphone.

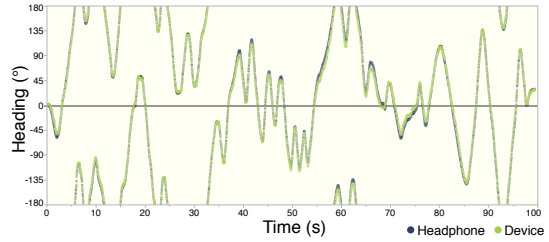


Figure 4.15: We compared the IMU of the Intelligent Headset (blue) and the iPhone 5S (green). Both report heading with a similar characteristic and update rate (Taken from [Heller, Borchers, 2015]).

4.6.1 Implementation

We used the Intelligent Headset and an iPhone 5S.

The implementation is very similar to the one used in section 4.4. To be able to track the user in a larger area, instead of using an optical tracking system, we measured head position at 34 Hz using a Ubisense[9] ultra wideband (UWB) location tracking system with an accuracy of around 5 cm. Head orientation was measured with the IMU of an Intelligent Headset, while device orientation was measured using the IMU of an Apple iPhone 5S. We compared both IMUs, which report changes in heading with an update rate of around 40 Hz, and found that they have a similar characteristic, with an average difference of only 4.8°(SD = 3.4°) (cf. Figure 4.15). The absolute average orientation error of the iPhone IMU is 4.25°(SD = 3.05°). The specified overall latency of the headphone orientation measurement is around 100 ms, which is noticeable [Brungart et al., 2005] but well below the limits of 372 ms defined in [Mariette, 2010].

Spatial audio rendering was implemented using the OpenAL framework in iOS 7.1, with the `ALC_EXT_MAC_OSX` extension enabled. As in the previous experiments (section 4.4), it uses a spatialization based on the spherical head model and includes interaural level and time difference, head filtering, and a frequency-dependent distance model as filters. To enhance front-back separation of sources, we added a low-pass filter to sources when they are behind the

[9] ubisense.net

listener. The intensity increases linearly from 0dB to -36dB for azimuth angles between 90°(side) and 180°(back). The minimum audible angle of the rendering is around 4°. This method, although less realistic than algorithms using individual, natural body cues in form of head-related transfer functions (HRTF), is a good representative of spatial rendering on mobile devices.

4.6.2 Evaluation

In our experiment, participants were instructed to navigate to single proximate sources with either head or device tracking enabled. We placed 24 loudspeakers spaced by 15°at a height of 150 cm forming a circle of 5 m diameter (Figure 4.16). As in this experiment we only wanted to compare the impact of the orientation measurement using the same rendering in both conditions, the loudspeakers did not play any sound but were mere physical representations of the virtual sound sources.

In the two conditions of our experiment, the audio rendering algorithm used the orientation either from the *head* or from the *device*. Participants started every trial standing in the center of the circle facing source no. 1 and were instructed to identify the currently active source as quickly as possible. Correct alignment of the heading information with the physical setup was verified before each trial. In a real scenario, e.g., a museum or a public place, users might not be able to get close to the sources. To account for this factor, and to make sure that the experiment revealed the impact of orientation measurement, participants were instructed to move only in an inner circle of 3 m diameter, such that they had to determine the exact sound source from a distance of approximately 1 m. We used an audio sample of a male voice reciting colors at a fast pace[10].

Participants had to localize sources using head-tracking or device-tracking.

We used a within-subjects design with a balanced order of conditions, and the order of active sound sources was randomized using Latin squares. Every participant had to nav-

[10] http://hci.rwth-aachen.de/public/AudioScope/Colors.au

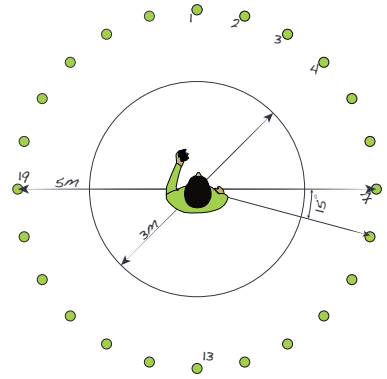


Figure 4.16: We placed 24 virtual sound sources, spaced by 15°in a circle of 5 m diameter. Participants had to start every trial standing in the center, facing source no. 1. They could move freely within the inner 3 m circle (Taken from [Heller, Borchers, 2015]).

igate to all 24 sources in the *head* and *device* measurement condition and had to complete a 10-trial training before each condition. We measured the time from users starting each trial by pressing the start button on the smartphone, until they confirmed standing in front of the audible source by pressing a second button. Participants had to name the source that they assumed was playing. We recorded the paths the users took to walk to the sources, along with the orientation fed into the rendering algorithm. After each condition, participants had to fill out a questionnaire about their perceived presence in the virtual environment [Witmer, Singer, 1998] on 5-point Likert scales.

In total, 20 users, 8 female, 12 male, aged 21 to 33 (average 26), participated in the study. None reported a hearing disorder or known problems with spatial hearing. Seven had prior experience with audio augmented reality systems.

4.6.3 Results

The average time users took to navigate to the sound source was 15.71 s (SD = 8.51) in the *head* condition and 17.22 s (SD = 9.72) in the *device* condition. A mixed model repeated measures ANOVA revealed this difference to be statistically significant ($F_{1,916} = 14.79, p < .0001$). At the same time, the rate of correctly recognized sources was 65% (SD = 0.28) for *head* tracking and 69% (SD = 0.26) for *device* tracking. This difference is not significant ($p = .91$) according to a Wilcoxon Signed Rank test. The recognition rates seem fairly low, but considering that the 15°spacing between our sources is in the range of the localization error of virtual sound sources [Middlebrooks, 1999; Wenzel et al., 1993] and that we used a rather simple rendering algorithm, this is not surprising. Most of the errors were only off by one source to the left or right. If we count these "off-by-one" answers as correct, then the recognition rates climb up to 97% (SD = 0.1) for *head* and 98% (SD = 0.1) for device tracking. No front-back confusions occurred. While people without prior experience with audio AR were significantly slower in the device condition (M = 15.97 s, SD = 8.5 vs. M = 18.8 s, SD = 10.4, $F_{1,622} = 13.207, p = .0003$), no significant difference between conditions could be found for participants with prior experience (*device*: M = 14.3 s, SD = 7.5; *head*: M = 15.2 s, SD = 8.47, $F_{1,334} = .8466, p = 0.358$). The average distance travelled only differs by half a step between conditions (*device*: M = 9.00 m, SD = 5.4; *head*: M = 8.7 m, SD = 6.32, $F_{1,961} = 4.5339, p = .0335$).

Novices are slower in the device condition.

No significant difference was found for participants with prior AAR experience.

Median ratings from the perceived presence questionnaire did not differ by more than one item on the 5-point Likert scale. Not surprisingly, *head* tracking (Mdn = 5, IQR = 1.75) was perceived more natural than *device* tracking (Mdn = 4, IQR = 1). For both conditions, the experience was rated to be consistent with the real world (*head*: Mdn = 4, IQR = 1.75; *device*: Mdn = 4, IQR = 2), and participants felt able to localize sounds well (*head*: Mdn = 4, IQR = 0; *device*: Mdn = 4, IQR = .75). Wilcoxon signed rank tests only revealed the ratings for the *natural interface* to be significantly better (Z=60, p=.022) for *head* tracking, all other ratings did not differ significantly between conditions.

Perceived presence is similar in both conditions.

The tracking technology did not seem to interfere with the experience as both the responsiveness (*head*: $Mdn = 5$, $IQR = 1$; *device*: $Mdn = 5$, $IQR = 1$) and the perceived delay (*head*: $Mdn = 4$, $IQR = 1.75$; *device*: $Mdn = 4$, $IQR = 2$) received similarly high ratings in both conditions. After the experiment, the participants felt proficient with the interface both with *head* tracking ($Mdn = 4$, $IQR = 1$) and *device* tracking ($Mdn = 4.5$, $IQR = 1$). Again, no significant difference was found between the conditions.

4.6.4 Discussion

Overall, the differences between both orientation tracking metaphors are fairly small. Out of the answers on the questionnaire, 85% of the ratings are above 3 out of 5 on a Likert scale (5 being the best). We are thus confident that the acceptance of the device orientation measurement is high. For people with prior audio AR experience, no significant difference in task completion time could be found which suggests that the metaphor is easy to adopt. The fact that the average task completion time was slightly longer (1.5 s) in the *device* condition is not critical in practice. Other studies have revealed that a longer task completion time can also be a result of people enjoying the experience [Vazquez-Alvarez et al., 2012], which in case of an audio guide for museums, is the primary focus. Furthermore, we observed that users experimented with the handling of the device tracking even though they completed the 10 training trials prior to the experiment.

If hardware requirements are an issue, using device tracking is a feasible alternative.

The low recognition rates show that the distance between the sources was at the limit of what can be differentiated with our rendering. As stated above, the 15°spacing is in the range of the localization error of virtual sound sources. Participants spent more time in a 1.5 m radius around the active source than in the rest of the field, which indicates that they took a long time to differentiate between two candidate sources. This problem can be solved by either placing the sources further apart, or by providing additional context, e.g., a beacon sound that relates to the physical object.

Some participants stated that they felt faster with the device tracking, and one mentioned the head tracking being more difficult immediately after switching conditions. On the other hand, some participants mentioned being confused by the metaphor of the virtual directional microphone, since when moving the smartphone to the right of a source, the left channel becomes louder.

The focus of this evaluation is the interaction metaphor, for which the use of an external location tracking system is acceptable. However, indoor location tracking is currently a focus of both researchers and smartphone manufacturers. Technologies such as Estimote beacons (estimote.com) support that we can expect significant improvements in accuracy in the near future, making smartphones a complete platform for MAARS.

4.7 Conclusion

In this chapter we augmented the auditory interaction bandwidth via virtual audio spaces the user can explore freely. Such audio spaces enable much more interactive listening experiences than the traditional stereo recording. Instead of being a static listener, one can navigate through the orchestra playing [Camurri et al., 2007]. But such audio spaces can also be used for non-visual navigation, both in small rooms and outdoors.

We increased the audio interaction bandwidth.

This increase in audio interaction bandwidth, however, needs to come with an appropriate controller. The number of parameters defining this environment, even in a simple implementation is substantially larger than what is necessary to control stereo playback. The most important dynamic parameters are listener position and orientation, since these determine what is to be heard even when the rest of the environment is static. We evaluated two possible controllers for these parameters that use a natural mapping: tracking the user's head and tracking the user's smartphone. Using head tracking results in the most natural interaction, since it is a simulation of our everyday interaction with spatial audio. Using device orientation, com-

To make spatial audio accessible, we increased the haptic input bandwidth to simulate realistic listening behavior.

municated through a proper metaphor, however, requires less hardware while not being perceived as significantly less natural.

The technical evolution has brought up off the shelf hardware which is capable of simulating a virtual audio space on a mobile device at the same quality level as complex hardware systems a decade ago. The drastically lower hardware requirements make spatial audio available to a broad audience. However, virtual audio spaces still offer interesting research questions, on, e.g., how people interact with moving virtual sound sources.

After having augmented the auditory modality in this chapter, we will now summarize and conclude the thesis, and give an outlook on future research directions.

Chapter 5

Summary and Future Work

The language of interaction with time-based media, like audio, has long been defined by technological constraints. Audio is dependent on a recording and playback device, as it only exists in the time-domain. Appearing at the end of the 19th century, audio recording is a comparatively new technology, that has no century-old ancestors. In contrast to, for example, drawing, the tools available to manipulate audio recordings are not the result of an evolution over centuries. Instead, since the first successful attempts to record and play back audio, several different media were used, but the playback devices were always designed around that medium. The interfaces of these devices are access to technical components such as motors or playhead mechanics.

Audio playback interfaces are designed around the medium.

When designing natural interfaces for tasks that have an "analog" ancestor, we can revert to these and simulate their behavior as good as we can. In the example of drawing, we have graphics tablets that not only sense the position of the pen on screen, but also its tilt, applied pressure, and relative rotation to approximate the reaction of a real pen or brush. The question this thesis investigates is, how a natural interface for audio playback might look like. In line with the research framework described by Jacob et al. [1993] and definitions by Valli [2008] and Dix [2004], we argue that an increase in interaction bandwidth results in more natural

How to design natural interfaces for audio playback?

interaction. However, the interaction should become simpler, which means that just displaying more information or providing more controls will not result in an adequate interface.

We did a systematic exploration of the modality space.

Therefore, this thesis provided a systematic exploration of the space defined by the three modalities involved in audio playback interaction: auditive, visual, and haptic. For each modality, we presented interfaces that simplify the interaction by building on natural human skills and at the same time, increase interaction bandwidth.

We developed wearable controls with an increased haptic interaction bandwidth for mobile music players.

Chapter 2 showed some examples of wearable controls for mobile music players which leverage our fine grained motor skills by building on the natural affordances of cloth. Pinstripe and Intuitex both build on folding a piece of fabric as interaction and sense the displacement of this fold in one or two dimensions. Fabritouch, a wearable touchpad, accounts for the prevalent gesture interaction users know from their smartphones and tablets.

We increased the visual interaction bandwidth of a DJ turntable to re-create visual information that was present on traditional vinyl records.

In the history of audio playback devices, the turntable is among those with the highest interaction bandwidth. Digital vinyl systems lifted this purely analog technology in the digital era, but at the expense of a loss of visual information that was present on traditional vinyl records. In chapter 3, we used an extension of its visual output bandwidth to bring back this information. This results in a fully embodied device, where haptic input and visual output are physically co-located.

We extended the audio interaction bandwidth with virtual audio spaces.

We concluded our exploration with the auditory modality. The common stereo recording does not make much use of the spatial perception of sound we are capable of. Instead, the position of the sound sources is fixed once recorded and if experienced through headphones, are mostly perceived as to be in the head. Spatial audio rendering allows us to modify the spatial arrangement of sound sources the moment we experience them. By simulating the physical effects that a sound signal is exposed to on the way from its origin to the human ear, these rendering algorithms create the impression that the sound emanates from a source located in the physical space. In combination with orienta-

tion and location tracking of the user, we can create experiences with virtual sound sources being perceived to be at fixed positions in the real world. To make the sheer amount of playback parameters accessible, we tracked the user's head position and orientation and fed this information to the rendering algorithm, simulating how the virtual sources would sound like in the real world.

5.1 Contributions

The general question regarding how to increase the interaction bandwidth in interfaces for audio playback can be solved in various ways. Professional audio equipment offers physical controls for a large number of parameters, which gives quick access but also requires space and a deep understanding of the interface and signal flow. The more interesting aspect is how to implement this bandwidth increase in a way that results in a natural interface which leverages our natural skills. This thesis approached this question through a systematic analysis of bandwidth increase in the different modalities involved: haptic, visual, and auditive.

5.1.1 Haptics

To increase the haptic interaction bandwidth for mobile music players, we presented three wearable controllers building on the natural affordances of cloth. Pinstripe was the first textile controller using a textile fold for a continuous linear value input. We evaluated the concept in conjunction with a music player, where either volume or the selected song in a playlist could be changed. Participants were able to successfully achieve the task and the prototype got positive ratings for its usability. The underlying concept of using the fold as input was not as obvious as expected, but after a brief explanation all participants immediately understood it. Intuitex extends this unique concept to a second dimension, increasing the input possibilities. Due

Pinstripe is easy to understand. Participants all successfully completed the tasks.

to technical limitations, the physical resolution of the two-dimensional input is comparatively low, which requires an adaptation of the software interface. In combination with menus that can be navigated using simple continuous gestures, like marking menus [Kurtenbach, Buxton, 1993], it still is a powerful wearable controller. Our third prototype, Fabritouch, is a wearable touchpad. The ideas to build a wearable touchpad has been around in the DIY community for some time, but none of these prototypes as been formally evaluated, making it difficult to estimate how well these are suited for deployment. We built our own specimen and evaluated its input capabilities under realistic conditions. Results show that the underlying support material has a significant impact on user's performance. A wearable touchpad designed to be integrated at the upper thigh should thus not be tested on a table. Furthermore, our results show that while in motion, only simple gestures like strokes should be used as input as garment shift and crumpling causes unstable sensor readings.

Only simple gestures should be used as control input while walking.

These three prototypes all increase the haptic interaction bandwidth in a way such that our natural manual capabilities are used to a larger extend than with simple binary buttons. Additionally, Pinstripe allows one to control a parameter in various granularities.

5.1.2 Visual

To explore the visual interaction bandwidth, we built on a device that already has a high haptic interaction bandwidth: the DJ turntable. When this purely analog technology felt increasingly out of date because the music production and distribution already had moved to digital, digital vinyl systems transformed the turntable from a playback device to a tangible controller. During this process of digitalization, information that was once visible on a traditional record, like track start and end, was lost. Although this information is displayed on the laptop, the DJ has to split attention between the turntable she manipulates and the visualization on the screen. We used top-projection to extend the visual output bandwidth of the turntable to re-

Haptic input and visual output are co-located in the same device.

create this information. In two lab studies, we evaluated the system with professional DJs. While we could not observe that participants paid significantly less attention to the computer screen, the overall feedback was very positive. In an online survey we evaluated the acceptance of such a system within the DJ community which was very well received. While our primary focus was to make the interaction simpler for the DJ by co-locating haptic input and visual output, it is also beneficial for the audience. Our system addresses the general problem in electronic music performances of the non-obvious connection between actions on stage and their audible result. A visualization like DiskPlay makes the actions of the DJ on the turntable more transparent.

An evaluation with professional DJs was very positive.

DiskPlay makes the actions of the DJ transparent to the audience.

5.1.3 Auditory

To extend the auditory interaction bandwidth we took a closer look at spatial audio displays. These leverage our natural capability to localize the origin of a sound. In combination with location and orientation tracking of the user, we can create experiences in which virtual sound sources are perceived as to be fixed in the physical space. While the processing power of a current smartphone is sufficient to render a scene with several simultaneous sources, tracking head orientation still requires dedicated hardware. We investigated if we can approximate head orientation measurement using device orientation measurement. We analyzed the differences in paths and head, body, and device orientation of participants navigating through a virtual audio space. Results show that navigation is possible even with simplified rendering algorithms. For highest realism, head-tracking should be used, but if this is not possible, for whatever reason, using device orientation is an adequate replacement. To avoid confusion when using device orientation, we introduced the metaphor of a virtual directional microphone. Participants of our study were as fast in detecting the origin of a sound source as with head tracking and the perceived presence in the virtual environment was equally high. Finally, we evaluated if placing the virtual sound sourced at different simulated heights can be used

Head tracking results in highest realism.

Device tracking is a feasible alternative with fewer hardware requirements.

to reduce their horizontal distance. Our results show that using elevation has no effect as most participants could not match the change in the audio output to different heights.

5.2 Future work

Each of the fields presented in this thesis offers interesting opportunities for follow-up research, however, at this point we will concentrate on the high-level questions.

How to quantize interaction bandwidth?

In this thesis and in the related literature, the term interaction bandwidth is used somewhat informally. To increase the generative power of our systematic exploration, a defined quantification along each modality is necessary. Figure 5.1 depicts the interaction bandwidth along each of the modalities we considered. The blue areas describe the interaction bandwidth of the original interface, whereas the green areas describe the bandwidth after augmentation. Right now, we can only argue why a certain interface reaches a specific level of interaction bandwidth, but there are no hard numbers. Such a visualization is a useful tool to find the modalities considered the least in a certain interface and look for potential extensions. Once this modality is identified, one can take a look at more specific design spaces such as the design space if input devices [Card et al., 1991].

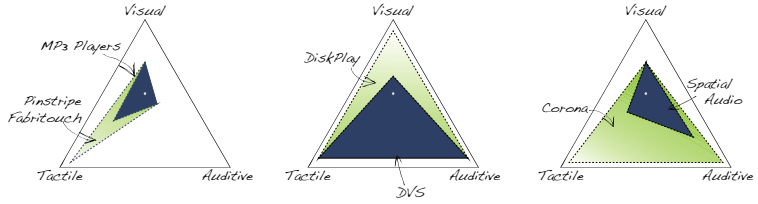


Figure 5.1: A graphical representation of the augmentation in interaction bandwidth along each of the modalities. The blue areas describe the interaction bandwidth of the original interface whereas the green space describes the interaction after augmentation.

While the haptic and visual modalities have received much attention in terms of structural analysis (e.g., [Card et al., 1991], [Reinecke et al., 2013]), there is no such overarching design framework for spatial audio displays. A formal taxonomy including aspects such as beacon sound type, continuous vs. interval beacons, conveyed information, passive or active interaction, would provide an overview and show possible opportunities for exploration.

A taxonomy could help discover underrepresented modalities.

Finally, we can apply the same systematic process to other areas of HCI. As mentioned in the introduction, we can look at drawing and analyze how large the interaction bandwidth is along the modalities involved. In the case of a graphics tablet with included display, the visual output bandwidth is already very high. But the haptic output bandwidth could be increased by adding a simulation of various surfaces, like paper, cardboard, or canvas.

Appendix A

Questionnaires

A.1 Pinstripe Questionnaire

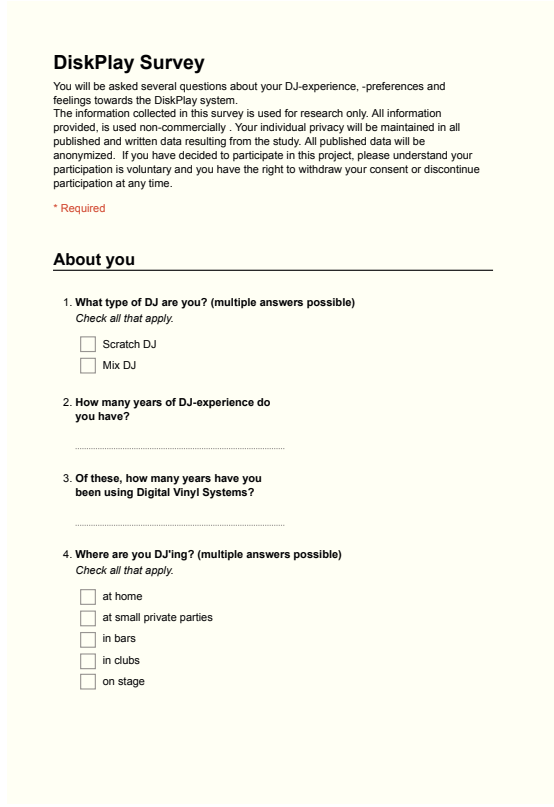
The Pinstripe questionnaire is a standard SUS questionnaire [Brooke, 1996] that we extended to include some questions specific to our project.

Age: _____	Participant #: _____
Handedness: ○ left ○ right ○ ambidextrous	Sex: ○ male ○ female
1. I think that I would like to use this system frequently ○ ○ ○ ○ ○ 1 2 3 4 5 agree strongly disagree strongly	9. I felt very confident using the system ○ ○ ○ ○ ○ 1 2 3 4 5 agree strongly disagree strongly
2. I found the system unnecessarily complex ○ ○ ○ ○ ○ 1 2 3 4 5 agree strongly disagree strongly	10. I needed to learn a lot of things before I could get going with this system ○ ○ ○ ○ ○ 1 2 3 4 5 agree strongly disagree strongly
3. I thought the system was easy to use ○ ○ ○ ○ ○ 1 2 3 4 5 agree strongly disagree strongly	11. I felt that I could control the volume precisely ○ ○ ○ ○ ○ 1 2 3 4 5 agree strongly disagree strongly
4. I think that I would need the support of a technical person to be able to use this system ○ ○ ○ ○ ○ 1 2 3 4 5 agree strongly disagree strongly	12. I felt that it was easy to switch between tracks ○ ○ ○ ○ ○ 1 2 3 4 5 agree strongly disagree strongly
5. I found the various functions in this system were well integrated ○ ○ ○ ○ ○ 1 2 3 4 5 agree strongly disagree strongly	13. I had difficulties navigating the graphical menu ○ ○ ○ ○ ○ 1 2 3 4 5 agree strongly disagree strongly
6. I thought there was too much inconsistency in this system ○ ○ ○ ○ ○ 1 2 3 4 5 agree strongly disagree strongly	14. I would be uncomfortable to use a final version of the system in public ○ ○ ○ ○ ○ 1 2 3 4 5 agree strongly disagree strongly
7. I would imagine that most people would learn to use this system very quickly ○ ○ ○ ○ ○ 1 2 3 4 5 agree strongly disagree strongly	15. I would buy clothing with this functionality to control my portable music player ○ ○ ○ ○ ○ 1 2 3 4 5 agree strongly disagree strongly
8. I found the system very cumbersome to use ○ ○ ○ ○ ○ 1 2 3 4 5 agree strongly disagree strongly	16. I would be willing to pay an extra _____ EUR for clothing that included this functionality

Figure A.1: The questionnaire used for the Pinstripe evaluation

A.2 DiskPlay Questionnaire

The DiskPlay questionnaire was published online and announced in various forums.



DiskPlay Survey

You will be asked several questions about your DJ-experience, -preferences and feelings towards the DiskPlay system.
The information collected in this survey is used for research only. All information provided, is used non-commercially . Your individual privacy will be maintained in all published and written data resulting from the study. All published data will be anonymized. If you have decided to participate in this project, please understand your participation is voluntary and you have the right to withdraw your consent or discontinue participation at any time.

* Required

About you

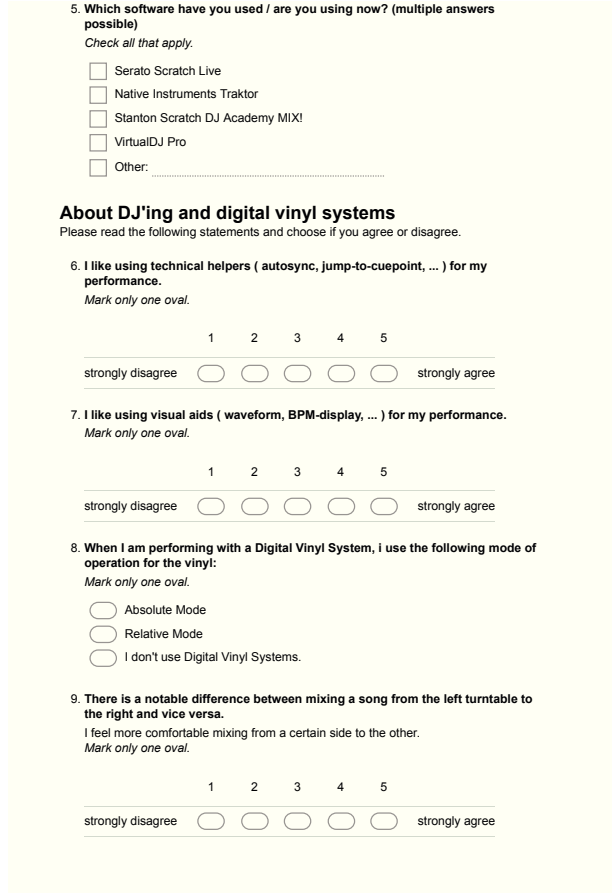
1. **What type of DJ are you? (multiple answers possible)**
   Check all that apply.
   - [ ] Scratch DJ
   - [ ] Mix DJ

2. **How many years of DJ-experience do you have?**

3. **Of these, how many years have you been using Digital Vinyl Systems?**

4. **Where are you DJ'ing? (multiple answers possible)**
   Check all that apply.
   - [ ] at home
   - [ ] at small private parties
   - [ ] in bars
   - [ ] in clubs
   - [ ] on stage

Figure A.2: The Diskplay Acceptance Questionnaire (Page 1)



5. **Which software have you used / are you using now? (multiple answers possible)**
Check all that apply.

- [ ] Serato Scratch Live
- [ ] Native Instruments Traktor
- [ ] Stanton Scratch DJ Academy MIX!
- [ ] VirtualDJ Pro
- [ ] Other: .....

About DJ'ing and digital vinyl systems

Please read the following statements and choose if you agree or disagree.

6. **I like using technical helpers (autosync, jump-to-cuepoint, ...) for my performance.**
Mark only one oval.

	1	2	3	4	5	
strongly disagree	◯	◯	◯	◯	◯	strongly agree

7. **I like using visual aids (waveform, BPM-display, ...) for my performance.**
Mark only one oval.

	1	2	3	4	5	
strongly disagree	◯	◯	◯	◯	◯	strongly agree

8. **When I am performing with a Digital Vinyl System, i use the following mode of operation for the vinyl:**
Mark only one oval.

- ◯ Absolute Mode
- ◯ Relative Mode
- ◯ I don't use Digital Vinyl Systems.

9. **There is a notable difference between mixing a song from the left turntable to the right and vice versa.**
I feel more comfortable mixing from a certain side to the other.
Mark only one oval.

	1	2	3	4	5	
strongly disagree	◯	◯	◯	◯	◯	strongly agree

Figure A.3: The DiskPlay Acceptance Questionnaire (Page 2)

10. **I consider in-track navigation with Digital Vinyl System to be harder than with traditional vinyl.**
Mark only one oval.

	1	2	3	4	5	
strongly disagree	◯	◯	◯	◯	◯	strongly agree

11. **When I am performing, it bothers me to look back and forth between computer screen and turntable.**
For example: When you are searching for certain positions in a track, checking remaining song duration or finding beats by looking at the waveform.
Mark only one oval.

	1	2	3	4	5	
strongly disagree	◯	◯	◯	◯	◯	strongly agree

12. **I regularly use cuepoints to help me find certain positions inside a track.**
Mark only one oval.

	1	2	3	4	5	
strongly disagree	◯	◯	◯	◯	◯	strongly agree

13. **The number of cuepoints i use per song is averaging at:**
Mark only one oval.

- ◯ 0
- ◯ 1-2
- ◯ 3-4
- ◯ 5 or more

About DiskPlay

Please read the following statements and choose if you agree or disagree.

14. **I would feel comfortable to perform with a system providing visual aids similar to DiskPlay.**
Mark only one oval.

	1	2	3	4	5	
strongly disagree	◯	◯	◯	◯	◯	strongly agree

Figure A.4: The DiskPlay Acceptance Questionnaire (Page 3)

15. **Systems similiar to DiskPlay provide too much visual help and are ultimately lowering the bar for DJ's.**
Mark only one oval.

	1	2	3	4	5	
strongly disagree	◯	◯	◯	◯	◯	strongly agree

16. **I think systems similiar to DiskPlay should not be used by DJ's.**
Mark only one oval.

	1	2	3	4	5	
strongly disagree	◯	◯	◯	◯	◯	strongly agree

17. **Did you set up the whole DiskPlay system?** *
Did you set it up as it was intended, using a projector and turntables?
Mark only one oval.

◯ Yes *Skip to question 19.*
◯ No *Skip to question 18.*

About DiskPlay

18. **Why did you not set up the DiskPlay system? (multiple answers are possible)**
Check all that apply.

☐ Missing hardware (projector / turntable / ...)
☐ Setup is too complicated
☐ Setup is too time consuming
☐ Other: .....

Skip to question 21.

About DiskPlay

19. **I had the impression that DiskPlay was able to help me navigate songs faster.**
Did the waveform, the cuepoint visualisation or the cuepoint progressbar help you find your target faster?
Mark only one oval.

	1	2	3	4	5	
strongly disagree	◯	◯	◯	◯	◯	strongly agree

Figure A.5: The DiskPlay Acceptance Questionnaire (Page 4)

20. **I had the impression that i spent less time looking for song information on the computer screen when using DiskPlay.**
Mark only one oval.

	1	2	3	4	5	
strongly disagree	◯	◯	◯	◯	◯	strongly agree

Help us to make DiskPlay better

21. **What did you not like about DiskPlay?**
Tell us about things we did wrong, things we should change or remove and problems you had with DiskPlay.

22. **What did you like about DiskPlay?**
Tell us about things we should keep, features you consider useful and how we could make them better.

23. **Would you be willing to provide detailed information about yourself and DiskPlay in a Skype interview?** *
The interview will be short (10-20 minutes). Please be assured that your individual privacy will be maintained, neither your picture nor name or skypename will be published or given to a third party.
Mark only one oval.

◯ Yes *Skip to question 24.*
◯ No *Stop filling out this form.*

Figure A.6: The DiskPlay Acceptance Questionnaire (Page 5)

A.3 Presence Questionnaire

The presence questionnaire in figure A.7 is based on the one by Witmer et al. [1998]. We started with versions with more questions, e.g., "how well were you able to control events", but participants struggled understanding these. In our setup, we simulate real-world behavior and there is no explicit interaction with the environment, thus it is hard to tell what an action is. Furthermore, the original questionnaire was designed to cover navigation through the virtual environment using a joystick for example. These questions regarding the controllers required additional explanation as the one-to-one mapping we used in our experiments (physical location of the listener corresponds to position of the virtual listener) was not perceived as being a controller. Figure A.7 shows the latest iteration as we will use it in future experiments.

Localization-Test Questionnaire

User ID: Age: Sex:

Do you have a hearing disorder?	Yes	No
Do you have problem with spatial hearing?	Yes	No
Do you have experience with Augmented Audio Reality?	Yes	No

Run 1:

	Responsive		**Neutral**		**Not at all**
1. How responsive was the environment to actions that you performed?					
	Natural		**Neutral**		**Artificial**
2. How natural did your interactions with the environment seem?					
	Consistent		**Neutral**		**Not at all**
3. How much did your experience in the virtual environment seem consistent with your real-world experiences?					
	Very Well		**Neutral**		**Not at all**
4. How well could you localize sounds?					
	Quickly		**Neutral**		**Slowly**
5. How quickly did you adjust to the virtual environment experience?					

Figure A.7: The presence questionnaire used for the audio augmented reality evaluations

Bibliography

Alais, David, Burr, David. "The ventriloquist effect results from near-optimal bimodal integration". *Current biology* 14.3 (2004), pp. 257–262.

Andersen, Tue Haste. "Mixxx: towards novel DJ interfaces". *NIME '03*. May 2003, pp. 30–35.

Arons, Barry. "A Review of The Cocktail Party Effect". *Journal of the American Voice I/O Society* 12 (1992), pp. 35–50.

Beamish, Timothy. *A Taxonomy of DJs*. 2001. URL: http://web.archive.org/web/20021016232519/http://www.cs.ubc.ca/~tbeamish/djtaxonomy/introduction.html.

Beamish, Timothy, Maclean, Karon, Fels, Sidney. "Manipulating music: multimodal interaction for DJs". *CHI '04*. Apr. 2004, pp. 327–334.

Bell, Paul. "Interrogating the live: a DJ perspective" (2010).

Berthaut, Florent, Marshall, Mark T, Subramanian, Sriram, Hachet, Martin. "Rouages: Revealing the Mechanisms of Digital Musical Instruments to the Audience." *NIME '13*. 2013, pp. 164–169.

Blauert, Jens. *Spatial Hearing: Psychophysics of Human Sound Localization*. 2nd ed. MIT Press, Oct. 1996.

Bohatsch, Jonas. *Vinyl+*. 2010. URL: http://www.jonasbohatsch.net/.

Brooke, John. "SUS - A quick and dirty usability scale". *Usability evaluation in industry* 189.194 (1996), pp. 4–7.

Brungart, Douglas S, Simpson, Brian D, Kordik, Alexander J. "The detectability of headtracker latency in virtual audio displays". *ICAD '05*. 2005, pp. 37–42.

Burger, Sebastian. "DiskPlay: Evaluation of an Augmented Reality Turntable for Musical Performance". Diploma Thesis. RWTH Aachen University, Apr. 2013.

Buxton, William. "Artists and the Art of the Luthier". *SIGGRAPH '97*. Vol. 31. 1. ACM, 1997, pp. 10–11.

Camurri, Antonio, Canepa, Corrado, Volpe, Gualtiero. "Active listening to a virtual orchestra through an expressive gestural interface: the orchestra explorer". *NIME '07*. 2007, pp. 56–61.

Card, Stuart K, Mackinlay, Jock D, Robertson, George G. "A Morphological Analysis of the Design Space of Input Devices". *ACM Trans. Inf. Syst.* 9.2 (1991), pp. 99–122.

Casiez, Géry, Roussel, Nicolas, Vogel, Daniel. "1€ filter: a simple speed-based low-pass filter for noisy input in interactive systems". *CHI '12*. 2012, pp. 2527–2530.

Dix, Alan, Finlay, Janet, Abowd, Gregory D, Beale, Russell. *Human-computer Interaction*. Pearson/Prentice-Hall, 2004.

Fukuchi, Kentaro. "Multi-track Scratch Player on a Multi-touch Sensing Device". *Entertainment Computing – ICEC '07*. Ed. by Lizhuang Ma, Matthias Rauterberg, Ryohei Nakatsu. Springer Berlin Heidelberg, 2007, pp. 211–218.

Gemperle, F, Kasabach, C, Stivoric, J, Bauer, Malcolm, Martin, R. "Design for Wearability". *ISWC '98*. Oct. 1998, pp. 116–122.

Hansen, Kjetil Falkenberg. "The acoustics and performance of DJ scratching. Analysis and modelling." PhD Thesis. KTH Stockholm, Feb. 2010.

Hansen, Kjetil Falkenberg, Bresin, Roberto. "Mapping strategies in DJ scratching". *NIME '06*. IRCAM — Centre Pompidou, June 2006.

Hansen, Kjetil Falkenberg, Bresin, Roberto. "The skipproof virtual turntable for high-level control of scratching". *Computer Music Journal* 34.2 (June 2010).

Harrison, Chris, Tan, Desney, Morris, Dan. "Skinput: Appropriating the Body As an Input Surface". *CHI '10*. 2010, pp. 453–462.

Heller, Florian. "Corona: Audio AR for historic sites". *AR[t]* 5 (May 2014), pp. 80–85.

Heller, Florian, Borchers, Jan. "AudioScope: Smartphones as Directional Microphones in Mobile Audio Augmented Reality Systems". *CHI '15*. Apr. 2015, pp. 949–952.

Heller, Florian, Borchers, Jan. "AudioTorch: Using a Smartphone As Directional Microphone in Virtual Audio Spaces". *MobileHCI '14*. 2014, pp. 483–488.

Heller, Florian, Borchers, Jan. "DiskPlay: in-track navigation on turntables". *CHI '12*. May 2012.

Heller, Florian, Borchers, Jan. "Visualizing Song Structure on Timecode Vinyls". *NIME '14*. June 2014, pp. 66–69.

Heller, Florian, Ivanov, Stefan, Wacharamanotham, Chat, Borchers, Jan. "FabriTouch: Exploring Flexible Touch Input on Textiles". *ISWC '14*. 2014, pp. 59–62.

Heller, Florian, Jevanesan, Jayan, Dietrich, Pascal, Borchers, Jan. "Where Are We? Evaluating the Current Rendering Fidelity of Mobile Audio Augmented Reality Systems". *MobileHCI '16: Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services*. Sept. 2016.

Heller, Florian, Knott, Thomas, Weiss, Malte, Borchers, Jan. "Multi-user interaction in virtual audio spaces". *CHI EA '09*. Apr. 2009.

Heller, Florian, Krämer, Aaron, Borchers, Jan. "Simplifying Orientation Measurement for Mobile Audio Augmented Reality Applications". *CHI '14*. 2014, pp. 615–624.

Heller, Florian, Lee, Hyun-Young Kriz, Brauner, Philipp, Gries, Thomas, Ziefle, Martina, Borchers, Jan. "An Intuitive Textile Input Controller". *MuC '15*. Berlin: De Gruyter Oldenbourg, Sept. 2015, pp. 263–266.

Holland, Simon, Morse, David R, Gedenryd, Henrik. "AudioGPS: Spatial Audio Navigation with a Minimal Attention Interface". *Pers. and Ubiqu. comp.* 6.4 (Jan. 2002), pp. 253–259.

Holleis, Paul, Schmidt, Albrecht, Paasovaara, Susanna, Puikkonen, Arto, Häkkilä, Jonna. "Evaluating capacitive touch input on clothes". *MobileHCI '08*. Sept. 2008.

Hook, Jonathan, Green, David, McCarthy, John, Taylor, Stuart, Wright, Peter, Olivier, Patrick. "A VJ centered exploration of expressive interaction". *CHI '11*. May 2011.

Hook, Jonathan, Olivier, Patrick. "Waves: multi-touch VJ interface". *ITS '10*. 2010, pp. 305–305.

Hürst, Wolfgang, Götz, Georg, Jarvers, Philipp. "Advanced User Interfaces for Dynamic Video Browsing". *Multimedia '04*. 2004, pp. 742–743.

Ivanov, Stefan. "TextiPad: Implementation and Evaluation of a Wearable Textile Touchpad". Master's thesis. RWTH Aachen University, Sept. 2012.

Jacob, Robert J K, Leggett, John J, Myers, Brad A, Pausch, Randy. "Interaction styles and input/output devices". *Behaviour & Information Technology* 12.2 (1993), pp. 69–79.

Jevanesan, Jayan. "Influence of Elevation on Localization Performance in Mobile Audio Augmented Reality Sys-

tems". Bachelor Thesis. RWTH Aachen University, Oct. 2015.

Kajastila, Raine, Lokki, Tapio. "Eyes-free methods for accessing large auditory menus". *ICAD '10*. 2010, pp. 223–230.

Karrer, Thorsten, Wittenhagen, Moritz, Heller, Florian, Borchers, Jan. "Pinstripe: eyes-free continuous input anywhere on interactive clothing". *UIST EA '10*. Oct. 2010.

Karrer, Thorsten, Wittenhagen, Moritz, Lichtschlag, Leonhard, Heller, Florian, Borchers, Jan. "Pinstripe: eyes-free continuous input on interactive clothing". *CHI '11*. May 2011.

Knott, Thomas. "CORONA - Implementation and Evaluation of Continuous Virtual Audio Spaces for Interactive Exhibits". Diploma Thesis. RWTH Aachen University, 2009.

Kohlrausch, A, Braasch, J, Kolossa, D, Blauert, J. "An Introduction to Binaural Processing". *The Technology of Binaural Listening*. Ed. by Jens Blauert. Springer Berlin Heidelberg, 2013, pp. 1–32.

Komor, Nicholas, Gilliland, S, Clawson, James, Bhardwaj, Manish, Garg, M, Zeagler, Clint, Starner, T. "Is It Gropable? – Assessing the Impact of Mobility on Textile Interfaces". *ISWC '09*. 2009, pp. 71–74.

Krämer, Aaron. "Orientation Measurement for Mobile Audio Augmented Reality Applications". Bachelor Thesis. Aachen: RWTH Aachen University, Apr. 2014.

Kurtenbach, Gordon, Buxton, William. "The Limits of Expert Performance Using Hierarchic Marking Menus". *INTERACT '93 and CHI '93*. 1993, pp. 482–487.

Lauten, Justus. "DiskPlay: An augmented In-Track Navigation Display for Digital Vinyl Systems". Bachelor Thesis. RWTH Aachen University, Sept. 2011.

Linz, Torsten, Kallmayer, Christine, Aschenbrenner, Rolf, Reichl, Herbert. "Embroidering electrical interconnects with conductive yarn for the integration of flexible electronic modules into fabric". *ISWC '05*. Fraunhofer IZM. Oct. 2005, pp. 86–89.

Lippit, TM. "Turntable music in the digital era: designing alternative tools for new turntable expression". *NIME '06*. 2006.

Lissermann, Roman, Huber, Jochen, Hadjakos, Aristotelis, Nanayakkara, Suranga, Mühlhäuser, Max. "EarPut: Augmenting Ear-worn Devices for Ear-based Interaction". *OZCHI '14*. 2014, pp. 300–307.

Loomis, Jack M, Hebert, Chick, Cicinelli, Joseph G. "Active localization of virtual sounds". *J. Acoust. Soc. Am.* 88 (1990), p. 1757.

Lopes, Pedro, Ferreira, Alfredo, Pereira, J A Madeiras. "Battle of the DJs: an HCI perspective of Traditional and Virtual and Hybrid and Multitouch DJing". *NIME '11*. Oslo, Norway, 2011.

Lyons, K, Gandy, M, Starner, T. "Guided by voices: An audio augmented reality system". *ICAD '03*. Citeseer, 2000.

Mackay, Wendy E. "Augmented reality: linking real and virtual worlds: a new paradigm for interacting with computers". *AVI '98*. 1998, pp. 13–21.

Majdak, Piotr, Walder, Thomas, Laback, Bernhard. "Effect of long-term training on sound localization performance with spectrally warped and band-limited head-related transfer functions". *J. Acoust. Soc. Am.* 134.3 (2013), pp. 2148–2159.

Marculescu, D, Marculescu, R, Zamora, N H, Stanley-Marbell, P, Khosla, P K, Park, S, Jayaraman, S, Jung, S, Lauterbach, C, Weber, Werner, Kirstein, T, Cottet, D, Grzyb, J, Troster, G, Jones, M, Martin, T, Nakad, Z. "Electronic textiles: a platform for pervasive computing". *Proc. IEEE* 91.12 (Dec. 2003), pp. 1995–2018.

Marentakis, Georgios N, Brewster, Stephen A. "Effects of feedback, mobility and index of difficulty on deictic spatial audio target acquisition in the horizontal plane". *CHI '06*. 2006, pp. 359–368.

Mariette, Nicholas. "Navigation Performance Effects of Render Method and Head-Turn Latency in Mobile Audio Augmented Reality". *ICAD '09*. Ed. by Sølvi Ystad, Mitsuko Aramaki, Richard Kronland-Martinet, Kristoffer Jensen. Springer Berlin Heidelberg, 2010, pp. 239–265.

Martin, A, Hurford, R, McCann, J, Smith, D. "Designing Wearables: A Brief Discussion". *Proc The Role of Design in ...* 2007.

McCann, J, Hurford, R, Martin. "A design process for the development of innovative smart clothing that addresses end-user needs from technical, functional, aesthetic and cultural view points". *ISWC '05*. 2005, pp. 70–77.

McGookin, David, Priego, Pablo. "Audio Bubbles: Employing Non-speech Audio to Support Tourist Wayfinding". *Haptic and Audio Interaction Design*. Springer Berlin Heidelberg, 2009, pp. 41–50.

Middlebrooks, John C. "Virtual localization improved by scaling nonindividualized external-ear transfer functions in frequency". *J. Acoust. Soc. Am.* 106.3 (1999), pp. 1493–1510.

Middlebrooks, John C, Green, David M. "Sound localization by human listeners". *Annual review of psychology* 42.1 (1991), pp. 135–159.

Mujibiya, Adiyan, Cao, Xiang, Tan, Desney S, Morris, Dan, Patel, Shwetak N., Rekimoto, Jun. "The Sound of Touch: On-body Touch and Gesture Sensing Based on Transdermal Ultrasound Propagation". *ITS '13*. 2013, pp. 189–198.

O'Donnell, Richard. "Prolog to Electronic textiles: a platform for pervasive computing". *Proc. IEEE* 91.12 (Dec. 2003), pp. 1993–1994.

O'Sullivan, Dan, Igoe, Tom. *Physical computing: sensing and controlling the physical world with computers*. Course Technology Press, Jan. 2004.

Pabst, Andreas, Walk, Roger. "Augmenting a rugged standard DJ turntable with a tangible interface for music browsing and playback manipulation". *IE '07*. Ulm, Germany, 2007, pp. 533–535.

Paterson, Natasa, Naliuka, Katsiaryna, Jensen, Soren Kristian, Carrigy, Tara, Haahr, Mads, Conway, Fionnuala. "Design, implementation and evaluation of audio for a location aware augmented reality game". *Fun and Games '10*. Sept. 2010.

Rantanen, J, Impiö, J, Karinsalo, T, Malmivaara, M, Reho, A, Tasanen, M, Vanhala, J. "Smart Clothing Prototype for the Arctic Environment". *Pers. and Ubiqu. comp.* 6.1 (2002), pp. 3–16.

Rayleigh O.M. Pres. R.S., Lord. "XII. On our perception of sound direction". *Philosophical Magazine*. 6th ser. 13.74 (1907), pp. 214–232.

Reeves, Stuart, Benford, Steve, O'Malley, Claire, Fraser, Mike. "Designing the spectator experience". *CHI '05*. 2005, pp. 741–750.

Reinecke, Katharina, Yeh, Tom, Miratrix, Luke, Mardiko, Rahmatri, Zhao, Yuechen, Liu, Jenny, Gajos, Krzysztof Z. "Predicting Users' First Impressions of Website Aesthetics with a Quantification of Perceived Visual Complexity and Colorfulness". *CHI '13*. 2013, pp. 2049–2058.

Rekimoto, J. "GestureWrist and GesturePad: unobtrusive wearable interaction devices". *ISWC '01*. 2001, pp. 21–27.

Sander, Christian, Wefers, Frank, Leckschat, Dieter. "Scalable Binaural Synthesis on Mobile Devices". *Audio Engineering Society Convention 133*. Oct. 2012.

Saponas, T Scott, Harrison, Chris, Benko, Hrvoje. "PocketTouch: through-fabric capacitive touch input". *UIST '11*. Oct. 2011.

Sauro, Jeff. *Measuring Usability with the System Usability Scale (SUS)*. Feb. 2011. URL: http://www.measuringu.com/sus.php.

Schlager, N. *How products are made: an illustrated guide to product manufacturing*. How Products are Made: An Illustrated Guide to Product Manufacturing. Gale Research Inc., 1994.

Schwarz, Julia, Harrison, Chris, Hudson, Scott, Mankoff, Jennifer. "Cord input: an intuitive, high-accuracy, multi-degree-of-freedom input method for mobile devices". *CHI '10*. Apr. 2010.

Shaw, EAG. "External ear response and sound localization". *Localization of sound: Theory and applications* (1982), pp. 30–41.

Stahl, Christoph. "The roaring navigator: a group guide for the zoo with shared auditory landmark display". *MobileHCI '07*. Sept. 2007.

Terrenghi, Lucia, Zimmermann, Andreas. "Tailored audio augmented environments for museums". *IUI '04*. Jan. 2004.

Thar, Jan. "Pinstripe: Integration & Evaluation of a Wearable Linear Input Controller for Everyday Clothing". Bachelor Thesis. RWTH Aachen University, Feb. 2013.

The Serato Face. 2013. URL: http://seratoface.tumblr.com/.

Thomas, B, Grimmer, K, Zucco, J, Milanese, S. "Where Does the Mouse Go? An Investigation into the Placement of a Body-Attached TouchPad Mouse for Wearable Computers". *Pers. and Ubiqu. comp.* 6.2 (Jan. 2002), pp. 97–112.

Toney, Aaron, Mulley, Barrie, Thomas, Bruce H, Piekarski, Wayne. "Social Weight: Designing to Minimise the Social Consequences Arising from Technology Use by the Mobile Professional". *Pers. and Ubiqu. comp.* 7.5 (Oct. 2003), pp. 309–320.

Tran, Tuyen V, Letowski, Tomasz, Abouchacra, Kim S. "Evaluation of acoustic beacon characteristics for navigation tasks". *Ergonomics* 43.6 (2000), pp. 807–827.

Valli, Alessandro. "The design of natural interaction". *Multimedia Tools and Applications* 38.3 (2008), pp. 295–305.

Vazquez-Alvarez, Yolanda, Oakley, Ian, Brewster, Stephen A. "Auditory display design for exploration in mobile audio-augmented reality". *Pers. and Ubiqu. comp.* 16.8 (2012), pp. 987–999.

Villar, Nicolas, Gellersen, Hans, Jervis, Matt, Lang, Alexander. "The ColorDex DJ system: a new interface for live music mixing". *NIME '07*. June 2007.

Vorländer, Michael. *Auralization: Fundamentals of Acoustics, Modelling, Simulation, Algorithms and Acoustic Virtual Reality*. 1st. Springer, 2007.

Wagner, Julie, Nancel, Mathieu, Gustafson, Sean G, Huot, Stéphane, Mackay, Wendy E. "Body-centric design space for multi-surface interaction". *CHI '13*. Apr. 2013.

Walker, Bruce N, Lindsay, Jeffrey. "Navigation Performance With a Virtual Auditory Display: Effects of Beacon Sound, Capture Radius, and Practice". *Human Factors: The Journal of the Human Factors and Ergonomics Society* 48.2 (2006), pp. 265–278.

Wardle, S A. "Position and velocity transducer using a phonograph disc and turntable". Sept. 2007.

Weigel, Martin, Lu, Tong, Bailly, Gilles, Oulasvirta, Antti, Majidi, Carmel, Steimle, Jürgen. "iSkin: Flexible, Stretchable and Visually Customizable On-Body Touch Sensors for Mobile Computing". *CHI '15*. 2015, pp. 2991–3000.

Wenzel, Elizabeth M, Arruda, Marianne, Kistler, Doris J, Wightman, Frederic L. "Localization using nonindividualized head-related transfer functions". *J. Acoust. Soc. Am.* 94.1 (1993), pp. 111–123.

Wenzel, Elizabeth M, Wightman, Frederic L, Kistler, Doris J. "Localization with non-individualized virtual acoustic display cues". *CHI '91*. 1991, pp. 351–359.

Wightman, Frederic L, Kistler, Doris J. "Headphone simulation of free-field listening. I: Stimulus synthesis". *J. Acoust. Soc. Am.* 85.2 (1989), pp. 858–867.

Witmer, Bob G, Singer, Michael J. "Measuring Presence in Virtual Environments: A Presence Questionnaire". *Presence: Teleoper. Virtual Environ.* 7.3 (1998), pp. 225–240.

Zhao, Shengdong, Dragicevic, Pierre, Chignell, Mark, Balakrishnan, Ravin, Baudisch, Patrick. "Earpod: eyes-free menu selection using touch input and reactive audio feedback". *CHI '07*. Apr. 2007.

Zotkin, Dmitry N, Duraiswami, Ramani, Davis, Larry S. "Rendering localized spatial audio in a virtual auditory space". *IEEE Transactions on Multimedia* 6.4 (2004), pp. 553–564.

Index

AAR ..... *see* Audio augmented reality
Attigo TT 58
Audio augmented reality 80

Beatmatching 57
Beatmix DJ 56
Beatmixing 57

Caplinq 41
CD 15
Cocktail Party Effect 76
Conveyor belt 58, 59, 65
Corona 82
Cue points 62, 66

Digital Vinyl System 53
DJ set 56
DJing 56–57
Durability 20
DVS ..... *see* Digital Vinyl System
DVS Setup 55

Fabritouch 39, 40
Fashion compatibility 19
Flexible PCB 32
Future work 118–119

Head-Related Transfer Function 76, 78, 98
Head-yaw 89
HRTF ..... *see* Head-related transfer function

ILD ..... *see* Interaural level difference
IMU ..... *see* Inertial measurement unit
Inertial measurement unit 98, 106
Intelligent Headset 106
Interactive clothing 18
Interaural level difference 77
Interaural time difference 77
Intuitex 36

Involuntary activation 21
iPod 15, 17
ITD . . . . . *see* Interaural time difference

Lead-in 52
Lead-out 52

MAARS . . . . . *see* Mobile audio augmented reality systems
Mobile Audio Augmented Reality Systems 105
MP3 15

Native Instruments Traktor Scratch 54

OpenAL 85, 106
Orbit 62

Piezoresistive foil 40, 41
Pinstripe 19

Quartz Composer 67

Random access media 15

Scratch DJ 56
Scratch Live 54
Scratching 53
Serato Scratch Live 54
Slipmat 56
Smartphone 16
Spatial hearing 77

Tape 14
Textile touchpads 40
Timecode record 54
Touchpad architecture 41
Traktor Scratch 54
Turntable 51

Ubisense 106
Ultra wide band 106

Ventriloquist effect 79
Vinyl record 52

Walkman 14
Wearability 19

Own Publications

Papers (Peer-reviewed, archival)

Hamdan, Nur Al-huda, Blum, Jeffrey, **Heller, Florian**, Kosuru, Ravi Kanth, Borchers, Jan. "Grabbing at an Angle: Menu Selection for Fabric Interfaces". *ISWC '16: Proceedings of the 2016 International Symposium on Wearable Computers*. Sept. 2016.

Heller, Florian, Borchers, Jan. "AudioScope: Smartphones as Directional Microphones in Mobile Audio Augmented Reality Systems". *CHI '15: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. Seoul, Republic of Korea, Apr. 2015.

Heller, Florian, Borchers, Jan. "Diskplay: In-Track Navigation on Turntables". *CHI '12: Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems*. May 2012.

Heller, Florian, Borchers, Jan. "Visualizing Song Structure on Timecode Vinyls". *NIME '14: Proceedings of the International Conference on New Interfaces for Musical Expression*. June 2014.

Heller, Florian, Ivanov, Stefan, Wacharamanotham, Chat, Borchers, Jan. "FabriTouch: Exploring Flexible Touch Input on Textiles". *ISWC '14: Proceedings of the 2014 ACM International Symposium on Wearable Computers*. Sept. 2014.

Heller, Florian, Jevanesan, Jayan, Dietrich, Pascal, Borchers, Jan. "Where Are We? Evaluating the Current Rendering Fidelity of Mobile Audio Augmented Reality Systems". *MobileHCI '16: Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services*. Sept. 2016.

Heller, Florian, Krämer, Aaron, Borchers, Jan. "Simplifying Orientation Measurement for Mobile Audio Augmented Reality Applications". *CHI '14: Proceedings of the 2014 ACM annual conference on Human Factors in Computing Systems*. Apr. 2014.

Karrer, Thorsten, Wittenhagen, Moritz, Lichtschlag, Leonhard, **Heller, Florian**, Borchers, Jan. "Pinstripe: Eyes-free Continuous Input on Interactive Clothing". *CHI '11*. May 2011.

Wacker, Philipp, Kreutz, Kerstin, **Heller, Florian**, Borchers, Jan. "Maps and Location: Acceptance of Modern Interaction Techniques for Audio Guides". *CHI '16*. May 2016.

Posters (Peer-reviewed, non-archival)

Hamdan, Nur Al-huda, **Heller, Florian**, Wacharamanotham, Chat, Thar, Jan, Borchers, Jan. "Grabrics: A Foldable Two-Dimensional Textile Input Controller". *CHI EA '16: Extended Abstracts of the SIGCHI Conference on Human Factors in Computing Systems*. May 2016.

Heller, Florian, Borchers, Jan. "AudioTorch: Using a Smartphone as Directional Microphone in Virtual Audio Spaces". *MobileHCI EA '14: Extended Abstracts of the 16th International Conference on Human-Computer Interaction with Mobile Devices & Services*. Sept. 2014.

Heller, Florian, Borchers, Jan. "PowerSocket: Towards On-Outlet Power Consumption Visualization". *CHI '11: Extended Abstracts of the CHI 2011 Conference on Human Factors in Computing Systems*. 2011.

Heller, Florian, Knott, Thomas, Weiss, Malte, Borchers, Jan. "Multi-User Interaction in Virtual Audio Spaces". *CHI EA '09: Extended Abstracts of the SIGCHI Conference on Human Factors in Computing Systems*. 2009.

Heller, Florian, Lee, Hyun-Young Kriz, Brauner, Philipp, Gries, Thomas, Ziefle, Martina, Borchers, Jan. "An Intuitive Textile Input Controller". *MuC '15: Mensch und Computer 2015 – Proceedings*. Sept. 2015.

Heller, Florian, Lichtschlag, Leonhard, Wittenhagen, Moritz, Karrer, Thorsten, Borchers, Jan. "Me Hates This: Exploring Different Levels of User Feedback for (Usability) Bug Reporting". *CHI EA '11: Ex- tended Abstracts of the SIGCHI Conference on Human Factors in Computing Systems*. 2011.

Heller, Florian, Oßmann, Lukas, Hamdan, Nur Al-huda, Brauner, Philipp, Heek, Julia Van, Scheulen, Klaus, Goßen, Laura, Witsch, Rouven, Möllering, Christian, Gries, Thomas, Ziefle, Martina, Borchers, Jan. "Gardeene! Textile Controls for the Home Environment". *MuC '16: Mensch und Computer 2016 – Proceedings*. Berlin, Sept. 2016.

Heller, Florian, Tsoleridis, Konstantinos, Borchers, Jan. "Counter Entropy: Visualizing Power Consumption in an Energy+ House". *CHI EA '13: Extended Abstracts*

of the 2013 ACM annual conference on Human Factors in Computing Systems. Apr. 2013.

Heller, Florian, Voelker, Simon, Wacharamanotham, Chat, Borchers, Jan. "Transporters: Vision & Touch Transitive Widgets for Capacitive Screens". *CHI EA '15: Extended Abstracts of the SIGCHI Conference on Human Factors in Computing Systems*. Apr. 2015.

Herkenrath, Gero, Huch, Carl, **Heller, Florian**, Borchers, Jan. "Geo-Sociograms: A Method to Analyze Movement Patterns and Characterize Tasks in Location-Based Multiplayer Games". *CHI EA '14: Extended Abstracts of the 2014 ACM Conference on Human Factors in Computing Systems*. Apr. 2014.

Karrer, Thorsten, Wittenhagen, Moritz, **Heller, Florian**, Borchers, Jan. "Pinstripe: Eyes-free Continuous Input Anywhere on Interactive Clothing". *UIST '10 Adjunct proceedings*. Oct. 2010.

Workshop papers (Juried, non-archival)

Heller, Florian, Borchers, Jan. "Corona: Audio Augmented Reality in Historic Sites". *MobileHCI 2011 Workshop on Mobile Augmented Reality: Design Issues and Opportunities*. Ed. by Sa, Marco de, Elizabeth F. Churchill, Katherine Isbister. Stockholm, Sweden, Aug. 2011.

Heller, Florian, Borchers, Jan. "Physical Interaction with Audio". *CHI '15 Workshop on Collaborating with Intelligent Machines*. Apr. 2015.

Magazine articles (Edited, non-archival)

Heller, Florian. "CORONA: Audio AR for historic sites". *AR[t] - Augmented Reality, Art and Technology* 5 (May 2014), pp. 80–85.

Heller, Florian, Borchers, Jan. "Physical prototyping of an on-outlet power-consumption display". *interactions* 19 (1 Jan. 2012), pp. 14–17.