

LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN
Department "Institut für Informatik"
Lehr- und Forschungseinheit Medieninformatik
Prof. Dr. Heinrich Hußmann

Masterarbeit

Guiding Multiple People in Front of Public Displays

Christian Becker
beckerc@cip.ifi.lmu.de

Bearbeitungszeitraum: 15.07.2015 bis 12.07.2016
Betreuer: Mohamed Khamis
Verantw. Hochschullehrer: Prof. Dr. Florian Alt

Danksagung

An dieser Stelle möchte ich mich ganz herzlich bei all denjenigen bedanken, die mich während der Anfertigung dieser Masterarbeit unterstützt und motiviert haben.

Ganz herzlich bedanke ich mich bei Mohamed Khamis, der meine Masterarbeit betreute und mir stets mit hilfreichen Anregungen und konstruktiver Kritik zur Seite stand.

Ebenso gilt mein Dank Prof. Dr. Florian Alt für seine Unterstützung und die vielen Ratschläge.

Bedanken möchte ich mich auch bei Dr. Andreas Bulling (Max-Planck Institut für Informatik der Universität des Saarlandes) der mir beratend zur Seite stand.

Ein besonderer Dank gilt auch Carina Hoppenz für die Unterstützung bei der Nutzung des linearen gemischten Modells.

Darüber hinaus bedanke ich mich ganz herzlich bei meinen Eltern und Freunden, die mich auf diesem Weg mit viel Geduld und Verständnis begleiteten und unterstützten.

Zusammenfassung

Public Displays bieten im öffentlichen Raum Zugang zu Unterhaltung, Informationen und Dienstleistungen. Ihre Verbreitung hat, überwiegend in Großstädten, in den letzten Jahren immer weiter zugenommen. Vor allem jene Systeme, die potentiellen Nutzern das gezielte Betrachten von Inhalten und oft auch deren Mitnahme durch Interaktion ermöglichen, stehen im Fokus der Forschung. Ebenso stellen Bildschirm-Blindheit, Interaktions-Blindheit und das Lotsen mehrerer potentieller Nutzer in den Sweetspot Herausforderungen dar, die es zu lösen gilt.

Wie verschiedene Arbeiten gezeigt haben, sind Nutzerrepräsentationen ein vielversprechendes Werkzeug, um auf öffentliche Bildschirme aufmerksam zu machen und deren Interaktivität zu vermitteln. Möglicherweise eignen sie sich ebenfalls gut dazu, potentielle Nutzer in den Sweetspot zu führen. Um dies herausfinden zu können, muss jedoch zunächst geklärt werden, wie gut potentielle Nutzer zwischen verschiedenen Nutzerrepräsentationen unterscheiden können. Dazu muss untersucht werden, welche Strategien zur Erkennung von Nutzerrepräsentationen angewandt werden und wie gut sich diese zur Erkennung der unterschiedlichen Repräsentationen eignen. Die Frage ist außerdem, ob und falls ja, inwiefern Repräsentationen auch dann erkannt werden, wenn gleichzeitig mehrere gleichartige Repräsentationen angezeigt werden. Um Lösungen für oben genannte und weitere Problemstellungen zu finden, wurden im Rahmen dieser Arbeit zwei Studien durchgeführt. Dabei wurden vier Nutzerrepräsentationen miteinander verglichen: das Spiegelbild, die Silhouette, das Skelett und als abstraktes Objekt das Quadrat. Aufbauend auf den Ergebnissen dieser Studien wurde ein Leitfaden entwickelt, welcher Anwendungsentwicklern und -designern dabei behilflich sein soll, die passende Nutzerrepräsentation für das jeweilige Anwendungsszenario zu finden.

Abstract

Public displays provide access to entertainment, information and services in public space. Their number increased in recent years, especially in major cities. Research mostly deals with interactive systems. These allow users to navigate through content and to take content away for later use. Display blindness, interaction blindness and guiding of multiple users to the sweetspot are challenges to get solved. As shown in previous work, user representations are a promising tool for drawing attention to and indicating interactivity of public displays. They might also be suitable for guiding users to the sweetspot. For proving this it must be examined to which degree users are able to differentiate between several user representations. Therefore it should be clarified which strategies are used to recognize these kind of representations and to what extent the strategies can be used to recognize different types of representations. The question also is, whether user representations are recognized if multiple representations of the same kind are shown at the same time. This work is dedicated to find solutions for the challenges mentioned above. Hence two studies were conducted comparing four user representations: the mirror image, the silhouette, the skeleton and an abstract object, the square. Based on the results there have been developed guidelines for application developers and designers to support finding the right user representation for the given application scenario.

Aufgabenstellung

As public displays become more ubiquitous and are offering interactive experiences to multiple passersby in parallel, there is a need to guide these passersby to the interaction “sweet spots”. However, while guiding a single passerby is straightforward, guiding multiple ones requires giving different instructions to each passerby. A prerequisite would be to assign each passerby to a representation on the display.

Scope of the thesis

The aim of this thesis is thus to study how well user representations are distinguishable on public displays for multiple users. Existing research have utilized different user representations to represent a single user on the display. However the use of a representation with multiple users and understanding how well users can distinguish themselves have not been studied before.

Tasks

- Review state of the art in user representations on public displays
- Identify the most promising user representations for public displays
- Identifying strategies employed by users in order to distinguish themselves when using the different representations on public displays
- Implementing a prototype application that can be used as a platform for testing the representation types analyzing the data and reporting the results
- Written thesis and presentation of the work

Ich erkläre hiermit, dass ich die vorliegende Arbeit selbstständig angefertigt, alle Zitate als solche kenntlich gemacht sowie alle benutzten Quellen und Hilfsmittel angegeben habe.

München, 10. Juli 2016

.....

Inhaltsverzeichnis

1	Einleitung	3
1.1	Motivation	3
1.2	Ziele dieser Arbeit	3
1.3	Aufbau dieser Arbeit	5
2	Verwandte Literatur	6
2.1	Definition des Begriffs Public Display	6
2.2	Gründe für Interaktion	7
2.3	Arten von Public Displays und historische Entwicklung	8
2.4	Interaktionstechniken	13
2.5	Sweetspot und Positionierung des Nutzers	17
2.6	Nutzerrepräsentationen	18
3	Implementierung	23
3.1	Aufbau	23
3.2	Kinect	23
3.2.1	Aufbau und Spezifikationen	23
3.2.2	Unterschiede zur ersten Generation	24
3.2.3	Programmierung mittels Kinect	25
3.3	Software-Prototyp	27
3.3.1	Herausforderungen	30
4	Vorstudie	33
4.1	Ziele	33
4.2	Nutzerrepräsentationen	33
4.3	Aufbau	34
4.4	Ablauf	35
4.5	Ergebnisse	35
4.5.1	Teilnehmer	36
4.5.2	Bewegungs-Strategien	36
4.5.3	Positions-Strategie	37
5	Hauptstudie	39
5.1	Ziele	39
5.2	Aufnahmen	39
5.3	Aufbau	40
5.4	Ablauf	41
5.5	Auswertung	43
5.5.1	Vorgehensweise	43
5.5.2	Ergebnisse	46
6	Diskussion	57
7	Leitfaden zur Implementierung von Anwendungen	59
8	Fazit	61
8.1	Zusammenfassung	61
8.2	Ausblick	62

1 Einleitung

Dieses Kapitel dient als Einführung in den Bereich Public Displays. Daher wird auf unterschiedliche Aspekte des Themenkomplexes eingegangen. Dazu gehören unter anderem eine Begriffsdefinition sowie eine historische Einordnung. Zudem werden verschiedene Problemstellungen, vor allem in Bezug auf Nutzerverhalten und Aufmerksamkeit, angesprochen, deren Lösung die Motivation darstellte, die zur Entstehung dieser Arbeit beitrug. Schließlich werden die Ziele der Arbeit definiert und es wird aufgezeigt, wie die vorliegende Arbeit im Detail aufgebaut ist.

1.1 Motivation

Public Displays haben in den letzten Jahren immer mehr Aufmerksamkeit auf sich gezogen und das Interesse von Wissenschaft und Industrie gleichermaßen geweckt. Seit man das Potential solcher digitaler Anzeigen erkannte, haben Verbreitung sowie Forschung in diesem Bereich stark zugenommen.

Katalysatoren, wie die in den letzten Jahren immer weiter gesunkenen Produktionskosten für LCDs [49], haben dazu geführt, dass Public Displays in vielen Innenstädten und an Flughäfen beziehungsweise Bahnhöfen zu sehen sind. Dort werden sie hauptsächlich zu Informations-, Werbe- und Unterhaltungszwecken eingesetzt. Es wird erwartet, dass ihre Verbreitung durch neuartige Technologien wie organische Leuchtdioden (OLED) oder elektronisches Papier in den nächsten Jahren weiter zunimmt.

Immer wieder werden auch solche Systeme entwickelt, die über die Möglichkeit der Interaktion verfügen und dadurch zusätzliche Mehrwerte bieten. Doch sind sich viele potentielle Nutzer der Interaktivität und Interaktionsmöglichkeiten oft nicht bewusst. Hinzu kommt, dass die Interaktion oft nur unter bestimmten Bedingungen funktioniert. Beispielsweise kann es erforderlich sein, dass ein festgelegter Abstand zum Display eingehalten wird oder sich Nutzer in einen bestimmten Punkt beziehungsweise Bereich, den sogenannten Sweetspot begeben müssen, damit die Interaktion optimal funktioniert.

Es gilt also, Nutzer auf entsprechende Funktionen aufmerksam zu machen und sie, sofern dies erforderlich ist, in den Sweetspot zu lotsen. Diese Aspekte stellen jedoch nach wie vor große Herausforderungen dar. Mit der Zeit wurden diesbezüglich unterschiedliche Ansätze entwickelt. Einer der vielversprechendsten Ansätze sieht vor, auf dem Display unterschiedliche Arten von Nutzerrepräsentationen anzuzeigen. Beispielsweise ist es möglich, unter Zuhilfenahme spezieller Kameras wie Kinect, den Körper eines Passanten als gehendes Skelett oder Silhouette darzustellen und so dessen Aufmerksamkeit zu erregen.

Daraus ergeben sich jedoch mehrere Fragen: Wie gut können diese Arten von Darstellungen durch potentielle Nutzer überhaupt erkannt und voneinander unterschieden werden? Welche Strategien (bewusst oder unbewusst) werden angewandt, um sie zu erkennen und wie gut eignen sich diese zur Erkennung? Kann möglicherweise eine besondere Eignung unterschiedlicher Darstellungen für unterschiedliche Anwendungsszenarien festgestellt werden? Diese und weitere Fragen waren Kern dieser Arbeit. In Kapitel 1.2 wird genauer darauf eingegangen.

1.2 Ziele dieser Arbeit

Ziel dieser Arbeit war es herauszufinden, welche Strategien Nutzer anwenden um Repräsentationen von sich selbst zu erkennen und wie gut sich diese Repräsentationen in Bezug auf Erkennungsgeschwindigkeit und Fehlerrate unter Anwendung der verschiedenen Strategien erkennen lassen. Dazu wurden die Repräsentationsformen Spiegelbild, Silhouette, Skelett und als abstrak-

tes Objekt das Quadrat miteinander verglichen. Ein Vergleich in dieser Zusammensetzung fand in anderen Arbeiten bisher noch nicht statt.

Von großer Relevanz war außerdem, zu untersuchen, inwiefern sich mehrere gleichartige Repräsentationen, die gleichzeitig angezeigt werden, auf die Erkennung der eigenen Repräsentation auswirken. Daher erfolgte der Vergleich der Repräsentationsformen immer mit jeweils 7 Teilnehmern beziehungsweise Repräsentationen gleichzeitig.

Des Weiteren sollte in Erfahrung gebracht werden, wo die Vor- und Nachteile der jeweiligen Darstellungen liegen. Aufbauend auf den Vor- und Nachteilen sowie den Ergebnissen zweier Studien galt es einen Leitfaden für Anwendungsentwickler und -designer zu erstellen. Dieser soll dabei helfen, die passende Nutzerrepräsentation für das jeweilige Anwendungsszenario zu finden.

1.3 Aufbau dieser Arbeit

Wie aus dem Inhaltsverzeichnis ersichtlich gliedert sich diese Arbeit, inklusive Einleitung, in folgende 8 Kapitel:

1. Einleitung
2. Verwandte Literatur
3. Implementierung
4. Vorstudie
5. Hauptstudie
6. Diskussion
7. Leitfaden zur Implementierung von Anwendungen
8. Fazit

Der Aufbau beziehungsweise weitere Verlauf der Arbeit wird in den folgenden Abschnitten skizziert.

Das 2. Kapitel befasst sich mit grundlegenden Aspekten des Themenbereichs Public Displays. Es wird beispielsweise geklärt, was Public Displays sind, welche Arten es gibt und welche Techniken zur Interaktion entwickelt wurden. Auch wird der Frage nachgegangen, wie potentielle Nutzer auf entsprechende Bildschirme aufmerksam gemacht und wie ihnen die Interaktivität von Public Displays vermittelt werden kann. Hierbei stehen die sogenannten Nutzerrepräsentationen im Mittelpunkt.

Im 3. Kapitel folgt eine Beschreibung der Implementierung, die notwendig war, um die in Kapitel 4 und 5 angesprochenen Studien durchführen zu können. Dabei wird unter anderem auf die Herangehensweise und etwaige Probleme, die es während der Implementierung zu lösen galt, eingegangen. Die Software-Entwicklung mit Kinect ist ebenfalls ein großer Bestandteil der Implementierung.

Die Kapitel 4 und 5 beschäftigen sich mit der Durchführung zweier Studien, welche unter anderem zum Ziel hatten, vier Nutzerrepräsentationen (Spiegelbild, Silhouette, Skelett und Quadrat) im Hinblick auf Erkennungsgeschwindigkeit und Fehlerquote miteinander zu vergleichen. Außerdem sollte herausgefunden werden, unter Zuhilfenahme welcher Strategien, Nutzer entsprechende Repräsentationen identifizieren. Weshalb die genannten Repräsentationsformen ausgewählt wurden und welche Eigenschaften sie aufweisen wird in Kapitel 4.2 beschrieben.

Das darauf folgende 6. Kapitel greift die Ergebnisse der Kapitel 4 und 5 auf, diskutiert diese und versucht Antworten auf eventuell damit verbundene Fragen zu liefern.

Es schließt sich Kapitel 7 an. Dieses dient der Präsentation eines Leitfadens für Anwendungsentwickler und -designer. Der Leitfaden soll Auskunft darüber geben, inwiefern sich die in der Arbeit besprochenen Nutzerrepräsentationen für unterschiedliche Verwendungszwecke beziehungsweise Anwendungsszenarien eignen und entsprechende Empfehlungen aussprechen.



Abbildung 2.1: Public Display in der Innenstadt von Oulu, Finnland [102, 55].

Die Arbeit endet mit dem 8. Kapitel, der Zusammenfassung. Diese beschreibt in einer gekürzten Form nochmals den Aufbau beziehungsweise Ablauf der Arbeit und hebt rückblickend die wichtigsten Aspekte und Ergebnisse hervor. Außerdem wird erläutert, inwiefern bestimmte Fragestellungen in nachfolgenden Arbeiten untersucht werden könnten.

2 Verwandte Literatur

Im nun folgenden Unterkapitel wird zunächst auf grundlegende Aspekte des Themenbereichs Public Displays eingegangen. Beispielsweise erfolgt eine Begriffsdefinition sowie eine historische Einordnung. Die Abschnitte 2.5 und 2.6 befassen sich näher mit den in der Einleitung bereits beschriebenen Problemstellungen. Zum Beispiel wird darauf eingegangen, welche Möglichkeiten es gibt, potentielle Nutzer auf Public Displays aufmerksam zu machen und diese in den Sweetspot zu leiten. Dabei stehen die sogenannten Nutzerrepräsentationen im Mittelpunkt. Anhand verwandter Arbeiten wird unter anderem untersucht, welche dieser Nutzerrepräsentationen bereits Verwendung fanden und inwiefern sich diese bewährt haben.

2.1 Definition des Begriffs Public Display

Als Public Displays bezeichnet man im öffentlichen Raum aufgebaute Bildschirme beziehungsweise digitale Anzeigen, die heutzutage in der Regel zur Darstellung von Werbung oder Informationen genutzt werden. Außerdem werden sie häufig genutzt, um verschiedene Arten von Dienstleistungen anzubieten, beispielsweise in Form von Bank- und -Fahrkartenautomaten oder interaktiven Wegweisern. Sie sind häufig an Flughäfen und Bahnhöfen anzutreffen oder auch in Innenstädten.

Anordnung, Ausrichtung und Größe dieser Bildschirme können sehr stark variieren. Beispielsweise gibt es Systeme, die aus einem einzelnen horizontal ausgerichteten Bildschirm bestehen (siehe Abbildung 2.1), aber auch jene, wie CityWall [57] oder ChainedDisplays [70], die mehrere horizontal oder vertikal ausgerichtete Bildschirme miteinander verbinden, um so eine möglichst große Interaktionsfläche zu schaffen und dadurch die gleichzeitige Interaktion mehrerer Nutzer zu ermöglichen beziehungsweise zu fördern. Das Resultat muss dabei nicht zwingend ein großer, planarer Bildschirm sein; im Gegenteil: wie im Fall von ChainedDisplays [70] sind auch konkave und hexagonale Konfigurationen möglich. Abbildung 2.2 zeigt diese drei Konfigurationen.



Abbildung 2.2: Unterschiedliche Konfigurationen derselben Bildschirmmenge: Planar/Flach (1), hexagonal (2) und konkav (3)[70].

Neben äußerlichen Unterschieden gibt es vor allem auch solche in Bezug auf die Art der Interaktion. Public Displays, die mit einem Touchscreen versehen sind, reagieren beispielsweise auf Berührung oder Druck [57, 31]. Andere Systeme wiederum setzen auf die Ausführung von Gesten. Die Erkennung dieser erfolgt dabei in der Regel entweder über eine Kamera mit Tiefensensor [70, 50, 56, 72, 48], die Beschleunigungssensoren eines Smartphones [11, 21] oder dessen Kamera [5, 33, 42, 61]. Auch gibt es Systeme, die auf Blick-Interaktion¹ basieren [77, 76, 41]. Das Kapitel 2.4 setzt sich mit einigen Interaktionsmethoden genauer auseinander.

Die Frage ist jedoch nicht nur, **wie** man mit Public Displays interagieren kann, sondern auch, **weshalb** man mit diesen überhaupt interagieren können sollte und welche Vorteile die Interaktion mit sich bringt. Das nun folgende Kapitel geht genau darauf ein.

2.2 Gründe für Interaktion

Basierend auf der Arbeit von Davies, Clinch und Alt [20] können vier wesentliche Gründe, die für die Interaktion sprechen, genannt werden. Diese sind:

- **Navigation:** Kann der Nutzer mit einem Public Display interagieren, so geht dies oft mit der Navigation innerhalb des Systems einher. Dadurch hat er die Möglichkeit, gezielt die Inhalte zu betrachten, die ihn interessieren. Je nach Zweck und Komplexität des Systems kann die Komplexität der Navigation dabei stark variieren.
- **Ausdrücken von Interesse:** Wie bereits erwähnt, können Nutzer durch Interaktion mit dem Display und etwaige damit verbundene Navigationsmöglichkeiten ausdrücken, welche Inhalte sie interessieren und welche nicht. Diese Informationen sind für Betreiber der Displays von großem Nutzen, da sie dadurch Inhalte an die Interessen der Nutzer anpassen können.
- **Mitnehmen von Inhalten:** Interaktion kann auch dazu genutzt werden, die Mitnahme von Inhalten zu unterstützen. Beispielsweise könnten URLs oder QR-Codes auf dem Display angezeigt werden, um Nutzern die Möglichkeit zu geben, die dargestellten Inhalte oder einen Teil davon auch auf ihrem Smartphone zu betrachten. Davies, Clinch und Alt zu Folge [20] könnte man dem Nutzer alternativ die Möglichkeit dazu geben, Inhalte direkt an die eigene oder eine andere E-Mail Adresse zu senden.

¹engl. gaze interaction



Abbildung 2.3: Das Telectroscope auf der Londoner Tower Bridge [80].

- **Hinzufügen von Inhalten:** Das Hinzufügen von Inhalten durch Nutzer des Systems kann ebenfalls eine große Rolle spielen. In einigen der in Kapitel 2.3 erwähnten Arbeiten, wie beispielsweise dem Wray Photo Display [67, 68], e-Campus [27] oder Plasma Poster Network [19], war dies elementarer Bestandteil der Systeme.

2.3 Arten von Public Displays und historische Entwicklung

In Kapitel 2.1 wurde bereits darauf hingewiesen, dass Anordnung, Ausrichtung und Größe von Public Displays keinesfalls fest definiert sind. Ebenso variiert auch der Einsatzzweck. Dies wird vor allem deutlich, betrachtet man im Folgenden die verschiedenen Arten von Public Displays sowie deren historische Entwicklung. Die den Beispielen zugrunde liegende Klassifizierung basiert auf der Arbeit von Davies et al. [20].

Media-Links

Die erste Arbeit, die man noch heute unter der Kategorie Public Display einordnet, wurde bereits im Jahr 1980 veröffentlicht. Das Hole-In-Space [81], eine als öffentliche Kommunikations-Skulptur beschriebene Installation, ermöglichte es Einwohnern New Yorks in Form eines Videostroms mit Einwohnern von Los Angeles zu kommunizieren. Die Kommunikation selbst erfolgte dabei über Bildschirme und Kameras, die in einem Schaufenster in New York City und einem Kaufhaus in Los Angeles für eine Dauer von drei Tagen aufgestellt waren. Der Videostrom wurde über eine Satellitenverbindung übertragen.

Paul St George, ein in London lebender Multimedia-Künstler, entwarf The Telectroscope, eine Kunst-Installation die eine Videoverbindung zwischen London und New York ermöglicht [80] und damit dem Hole-In-Space ähnelt (siehe Abbildung 2.3).

In den kommenden Jahren folgten weitere sogenannte Media-Links. Sie alle ermöglichten Videoverbindungen über größere Entfernungen. Allerdings wurde jedes System für einen ganz bestimmten Einsatzzweck entwickelt.

Die Bellcore Labs, später bekannt als Telcordia Technologies, entwickelten beispielsweise ein System namens VideoWindow [25], welches Forscher über zwei Stockwerke eines Gebäudes miteinander verband. Um das Videobild anzuzeigen, wurden große Projektionsflächen in Gemeinschaftsbereichen genutzt.

Ambient-Displays (Umgebungs-Bildschirme)

Im Laufe der nächsten zehn Jahre entdeckte man neue Verwendungsmöglichkeiten für Public Displays. Die 90-er Jahre waren geprägt von tragbaren Bildschirmen beziehungsweise Anzeigen und auch solchen, die dafür gedacht waren, Statusinformationen ihrer Umgebung zu übermitteln.



Abbildung 2.4: Das Living Plant Display. Die Recycling-Tonne auf der rechten Seite wird intensiver genutzt. Daher ist die Pflanze der rechten Lampe stärker ausgesetzt und ihre Blätter richten sich nach rechts aus [29].



Abbildung 2.5: Breakaway in aufrechter Position [32].

Böhlen und Mateas [9] entwickelten beispielsweise ein System in Form einer Zimmerpflanze, welches durch Veränderung der Haltung über eingegangene E-Mails informierte. Die Idee, Pflanzen zu nutzen, um Informationen aus der Umgebung zu übermitteln, wurde in nachfolgenden Arbeiten aufgegriffen und weiterentwickelt. LaughingLily [4] gibt Auskunft darüber, ob und wenn ja welche Art einer Konversation gerade im Gange ist. Sie kann unterscheiden zwischen Stille, einer produktiven Konversation sowie Argumentation.

Ähnliche Arbeiten wie Living Plant Display [29], Spore 1.1 [22] und PlantDisplay [39] haben zum Ziel, die Lebensbedingungen realer Pflanzen zu verändern, um durch das sich daraus verändernde Wachstum beziehungsweise die sich ergebende Gesundheit Informationen zu übermitteln. Im Fall des Living Plant Display [29] beispielsweise wurde zunächst eine künstliche, später eine echte Pflanze in einer Cafeteria platziert. Rechts und links der Pflanze befanden sich neben einem Recycling- und Müllcontainer auch jeweils eine Lampe. Je mehr Menschen den Recyclingcontainer nutzten, um Abfall zu entsorgen, desto intensiver leuchtete die Lampe auf der rechten Seite (siehe Abbildung 2.4). Mit dem Müllcontainer verhielt es sich ähnlich mit dem Unterschied, dass in diesem Fall die linke Lampe intensiver leuchtete. Die Blätter der Pflanze richteten sich langfristig zu der Lampe hin aus, die intensiver leuchtete. Dadurch war es möglich, allein durch betrachten der Blätter und deren Ausrichtung einzuschätzen, zu welchem Grad in der Cafeteria recycled wurde.

Weitere Beispiele für Umgebungs-Bildschirme stammen von Rodenstein [59], Jafarinaimi

et al. [32] und Rogers et al. [60]. Ersterer experimentierte mit physischen Fenstern und nutzte diese mittels Projektion zur Anzeige von Wetterinformationen, da sie, laut Rodenstein, über mehrere besondere Eigenschaften verfügten. Zum einen seien Menschen bereits daran gewöhnt, Fenster an- und durch sie hindurchzuschauen, da diese eine physische Schnittstelle darstellten (beispielsweise eben diese zwischen einem Wohnraum und einem Garten oder der Außenwelt im Allgemeinen). Gerade diese Eigenschaft der Schnittstelle mache sie einerseits informativ, andererseits aber auch erholsam.

Bei Breakaway [32] (siehe Abbildung 2.5) handelt es sich um eine kleine Skulptur für den Schreibtisch. Sie agiert als Ambient-Display und soll Nutzer, die oft und lange am Schreibtisch sitzen, dazu ermutigen öfter Pausen zu machen. Breakaway kann durch Sensoren im Schreibtisch-Stuhl ermitteln, wie lange der Nutzer schon am Tisch sitzt. Ist dies schon lange der Fall, lässt sie sich hängen. Ist der Nutzer erst kürzlich aufgestanden und hat sich dann wieder hingesetzt, signalisiert sie das durch eine aufrechte Haltung.

Rogers et al. verfolgten mit The Clouds [60] ein ähnliches Ziel. Hierbei handelte es sich um eine physische Installation im Atrium eines Bürogebäudes, welche aus zwei Gruppen jeweils gleich eingefärbter Kugeln bestand. Diese hingen von der Decke herab und sollten den Eindruck von Wolken vermitteln. Der relative Abstand der beiden Kugel-Gruppen gab an, in welchem Verhältnis Menschen, die das Gebäude betreten, den Aufzug beziehungsweise die Treppen nahmen, um in obere Stockwerke zu gelangen. Die Kugeln waren so angebracht, dass sie aus dem ganzen Atrium sichtbar waren und sofort ins Auge sprangen, sobald man das Gebäude betrat. Ziel des Projekts war es, Menschen bei der entsprechenden Entscheidung zu unterstützen und sie zu ermutigen, zunächst zu reflektieren.

Bereits jetzt lässt sich also sagen, dass Public Displays keinesfalls 'nur' elektronische Bildschirme darstellen, die mit Inhalten 'gefüttert' werden. Ganz im Gegenteil: Es existiert eine große Vielfalt unterschiedlicher Anzeigen; viele davon wurden für einen ganz speziellen Verwendungszweck konstruiert, andere wiederum waren beziehungsweise sind eher generischer Natur. Diese soeben beschriebene Vielfalt setzt sich unter anderem in Form der tragbaren Bildschirme fort.

Thinking Tags [12, 13, 15] stellt eine der frühesten Arbeiten in diesem Bereich dar. Ziel war es, bei Veranstaltungen die Kommunikation unter den Teilnehmern zu fördern. Dazu wurden konventionelle Namensschilder um fünf Farb-LEDs ergänzt. Auch wurden die Namensschilder so umgebaut, dass eine Kommunikation mittels Infrarot möglich war. Standen nun zwei Teilnehmer einer Veranstaltung nahe genug beieinander, tauschten ihre Namensschilder Antworten zu fünf zuvor beantworteten Fragen aus. Stimmten die Antworten beider Teilnehmer zu einer Frage überein, leuchteten die LEDs an den entsprechenden Positionen auf den Schildern grün. Stimmten sie nicht überein, leuchteten sie rot. Dies sollte eine schnelle Identifikation potentieller Gesprächsthemen ermöglichen.

Die Idee wurde weitergeführt und es entstand MemeTags [12, 14]. Dieses System setzte nicht auf LEDs sondern auf Flüssigkristallbildschirme (LCDs). Dadurch konnten die Namensschilder die Namen ihrer Träger untereinander austauschen, welche dann in einen vordefinierten Text eingesetzt und auf dem LCD dargestellt wurden. Zur Übermittlung der Daten wurde auch hier Infrarot eingesetzt.

Public Displays für den Arbeitsplatz

Zu Beginn der 2000-er Jahre änderte sich die Ausrichtung erneut und es kam die Frage auf, wie Public Displays am Arbeitsplatz eingesetzt werden könnten beziehungsweise inwiefern diese dort von Nutzen sein könnten. Dabei fokussierte man sich vor allem auf Türschilder, Besprechungs- und Gemeinschaftsräume.

Frühe Arbeiten wie jene von Mankoff und Schilit [43] oder Ngyuen et al. [53] nutzten unter anderem berührungsempfindliche Bildschirme, um in Büroumgebungen ortsbasierte Aufgaben schnell erledigen, Informationen beschaffen und Nachrichten hinterlassen zu können. Die sogenannten Palplates [43] beispielsweise wurden in beziehungsweise vor Küchen oder Besprechungsräumen angebracht. Abhängig von der am entsprechenden Ort häufig zu erledigenden Aufgabe (zum Beispiel Reservieren des Besprechungsraums oder Bestellen von Kaffee) wurde die Benutzeroberfläche der Bildschirme so angepasst, dass die jeweilige Aufgabe möglichst schnell erledigt werden konnte.

Hermes [18] wiederum widmete sich ganz der Kommunikation. Im Rahmen dieses an der Lancaster-Universität in England durchgeführten Projekts wurden vor einigen Büros PDAs angebracht. Auf den Bildschirmen der PDAs konnten Mitarbeiter und Besucher Nachrichten hinterlassen. Die Eingabe der Nachrichten erfolgte entweder direkt am Gerät oder per Web-Interface beziehungsweise SMS/MMS.

Doch man beschäftige sich nicht nur mit Kommunikation, Informationsbeschaffung und der effizienteren Abarbeitung von Aufgaben. Mit der Zeit kamen auch in diesem Bereich neue Anwendungsgebiete hinzu. Beispielsweise gab es Arbeiten, die der Frage nachgingen, wie das Bewusstsein sowie das Gemeinschaftsgefühl am Arbeitsplatz gestärkt werden können.

Einen Lösungsansatz stellt das Plasma Poster Network [19] dar. Entwickelt wurde es vom FX Palo Alto Laboratory, einem zu Fuji Xerox gehörenden Forschungslabor. Wie der Name schon andeutet, bestand das System aus drei Plasma-Bildschirmen. Diese wurden in verschiedenen Räumen des Labors aufgebaut. Das laboreigene Intranet sorgte dafür, dass sie mit Inhalten gespeist wurden. Außerdem war es möglich, Inhalte per E-Mail und über das Internet einzureichen. Nutzer des Systems berichteten, dass sie hauptsächlich solche Inhalte eingereicht hätten, von denen sie dachten, dass sie für andere Nutzer von Interesse wären. Laut der Autoren ließen ihre Beobachtungen den Schluss zu, dass die soziale Interaktion der Mitarbeiter des Labors durch das Projekt gefördert worden sei.

Im Rahmen eines anderen Projekts, genannt USIAumniFaces [62], wurde ein großer Bildschirm genutzt, um die Interaktion mit einem digitalen Jahrbuch zu ermöglichen. Eingesetzt bei einem Absolvententreffen führte es dazu, dass Erinnerungen geweckt beziehungsweise ausgetauscht wurden und so die Konversation unter den Teilnehmern gefördert werden konnte.

Netzwerkbasierete und langfristige Installationen

Die späten 2000-er Jahre waren und sind geprägt von großflächigen Displays, die oft in Form von Netzwerken für mehrere Jahre installiert und genutzt werden können. Ein typisches Beispiel hierfür sind die INFOSCREEN-Installationen [86] (siehe Abbildung 2.6). Dabei handelt es sich um großflächige, nicht interaktive Bildschirme, welche in Deutschland vornehmlich an U- und S-Bahnsteigen zu finden sind. Dargestellt wird ein gemischtes Programm aus Information, Unterhaltung und Werbung.

Ein weiteres Beispiel für diese Art von Public Displays ist The Wray Photo Display [67, 68]. Dieses wurde 2006 im Postamt von Wray, einem kleinen Dorf in Nordengland, installiert. Auf dem Bildschirm zu sehen sind Fotos und kurze Videoclips des Dorfes. Diese wurden im Lauf der Zeit von den Bewohnern aufgenommen und dem System hinzugefügt. Kommentare von Dorfbewohnern in einem danebenliegenden Notizbuch haben gezeigt, dass das System sehr positiv aufgenommen wurde und nach Ansicht der Bewohner eine Bereicherung für das Leben im



Abbildung 2.6: Ein INFOSCREEN am Münchner Hauptbahnhof [86].

Wray darstellt - sowohl für diejenigen, die seit ihrer Geburt im Dorf leben als auch für diejenigen, die erst später hinzugezogen sind.

Installationen in und durch Universitäten

Natürlich gab es auch entsprechende (Forschungs-)Projekte an Universitäten. Eines der ersten dieser Projekte war FLUMP (FLexible Ubiquitous Monitor Project) [24] der Universität Lancaster. Es diente den Mitarbeitern der Universität als multimediales Informationssystem und bestand aus mehreren sogenannten FLUMP stations. Diese wiederum setzten sich aus mehreren CRT-Monitoren und einem Computer zusammen und wurden an diversen Orten innerhalb des Fachbereichs für Informatik aufgebaut. Näherte sich ein Mitarbeiter einer FLUMP station, erkannte dies das System durch ActiveBadges [74] und die Monitore zeigten die persönliche HTML-Seite des Mitarbeiters an. Dieser musste im Voraus festlegen, welche Informationen die Seite enthalten solle. Insgesamt standen fünf Informationsdienste zur Verfügung: Time (aktuelle Uhrzeit und aktuelles Datum), Mail (Anzahl der gespeicherten, neuen und ungelesenen E-Mails), Timetable (kommende Termine), Fylde (gibt Auskunft darüber, ob das nächstgelegene Café im Fylde College offen oder geschlossen ist beziehungsweise wie lange es dauert, bis es wieder öffnet/schließt) und Matt (Darstellung eines Cartoons).

Das e-Campus System [27] - welches ebenfalls von Mitarbeitern der Universität Lancaster entwickelt wurde - besteht aus einem Verbund von Bildschirmen, die sich über den ganzen Campus der Universität erstrecken. Neben seiner Funktion als digitales Beschilderungssystem und Notfall-Warnsystem wurde es auch schon im Rahmen von Kunst-Festivals oder für diverse Studenten- und Forschungsprojekte genutzt.

Wie bei anderen Systemen, beispielsweise Wray Photo Display, Plasma Poster Network oder FLUMP, war es also auch hier möglich, Inhalte einzureichen und darzustellen. Der wesentliche Unterschied bestand jedoch darin, dass mit der Zeit ein Berechtigungssystem entwickelt wurde, welches zwischen dem Erstellen von Inhalten und Steuern/Kontrollieren der Bildschirme unterschied. Außerdem wurden sogenannte e-Channel eingeführt. Dabei handelte es sich um virtuelle Container, in denen Inhalte eines Themas aber unterschiedlicher Medien zusammengefasst werden konnten. Diese Container konnten dann von einem Bildschirm oder mehreren abonniert werden. Dadurch wurden neue Inhalte im Container automatisch in die Warteschlange darzustellender Inhalte eingereiht.

Ein ähnliches System, genannt iDisplays, wurde an der Universität Münster entwickelt [49, 51]. Im Gegensatz zu den oben genannten Beispielen lag der Fokus in diesem Fall auf der Aufmerksamkeit der Nutzer. Man wollte herausfinden, wie Inhalte präsentiert werden müssten,

damit diese von Menschen wahrgenommen werden, vor allem von jenen, die die Bildschirme und damit auch die Inhalte oft passieren (siehe auch Kapitel 2.5).

Müller et al. [51] argumentieren, dass Slideshows und Lauftexte (die zur Entstehungszeit der Arbeit am häufigsten verwendeten Darstellungsformen für Inhalte auf Public Displays) dafür ungeeignet seien, da Menschen, die die Inhalte oft passierten, nur schwer sagen könnten, ob die Informationen aktualisiert worden seien. Daher entwickelte man ein zweigeteiltes System, welches zum einen aus einem News Display und zum anderen aus einem Reminder Display bestand. Ersteres wurden dazu genutzt, Neuigkeiten anzuzeigen und das so schnell und früh wie möglich. Letzteres zeigte jene Informationen an, die vom System zum aktuellen Zeitpunkt als wichtig erachtet wurden, beispielsweise solche, die an eine Veranstaltung erinnern, welche in wenigen Minuten beginnen sollte.

2.4 Interaktionstechniken

In Kapitel 2.2 wurde bereits darauf eingegangen, welche Vorteile die Möglichkeit der Interaktion im Kontext von Public Displays mit sich bringt. Die Frage ist nun, **wie** mit Public Displays interagiert werden kann. Im folgenden Kapitel werden daher die gängigsten Interaktionstechniken vorgestellt. Es wird erklärt, welche Funktionsprinzipien diesen zu Grunde liegen und welche Vor beziehungsweise Nachteile die jeweiligen Techniken haben.

Kurdyukova, Obaid und André [38] zu Folge lassen sich Interaktionstechniken für Public Displays in drei Klassen unterteilen:

- Direkte Interaktion
- Körperliche Interaktion
- Interaktion mittels mobiler Endgeräte

Direkte Interaktion

Direkte Interaktion kann auf unterschiedlichen Wegen erfolgen. Möglich ist beispielsweise die Nutzung mobiler Endgeräte in Kombination mit Kurzstrecken-Funktechnologien wie NFC [65, 17].

In der Vergangenheit fanden jedoch vor allem Touchscreens in entsprechenden Systemen Verwendung [55, 57, 31]. Die Interaktion wird in diesem Fall als natürlich [64], schnell, zuverlässig und einfach empfunden [63]. Sie kann erfolgen, ohne dass besondere Abläufe, wie beispielsweise das Ausführen von Gesten erlernt werden müssen. Dies ist gerade im Kontext von Public Displays wichtig, da eine Interaktion hier nur von kurzer Dauer ist [50] und Nutzer oft aufgeben, sofern die Interaktion nicht sofort erfolgreich verläuft [44].

Da im Zuge der Interaktion jedoch ein physischer Kontakt zum Bildschirm hergestellt werden muss, kann dies zu hygienischen Bedenken führen [20, 54]. Von Vorteil wäre es daher, wenn entsprechende Bildschirme in einer Höhe montiert werden könnten, die es mehreren potentiellen Nutzern ermöglicht, gleichzeitig uneingeschränkter Blickkontakt zum Bildschirm aufzubauen. Dies wäre auch deshalb wünschenswert da interaktive Displays so vor Vandalismus geschützt werden könnten [20]. Durch den notwendigen physischen Kontakt ist es jedoch nicht ohne Einschränkungen möglich.

Abhängig von der im Touchscreen verbauten Technologie (resistiv, kapazitiv oder optisch), können sich Unterschiede in der Funktionsweise und im Nutzererlebnis ergeben. Auf diese und die daraus resultierenden Vor- und Nachteile bezüglich der Interaktion wird nun näher eingegangen.

Resistive Touchscreens nutzen zwei stromleitende Beschichtungen, die auf die Front des Bildschirms aufgebracht werden. Bei der Interaktion mit dem Bildschirm wird ein leichter Druck auf diesen ausgeübt. Dadurch werden die Beschichtungen für einen kurzen Moment miteinander verbunden. Dies hat zur Folge, dass ein Stromkreis geschlossen und Spannung erzeugt wird. Anhand der gemessenen Spannungshöhe und unter Zuhilfenahme eines zusätzlichen elektronischen Bauteils - des Controllers - kann berechnet werden, an welchem Punkt Druck auf den Bildschirm ausgeübt beziehungsweise mit dem Bildschirm interagiert wurde [82]. Diese Information wiederum kann genutzt werden, um mittels Software auf die Interaktion zu reagieren. Hat das System durch Berechnung des Druckpunkts beispielsweise erkannt, dass vom Nutzer ein Menüpunkt ausgewählt wurde, kann der Menüpunkt visuell hervorgehoben werden, um dem Nutzer Rückmeldung darüber zu geben, welches Element der Benutzeroberfläche ausgewählt wurde.

Da die Interaktion mit resistiven Touchscreens auf Ausübung von Druck basiert, ist es möglich, im Gegensatz zu kapazitiven Touchscreens, entsprechende Bildschirme auch mit Handschuhen, Prothesen oder beispielsweise Eingabestiften zu bedienen.

Des Weiteren sind sie widerstandsfähiger gegenüber Chemikalien sowie unsachgemäßem Gebrauch. Genannte Komponenten in herkömmlichen Bildschirmen nachzurüsten ist zudem sehr einfach und dadurch kostengünstig [82, 20]. Des Weiteren können entsprechende Touchscreens eine sehr hohe Genauigkeit aufweisen [20].

Gleichzeitig zeigt diese Art von Touchscreens jedoch eine - im Vergleich zu kapazitiven Touchscreens - verringerte Responsivität bei Berührung [20].

Kapazitive Touchscreens verfügen im Gegensatz zu resistiven Touchscreens über lediglich eine stromleitende Beschichtung. Durch an den Rändern der Beschichtung angebrachte Elektroden wird Spannung erzeugt, welche dafür sorgt, dass ein elektrisches Feld entsteht. Wird der Bildschirm mit einem stromleitenden Objekt, wie beispielsweise einem menschlichen Finger berührt, wird Spannung erzeugt, welche in den Ecken des Bildschirms gemessen werden kann. Die Höhe dieser Spannung und deren Verteilung über die vier Messpunkte ist direkt abhängig vom Berührungspunkt. Daher kann dieser auf Basis der Messwerte berechnet werden [83].

Da bei kapazitiven Touchscreens, im Gegensatz zu resistiven Touchscreens, kein Druck ausgeübt werden muss, kann dies zu einem besseren Nutzererlebnis führen. Gleichzeitig besteht hier der Vorteil, dass mehr als zwei Finger gleichzeitig erkannt werden können, was dazu führt, dass entsprechende (Mehr-)Finger-Gesten möglich sind.

Eine Einschränkung besteht jedoch auch. Zur Interaktion ist die Berührung mit einem stromleitenden Objekt (beispielsweise einem menschlichen Finger) notwendig [20]. Daher kann diese Art berührungsempfindlicher Bildschirme in der Regel nicht mit Handschuhen, Eingabestiften oder Prothesen bedient werden.

Optische Touchscreens nutzen optische Sensoren am Rand des Bildschirms, wie beispielsweise kleine Kameras, um eine Berührung zu erkennen und den Berührungspunkt zu berechnen [20].

Wie bei resistiven und kapazitiven Touchscreens gibt es auch hier unterschiedliche Umsetzungen. Ein mögliches Funktionsprinzip wird im Folgenden exemplarisch anhand einer inzwischen weit verbreiteten Technik, der sogenannten FTIR (verhinderte Totalreflexion) beschrieben.

FTIR basiert auf dem physikalischen Effekt der Totalreflexion. Dieser lässt sich laut Han [28] wie folgt erklären: Wird Licht durch ein transparentes Medium mit einem geringeren Brechungsindex als Luft (zum Beispiel Acrylglas) geschickt, wird an der Schnittstelle Luft - Medium (im Beispiel: Luft - Acrylglas), abhängig vom Eintrittswinkels des Lichts, ein größerer oder kleinerer Teil des Lichts gebrochen. Liegt der Eintrittswinkel über einem kritischen Wert, findet innerhalb des Mediums eine Totalreflexion statt. Berührt man die Schnittstelle (im Beispiel also die Oberfläche der Acrylglasplatte) mit dem Finger, unterbricht dies die Totalreflexion im Innern des Mediums. Am Berührungspunkt tritt, aufgrund der unterschiedlichen Brechungsindizes von Medium und Finger, Licht aus dem Medium aus, welches mit einer Kamera erfasst werden kann. Dadurch kann bestimmt werden, an welchem Punkt die Oberfläche berührt wurde.

Optische Touchscreens bieten aufgrund der optischen Erfassung wesentliche Vorteile gegenüber den beiden anderen beschriebenen Verfahren. Beispielsweise ist es dadurch nicht notwendig Druck aufzuwenden, was zu einem besseren Nutzungserlebnis führt. Eine Interaktion mit Prothesen oder Handschuhen ist ebenso möglich. Durch die gute Skalierbarkeit FTIR-basierter Systeme eignen diese sich außerdem sehr gut für Multi-Touch und die gleichzeitige Interaktion mehrerer Nutzer. Allerdings ist die Nutzung von Fingergesten in diesem Fall problematisch, da Berührungspunkte zwar korrekt erkannt aber den Nutzern nicht eindeutig zugeordnet werden können [28].

Körperliche Interaktion

Körperliche Interaktion kann Kurdyukova, Obaid und André [38] zu Folge durch die Ausführung räumlicher Gesten, Analyse der Körperhaltung oder räumliche Nähe erfolgen.

Vor allem **Gesten**, also ausgeführte Bewegungen des Körpers die Informationen beinhalten [40], stellen eine häufig genutzte Interaktionsmethode im Zusammenhang mit Public Displays dar. Die Erfassung solcher Gesten erfolgt in der Regel entweder über Kameras mit Tiefensensor (zum Beispiel Kinect) oder Zusatzgeräte mit Beschleunigungs- oder Positionssensoren (zum Beispiel Wii-Remote) [20].

Laut Walter et al. [72] ergeben sich mehrere Vorteile im Vergleich zu Touchscreens. Beispielsweise kann die Interaktion bereits aus der Ferne erfolgen; der Bildschirm muss, im Gegensatz zu Touchscreens, nicht berührt werden. Dies hat nicht nur Vorteile in Bezug auf das Nutzererlebnis; es werden auch hygienische Bedenken ausgeräumt. Außerdem kann der entsprechende Bildschirm dadurch möglicherweise besser vor äußeren Einflüssen wie Vandalismus geschützt werden [5]. Wie Müller et al. [50] gezeigt haben, kann es auch vorkommen, dass Passanten durch eine unbeabsichtigte Interaktion aus der Ferne auf ein entsprechendes Display aufmerksam werden. Ebenso ist es möglich, dass Nutzer durch besonders ausdrucksstarke und/oder performative Gesten auf sich selbst und damit auch auf das Display aufmerksam machen [58].

Mit der anwendungsseitigen Unterstützung von Gesten gehen jedoch auch einige Herausforderungen einher. Beispielsweise existieren keine fest definierten Gesten-Sets oder Richtlinien für Entwickler, die anwendungsübergreifend gelten. Je nach Anwendungsfall kann es nötig sein, dass ein neues Gesten-Set definiert werden muss, soll eine Anwendung oder ein Public Display durch die Ausführung von Gesten gesteuert werden können [52]. Potentielle Nutzer müssen darüber informiert werden, dass das vorgestellte System die Nutzung von Gesten unterstützt, um welche es sich dabei handelt und wie diese ausgeführt werden [72].

Eine weitere Herausforderung stellt die Erkennung der Gesten dar. Damit diese korrekt erfolgen kann, müssen Beginn, Fortsetzung und Ende der Geste klar voneinander getrennt sein [6, 26, 75]. Die Trennung dieser Schritte stellt allerdings ein großes Problem dar, da keine natürlichen oder vordefinierten Trenner existieren [20]. Die in verschiedenen Arbeiten beschriebene ergänzende Nutzung weiterer Modalitäten [66, 47] (wie zum Beispiel Interaktion mittels Blickkontakt oder Berührung) wirft wiederum neue Fragen auf [20].



Abbildung 2.7: Point & Shoot wird genutzt, um Bildelemente auszuwählen (hier: Puzzlestücke). Dazu wird das Kamerabild im ersten Schritt auf das gewünschte Element ausgerichtet (Bild links). Im zweiten Schritt wird die Auswahl bestätigt. Daraufhin erscheint ein virtuelles Raster mit Codes auf dem Bildschirm (Bild Mitte). Dieses dient der Berechnung der Zielkoordinaten. Schließlich verschwindet das Koordinatensystem und das gewünschte Element wird ausgewählt (Bild rechts). [5]

Interaktion mittels mobiler Endgeräte

Die Interaktion unter Zuhilfenahme mobiler Endgeräte wie beispielsweise Smartphones stellt eine Alternative zu direkten und körperlichen Interaktionstechniken dar. Sie bietet den Vorteil, dass potentielle Nutzer bereits an entsprechende Endgeräte gewöhnt und mit deren Nutzung vertraut sind. Außerdem haben viele von ihnen diese immer dabei, sie sind also stets verfügbar. Weitere Vorteile liegen vor allem darin, unter minimalem physischen Aufwand aus größerer Entfernung mit dem entsprechenden Bildschirm interagieren zu können [5]. Rukzio et al. [64] zu Folge sehen Nutzer vor allem Nachteile im konstanten Fokus-Wechsel zwischen mobilem Endgerät und Public Display. Im Folgenden wird nun näher auf entsprechende Arbeiten eingegangen.

Erfolgt die Interaktion über ein mobiles Endgerät, spielt die im Endgerät eingebaute Kamera in vielen Fällen eine große Rolle [5, 42, 61, 10]. Ballagas, Rohs und Sheridan [5] beispielsweise nutzten diese, um das dazugehörige Telefon als Eingabegerät im dreidimensionalen Raum zu verwenden. Die Sweep genannte Technik ermöglichte es, durch das Verarbeiten der durch die Kamera aufgezeichneten Bilder und das Vergleichen dieser mit relativen Bewegungsdaten entlang der drei Achsen, einen Cursor auf dem Bildschirm zu steuern.

Eine weitere Technik - genannt Point & Shoot - erlaubte das Selektieren von Bildelementen durch das Ausrichten der Kamera (siehe Abbildung 2.7). Wie in Arbeiten von Madhavapeddy et al. [42] und Rohs [61] kommen hierbei visuelle Codes zum Einsatz. Diese werden von der Kamera des Endgeräts erfasst und dienen der Bestimmung der Zielkoordinaten. Dadurch kann ermittelt werden, welches Element auf dem Bildschirm ausgewählt wurde.

Boring et al. [10] verfolgen einen ähnlichen Ansatz. Wesentliche Unterschiede liegen jedoch darin, dass das System für die Nutzung durch mobile Endgeräte entworfen wurde, die mit einem Touchscreen ausgestattet sind. Der Touchscreen erlaubt das Auswählen der entsprechenden Elemente durch direkte Interaktion, konkret durch berühren beziehungsweise antippen. Die Auswahl wird dann auf den 'großen Bildschirm projiziert'. Im Gegensatz zu Sweep beziehungsweise Point & Shoot [5] ermöglicht Touch Projector [10] das Zoomen, also die Vergrößerung eines Bildausschnitts. Dabei wird die Auswahl von Bildelementen aus größerer Entfernung erleichtert. Zudem kann das auf dem Touchscreen dargestellte Kamerabild eingefroren werden, um das Auswählen zu vereinfachen, während der Zoom aktiv ist.

2.5 Sweetspot und Positionierung des Nutzers

Wie aus Kapitel 2.4 hervorgeht, existieren unterschiedliche Techniken, die die Interaktion mit Public Displays ermöglichen. Diese unterscheiden sich unter anderem in den Anforderungen, die für den jeweiligen Interaktionsraum gelten. Beispielsweise erfordern Touchscreens eine Berührung zur Interaktion. Folglich muss sich der Nutzer in unmittelbarer Nähe zum Bildschirm befinden. Ähnliches gilt auch für die Interaktion mittels Blickkontakt. Es wird vorausgesetzt, dass sich der Nutzer in einer Entfernung von 70 cm [35] vor dem Bildschirm befindet. Um die Augen des Nutzers und damit dessen Blickpunkt zuverlässig erfassen zu können, ist es zudem von Nöten, dass sich der Kopf stets in einem virtuellen Feld von 44×22 cm beziehungsweise 30×22 cm Größe befindet² [69].

Entsprechende Anforderungen gelten auch für die Interaktion mittels Gesten. Erfolgt die Erfassung der Gesten beispielsweise mit Kinect der zweiten Generation, darf der Abstand zu Kinect 0,5 m nicht unter- und 4,5 m nicht überschreiten. Ebenso sind die Winkel des horizontalen sowie vertikalen Erfassungsfelds festgelegt. Diese betragen 70 beziehungsweise 60° [95].

Nutzer neigen dazu, sich vor einem Public Display in den Punkt zu begeben, in dem die Wahrnehmung des Inhalts optimal möglich ist [7, 3]. Dieser Punkt wird Sweetspot genannt. Da Größe und Lage des Sweetspots jedoch von der Interaktionstechnik abhängen, kann es notwendig sein, Nutzer in diesen Punkt zu führen [3].

Voraussetzung hierfür ist jedoch, dass das Display die Aufmerksamkeit potentieller Nutzer auf sich zieht. Es hat sich gezeigt, dass Public Displays von potentiellen Nutzern oft übersehen oder bewusst ignoriert werden [51, 30]. Dieser Effekt - genannt Bildschirm-Blindheit³ - hat mehrere Ursachen. Zum einen erfahren viele Menschen heutzutage, vor allem in städtischen Umgebungen, eine Informationsüberflutung [46]. Eppler und Mengis [23] zu Folge bedeutet dies, dass eine Person mit zu vielen Informationen konfrontiert wird und nicht mehr dazu in der Lage ist, diese Menge an Informationen zu verstehen beziehungsweise mit ihr umzugehen. Resultat dessen ist eine sehr selektive Wahrnehmung [30]. Dabei werden in vielen Fällen - ähnlich der Werbebanner im Internet - auch Public Displays ausgefiltert, da erwartet wird, dass diese uninteressante Inhalte wie zum Beispiel Werbung anzeigen [51].

Werden sie doch bemerkt, gehen potentielle Nutzer oft davon aus, dass diese keine Interaktionsmöglichkeiten bieten. Die sogenannte Interaktions-Blindheit⁴ [55] ist jedoch auch nicht weiter verwunderlich, da die meisten Public Displays nach wie vor tatsächlich keine Möglichkeiten zur Interaktion bieten [30].

Es gilt also, potentielle Nutzer (vor allem Passanten) auf die entsprechenden Bildschirme aufmerksam zu machen, die Möglichkeit der Interaktion zu vermitteln und sie schließlich zur Interaktion zu motivieren [50]. Die Frage ist nun, wie dies bewerkstelligt werden kann.

Lenkung der Aufmerksamkeit

Wie in verschiedenen Arbeiten gezeigt wurde [51, 30], gibt es eine Reihe von Faktoren, die die Aufmerksamkeit potentieller Nutzer lenken und so möglicherweise dazu beitragen können, dass diese Public Displays betrachten. Darunter fallen beispielsweise die Größe des Displays, dessen Farbvielfalt und die Dauer, während welcher es für potentielle Nutzer sichtbar ist [51]. Huang et al. [30] nennen weitere Faktoren wie zum Beispiel den Winkel, in dem man sich dem Display nähert, die Höhe in der es installiert wurde, sowie den Grad der Ablenkung (zum Beispiel durch andere Displays).

²Die Maße beziehen sich auf Eye-Tracker des Herstellers Tobii. Sie hängen davon ab, ob das Gerät mit 60 Hz oder 120 Hz läuft.

³engl. display blindness

⁴engl. interaction blindness

Kommunizieren von Interaktivität

Müller et al. [50] identifizierten insgesamt 6 Techniken, um die Interaktivität von Public Displays zu kommunizieren. Diese lauten wie folgt:

- **Aufruf zur Interaktion⁵**: Ein Aufruf zur (Inter-)Aktion, kann beispielsweise durch das Darstellen eines entsprechenden Textes auf dem Display erfolgen. Erfolgt die Interaktion durch das Ausführen von Gesten, könnte der Nutzer gebeten werden, eine bestimmte Geste auszuführen, um die Interaktion zu starten.
- **Anziehungs-Sequenz⁶**: Hierbei handelt es sich um die automatisch ablaufende Abfolge unterschiedlicher Objekte, wie beispielsweise Fotos [37]. Ebenso kann eine sich wiederholende Bewegung physischer Objekte genutzt werden [34].
- **Analoge Beschilderung⁷**: Neben den bereits genannten Techniken ist es auch möglich, analoge beziehungsweise physische Beschriftungen, Auszeichnungen oder Beschilderungen zu nutzen, um die Interaktivität zu kommunizieren. Peltonen et al. [57] nutzten auf Papier ausgedruckte Instruktionen. Doch auch in anderen Arbeiten fand diese Technik bereits Anwendung [37, 44].
- **Aufforderung zur Interaktion durch andere**: Alternativ oder ergänzend können potentielle Nutzer auch angesprochen und so auf die Möglichkeit der Interaktion aufmerksam gemacht beziehungsweise zur Interaktion eingeladen werden [31]. Im Optimalfall geschieht dies durch aktive Nutzer des Systems [44, 57].
- **Vorkenntnisse**: Hat ein Nutzer bereits mit einem bestimmten oder ähnlichen System interagiert und sich dadurch Vorkenntnisse angeeignet, führt dies möglicherweise dazu, dass er eher oder erneut mit einem System interagiert [44].
- **Honigtopf-Effekt⁸**: Sind bereits einige Nutzer auf ein Display aufmerksam geworden, kann es sein, dass noch mehr Nutzer durch diese Ansammlung angezogen werden und mit dem Display interagieren möchten. Dies bezeichnet man als Honigtopf-Effekt [16].

2.6 Nutzerrepräsentationen

Eine weitere, vielversprechende Technik um Interaktion zu kommunizieren, stellen Nutzerrepräsentationen dar [50, 2]. Auf diese wird im folgenden Kapitel näher eingegangen.

Technische Errungenschaften der letzten Jahre in den Bereichen Bilderfassung sowie -verarbeitung führten zu neuartigen Technologien, welche unter anderem die automatisierte optische Erfassung des dreidimensionalen Raums ermöglichten. Unter Zuhilfenahme spezieller Kameras wurde es so möglich, zwischen Vorder- und Hintergrund eines Bildes zu unterscheiden und unterschiedliche Objekte im Raum automatisiert zu identifizieren. Ebenso konnte man dadurch den menschlichen Körper beziehungsweise dessen Bewegungen erfassen und diese auf andere Darstellungen übertragen. Dies führte zur Entwicklung unterschiedlicher Formen von Nutzerrepräsentationen. Hierbei handelt es sich um bildliche Darstellungen des Nutzers, die dessen Bewegungen (oft aber nicht zwingend) in Echtzeit abbilden. Abbildung 2.8 zeigt drei

⁵engl. call-to-action

⁶engl. attract sequence

⁷engl. analog signage

⁸engl. honeypot effect

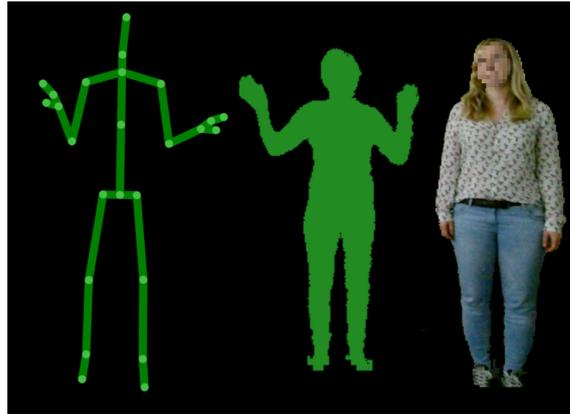


Abbildung 2.8: Drei typische Nutzerrepräsentationen: Skelett, Silhouette und Spiegelbild (von l. nach r., eigene Grafik)

dieser Repräsentationsformen.

Im Allgemeinen unterscheidet man zwischen zwei Klassen von Repräsentationen. Die Repräsentationen der beiden Klassen unterscheiden sich vor allem durch die Art und Weise, wie sie von ihren Nutzern erkannt beziehungsweise identifiziert werden können.

Die erste Klasse enthält all jene Repräsentationsformen, die allein durch ihre äußeres Erscheinungsbild (also Form, Größe und Aussehen) von Nutzern erkannt werden können (appearance matching). Die zweite Klasse hingegen beinhaltet jene Repräsentationsformen, die auf einem Vergleich von Bewegungen basieren. Konkret muss hier der jeweilige Nutzer seine Körperbewegungen mit denen der entsprechenden Repräsentation(en) vergleichen (kinesthetic matching) [2].

Nutzerrepräsentationen waren und sind Gegenstand einer Vielzahl von Forschungsarbeiten im Bereich Public Displays [2, 50, 73, 8, 70, 71, 56, 48, 72]. Diese hatten unter anderem zum Ziel herauszufinden, inwiefern sich Nutzerrepräsentationen zur Kommunikation von Interaktivität eignen und wie potentielle Nutzer auf diese reagieren. Verschiedene Repräsentationsformen wurden miteinander verglichen, um die Vor- und Nachteile im Einzelnen herauszuarbeiten. Im Fokus der Forschung stehen und standen dabei vor allem die drei folgenden Repräsentationsformen: das Skelett, die Silhouette und das Spiegelbild [50, 56, 1, 2, 71].

Müller et al. [50] beschäftigten sich mit den Fragen, wie Passanten die Interaktivität von Public Displays bemerken und wie schnell diese registriert wird. Man fand heraus, dass sich Nutzerrepräsentationen, insbesondere die Repräsentationsformen Silhouette und Spiegelbild, besonders gut dazu eignen, potentiellen Nutzern die Möglichkeit der Interaktion zu kommunizieren. In einer ersten (Labor-)Studie wurden diese auf appearance matching basierenden Repräsentationsformen mit zwei auf kinesthetic matching basierenden Repräsentationsformen verglichen. Dabei handelte es sich um eine zweidimensionale Avatar-Darstellung bestehend aus Kopf, Torso und Händen sowie eine abstrakte Darstellung, welche den Kopf, abstrakte Augen und einen abstrakten Mund beinhaltete. Die Teilnehmer wurden gebeten, ein Display seitlich zu passieren, dieses zu betrachten und zu entscheiden, ob die darauf dargestellte Nutzerrepräsentation auf ihre Bewegungen reagierte oder nicht. Unter beiden Bedingungen (interaktiv und nicht interaktiv) waren Spiegelbild und Silhouette ähnlich effizient und effizienter als Avatar und abstrakte Repräsentation.

In einer von den Autoren im Anschluss durchgeführten Feldstudie interagierten deutlich mehr Passanten mit entsprechenden Displays, wenn Silhouetten (+47 %) oder Spiegelbilder (+90 %)

zur Kommunikation der Interaktivität genutzt wurden anstelle herkömmlicher Techniken wie beispielsweise dem Aufruf zur Interaktion.

In einer Studie von Parra, Klerkx und Duval [56] konnten diese Ergebnisse nur zum Teil bestätigt werden. Ziel der Studie war eine Quantifizierung der durch Michelis und Müller [45] beschriebenen Interaktionsphasen. Von Interesse war hierbei vor allem die sogenannte Konversionsrate⁹, der Anteil jener Nutzer, die tatsächlich in die nächste Phase übergangen (also beispielsweise auf das Display aufmerksam wurden, sich diesem näherten und dann damit interagierten). Zudem wollten die Autoren auf eine Kampagne des ERC (European Resuscitation Council¹⁰) namens European Restart A Heart Day aufmerksam machen. Diese hat zum Ziel, die Allgemeinheit über Reanimationsmaßnahmen im Falle eines Herzstillstands zu informieren und zum Umsetzen dieser Maßnahmen zu bewegen. Daher wurden die durch Michelis und Müller [45] beschriebenen Interaktionsphasen um eine weitere ergänzt. Sie bestand in der Darstellung eines Videos, welches die Anwendung entsprechender Maßnahmen erklärte.

Die Studie lief einen Tag und wurde mittels zweier Public Displays durchgeführt, welche am Hauptbahnhof Brüssel aufgestellt wurden. Um Interaktivität zu kommunizieren, wurden zwei Software-Prototypen entwickelt, welche im Wechsel zeitweise entweder eine Silhouette oder ein Spiegelbild darstellten.

Insgesamt wurden 2.348 von 10.157 Passanten in dieser Zeit auf das System aufmerksam. Dies entspricht 23,12 % und stützt die Aussage von Huang, Koster und Borchers [30] wonach ein Großteil der Menschen entsprechende Bildschirme nicht wahrnimmt (Stichwort Bildschirm-Blindheit). Im Unterschied zur Feldstudie von Müller et al. [50] hatten beide Repräsentationsformen eine ähnlich große Anzahl an Interaktionen zur Folge (10,84 % im Fall des Spiegelbildes und 9,13 % im Fall der Silhouette). Hiervon sahen lediglich 11,27 % das Instruktionsvideo. Das bedeutet, dass 11,27 % alle Interaktionsphasen durchliefen und zeigt, dass der Großteil der Interaktionen nur von sehr kurzer Dauer war beziehungsweise ist. Huang, Koster und Borchers [30] machten ähnliche Beobachtungen.

Ackad, Kay und Tomitsch [2] verglichen die Repräsentationsformen Skelett (kinesthetic matching) und Silhouette (appearance matching) miteinander. Dies geschah ebenfalls im Rahmen einer (Feld-)Studie. Über eine Dauer von 56 Tagen wurden die Interaktionen mit einem Public Display protokolliert. Man fand heraus, dass im Fall der Silhouette in signifikant mehr Sitzungen Personen mit dem Display interagierten (84 % gegenüber 70 % bei der Repräsentationsform Skelett). Gleichzeitig führte die Repräsentationsform Silhouette zu deutlich mehr Sitzungen, welche ausschließlich sachliche Interaktionen beinhalteten (70 % gegenüber 34 %).

Im Fall des Skeletts verhielt sich dies umgekehrt. Hier erfolgten mehr spielerische Interaktionen mit dem Display. Zudem waren die Sitzungen, in denen ausschließlich eine spielerische Interaktion stattfand, im Vergleich zur Silhouette deutlich länger (88,1 gegenüber 28,9 Sekunden). Den Autoren zu Folge gab es außerdem große Unterschiede zwischen den beiden Repräsentationsformen hinsichtlich spielerischer Interaktionen. Im Fall des Skeletts taten beispielsweise einige Nutzer so, als würden sie boxen. Andere wiederum tätschelten den virtuellen Kopf ihres Skeletts. In Bezug auf die Silhouette beschränkten sich spielerische Interaktionen auf die Nutzung unterschiedlicher Gesten, die wiederum dazu dienten, durch die dargestellten Inhalte zu navigieren.

Diese Erkenntnisse werden, vor allem in Bezug auf die Darstellung des Skeletts, durch verschiedene Arbeiten [71, 1] gestützt.

Tomitsch et al. [71] fanden beispielsweise heraus, dass während länger andauernder Interaktionsphasen mit einem Public Display oft nur wenige Gesten ausgeführt wurden. Stattdessen wurde in

⁹engl. conversion rate

¹⁰Europäische Arbeitsgruppe zur Reanimation [84]

vielen Fällen ein spielerisches Verhalten beobachtet. Die Autoren schreiben dies unter anderem der eingesetzten Nutzerrepräsentation zu. Es handelte sich auch hierbei um die Darstellung des Skeletts.

Auch Ackad et al. [1] konnten bereits in einer früheren Studie zeigen, dass die Darstellung des Skeletts die Verspieltheit fördert.

Die oben genannten Arbeiten haben gezeigt, dass Nutzerrepräsentationen eine interessante Technik darstellen, um potentielle Nutzer auf Public Displays aufmerksam zu machen und diese zur Interaktion zu bewegen. Die Frage ist, ob sich Art von Repräsentationen auch dazu eignet Nutzer in den Sweetspot zu führen. Voraussetzung hierfür ist jedoch, dass Nutzer ihre entsprechenden Repräsentationen auch erkennen können. Vor allem in diesem Zusammenhang besteht noch Forschungsbedarf. Beispielsweise konnte bisher noch nicht geklärt werden, welche Strategien Nutzer anwenden, um ihre Repräsentationen zu erkennen und inwiefern sich diese Strategien zur Erkennung unterschiedlicher Repräsentationen eignen. Mittels zweier in dieser Arbeit beschriebener Studien sollten Antworten auf diese Fragen gefunden werden.

Der wesentliche Unterschied dieser Arbeit zu bereits existierenden Arbeiten liegt darin, dass:

- die Repräsentationsformen Spiegelbild, Silhouette, Skelett und das Quadrat als abstraktes Objekt miteinander verglichen wurden. Ein Vergleich in dieser Zusammensetzung wurde bisher noch nicht durchgeführt.
- der Vergleich der Repräsentationsformen mit der größtmöglichen Anzahl gleichzeitiger Teilnehmer (7 Teilnehmer in der Hauptstudie) durchgeführt wurde, um untersuchen zu können, inwiefern sich mehrere gleichartige Repräsentationen, die gleichzeitig angezeigt werden, auf die Erkennung der eigenen Repräsentation auswirken.

Die oben erwähnten Studien werden in den Kapiteln 4 und 5 näher betrachtet. Zunächst erfolgt jedoch die Erläuterung eines Software-Prototyps, welcher benötigt wurde, damit die Studien durchgeführt werden konnten.

3 Implementierung

Um die bereits beschriebenen Nutzerrepräsentationen (Quadrat, Skelett, Silhouette und Spiegelbild) im Rahmen der Vor- beziehungsweise Hauptstudie testen zu können, wurde ein Software-Prototyp entwickelt. Im nun folgenden Kapitel wird näher auf dessen Aufbau und Funktionsweise eingegangen. Des Weiteren wird beschrieben, wie Kinect, was essentiell für die Erfassung und Darstellung der Nutzerrepräsentationen war, in den Entwicklungsprozess eingebunden wurde.

3.1 Aufbau

Die Entwicklung des im Rahmen dieser Arbeit entstandenen Software-Prototyps erfolgte in vier Schritten. Im ersten Schritt befasste man sich mit der optischen Erkennung sich bewegender Nutzer und Verarbeitung der durch Kinect gelieferten Bildinformationen. Im zweiten darauf aufbauenden Schritt stand die Nutzung der von Kinect gelieferten Bildinformationen und die Darstellung der festgelegten Nutzerrepräsentationen im Mittelpunkt. Es folgten die technische Umsetzung der Erfassung und Verarbeitung von Eingaben sowie die Protokollierung von Messwerten. Nachfolgend wird beschrieben wie diese Schritte technisch umgesetzt wurden.

Zunächst erfolgt jedoch eine Auseinandersetzung mit Kinect, um ein grundlegendes Verständnis für die technische Umsetzung der ersten beiden Schritte entwickeln zu können.

3.2 Kinect

Kinect ist ein von Microsoft entwickeltes Kamerasystem [95]. Im Gegensatz zu herkömmlichen Videokameras ist Kinect mit einem Tiefen-Sensor ausgestattet, der es ermöglicht Tiefen-Informationen in Bildern zu erfassen. Unter Zuhilfenahme verschiedener bildverarbeitender Verfahren können diese Informationen genutzt werden, um den menschlichen Körper beziehungsweise die dem Skelett zu Grunde liegenden Gelenke im Bild zu erkennen und als Eingabewerkzeug für verschiedene Arten von Software zu nutzen. Dies wurde in der Vergangenheit vor allem genutzt, um Videospiele - ähnlich den Systemen Wii-Fernbedienung (Plus) [99] von Nintendo und PS Move [100] von Sony - durch Körperbewegungen steuern zu können. Daher wurde Kinect zunächst auch nur als Zubehör für die Microsoft-eigenen Konsolen Xbox 360 und Xbox One hergestellt.

Durch Freigabe entsprechender SDKs¹¹ [96, 97] (siehe Abschnitt SDK) und Laufzeitbibliotheken ist es seit einiger Zeit jedoch auch möglich, Kinect für die Entwicklung Windows-basierter Software zu nutzen. Im Folgenden wird näher auf den Aufbau sowie die Spezifikationen von Kinect eingegangen.

3.2.1 Aufbau und Spezifikationen

Wie bereits erwähnt verfügt Kinect über einen Sensor zur Erfassung von Tiefen-Informationen in Bildern. Dieser Sensor besteht aus zwei Komponenten, welche in der Front des Gehäuses eingelassen sind. Zum einen handelt es sich hierbei um einen Infrarot-Emitter beziehungsweise -Projektor, zum anderen um eine Infrarot-Kamera. Letztere kann durch den Infrarot-Projektor ausgestrahltes und von Objekten reflektiertes Licht wahrnehmen. Dadurch kann bestimmt werden, ob sich Objekte in Sichtweite von Kinect befinden und wie weit diese entfernt sind. Durch die Nutzung des Infrarot-Lichts funktioniert dies sowohl in hellen als auch in dunklen Umgebungen. Die Auflösung der Infrarot-Kamera der aktuellen zweiten Kinect-Generation beträgt 512×424

¹¹Software-Development-Kit. Es handelt sich um ein Softwarepaket für Entwickler, welches alle notwendigen Bestandteile enthält, um auf Software- oder auch Hardwarekomponenten entsprechender Systeme zuzugreifen.

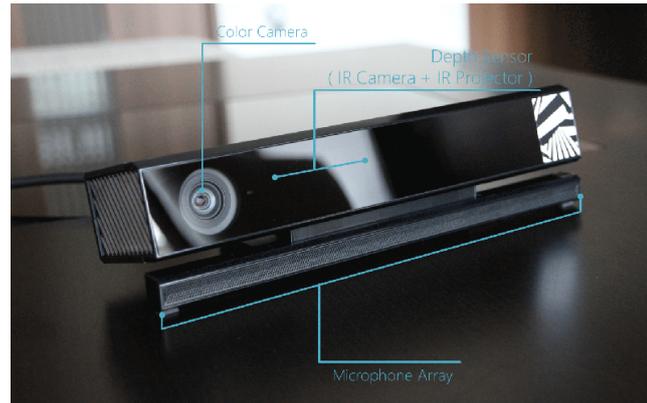


Abbildung 3.1: Übersicht über Ausrichtung und Lage der Kinect-Sensoren (aktuelle Generation) [88]

Pixel. Die Erfassung erfolgt mit 30 Hz. Die Größe des Sichtfelds wird mit 70° in der Breite und 60° in der Höhe bemessen [95].

Wie aus Abbildung 3.1 ersichtlich ist, verfügt Kinect zudem über eine Farb-Kamera (RGB) und ein Mikrofon-Array mit vier Mikrofonen [79]. Letzteres dient unter anderem der Spracheingabe und -erkennung. Die RGB-Kamera kann einerseits für alltägliche Dinge wie Videotelefonie [95] oder Live-Streaming [101] verwendet werden. Andererseits können die von ihr gelieferten Bilder aber auch mit denen des Tiefen-Sensors verrechnet werden, um so verschiedene Nutzerrepräsentationen wie beispielsweise Spiegelbilder zu erzeugen. Die Bilder der RGB-Kamera haben eine Auflösung von 1080 Pixel in der Höhe und 1920 Pixel in der Breite. Aufnahmen erfolgen ausschließlich in Vollbildern (progressiv) und bei 30 Hz beziehungsweise 15 Hz (je nach Menge des zur Verfügung stehenden Lichts) [95].

3.2.2 Unterschiede zur ersten Generation

Wie bereits angedeutet wurde, unterscheiden sich Kinect der ersten und zweiten (aktuellen) Generation hinsichtlich verschiedener technischer Eigenschaften. Weshalb diese Eigenschaften im Rahmen dieser Arbeit relevant sind und daher die Kinect der zweiten Generation für die Durchführung der beschriebenen Studien gewählt wurde, wird im Folgenden erläutert.

Anzahl der gleichzeitig erfassbaren Skelette

Betrachtet man die technischen Daten beider Kinect-Generationen, fällt auf, dass Kinect der zweiten Generation deutlich mehr Skelette gleichzeitig erfassen kann als Kinect der ersten Generation (6 gegenüber 2) [95]. Dies war im Kontext der Arbeit insofern wichtig, als die Teilnehmer der Hauptstudie (siehe Kapitel 5) gleichzeitig zwischen mehreren gleichartigen Nutzerrepräsentationen unterscheiden sollten.

Größe des Sichtfelds

Mit der Anforderung, möglichst viele Skelette gleichzeitig erfassen zu können, stiegen automatisch auch die Anforderungen an die Größe des Sichtfelds. Da die aktuelle Kinect-Generation auch in diesem Fall deutliche Vorteile bietet ($70^\circ \times 60^\circ$ gegenüber $57^\circ \times 53^\circ$) [95, 90], fiel die Wahl auf diese.

3.2.3 Programmierung mittels Kinect

Wie in Kapitel 3.2 bereits erwähnt stellt Microsoft für die Entwicklung mit Kinect SDKs bereit. Diese enthalten unter anderem sogenannte APIs¹², die den softwareseitigen Zugriff auf die in Kinect verbauten Mikrofone und Kameras ermöglichen. Wie diese SDKs aussehen und wie sie genutzt werden können, um auf Kinect zuzugreifen, wird im Folgenden erläutert.

Frame-Source-Typ	Auflösung	Information
BodyFrameSource	keine	räumliche Position und Ausrichtung der 25 Gelenke für max. 6 Skelette/Körper, außerdem: Hand-Status (geöffnet/geschlossen/Lasso) für max. 2 Skelette/Körper und Torso-Neigungsinformationen
BodyIndexFrameSource	512 px × 424 px	Wert zwischen 0 und 5, der für jeden Pixel angibt, zu welchem Körper/Skelett er gehört. Ist der Wert größer 5, gehört der entsprechende Pixel zum Hintergrund.
ColorFrameSource	1920 px × 1080 px	2D-Farbbild
DepthFrameSource	512 px × 424 px	Abstand von Kinect-Brennebene in Millimeter
InfraredFrameSource	512 px × 424 px	Infrarot-Intensitätswerte

Tabelle 3.1: Übersicht der FrameSource-Typen und Eigenschaften der von diesen zur Verfügung gestellten Frames [79].

SDK, Frames und FrameSources

Die von Microsoft angebotenen SDKs lassen sich in zwei Gruppen unterteilen: Alle SDKs mit Version 1.8 und darunter wurden für die erste Kinect-Generation entwickelt. Sie sind mit der zweiten Generation nicht kompatibel und daher an dieser Stelle nicht von Relevanz.

Betrachtet wird in dieser Arbeit ausschließlich das SDK in Version 2.0. Dieses ermöglicht die Entwicklung nativer Software wie auch solcher für Microsofts .NET-Plattform und den Windows Store. Unterstützte Programmiersprachen sind unter anderem C# und C++ [79]. Im praktischen Teil dieser Arbeit sowie in den folgenden Codebeispielen wurde ausschließlich C# Code verwendet.

Das Kinect-SDK in Version 2.0 setzt sich unter anderem aus den APIs Face, Kinect.Input und Kinect zusammen [91]. Ersteres ermöglicht neben der Erkennung von Gesichtern auch deren Erfassung innerhalb eines Videostroms sowie deren virtuelle dreidimensionale Modellierung [92]. Kinect.Input bietet Schnittstellen an, um Kinect als Maus-Ersatz zu verwenden und ein (Windows)-System durch Handbewegungen oder beispielsweise Gesten zu steuern [94]. Kinect stellt das im Rahmen dieser Anwendung relevanteste API dar. Es werden Schnittstellen, sogenannte FrameSources, bereitgestellt, um auf verschiedenartige **Frames** zuzugreifen [93]. Bei letzteren handelt es sich um Einzelbilder aus den von Kinect zur Verfügung gestellten Videoströmen [91]. Insgesamt gibt es sieben unterschiedliche **Frame-Typen** sowie **FrameSources**. Tabelle 3.1 benennt letztere und stellt Auflösungen sowie Informationen zu den entsprechenden Frame-Typen dar.

¹²Application Programming Interface. Softwareseitige Programmierschnittstelle für Entwickler um beispielsweise auf Software- oder Hardware-Komponenten zuzugreifen.

Frame-Zugriff

Die Frage ist, wie auf die erwähnten Frames zugegriffen und die gewünschten Informationen (wie beispielsweise die Position des Skeletts oder einzelner Gelenke, Entfernung oder Körpergröße) abgefragt werden können.

Bevor diese Frage beantwortet werden kann, muss man sich zunächst mit verschiedenen in der Kinect-API enthaltenen Klassen und Methoden auseinandersetzen. Die Klasse **KinectSensor** nimmt dabei den größten Stellenwert ein. Sie repräsentiert einen physischen Kinect-Sensor. Über die Klassenmethoden **Open()** und **Close()** können alle angebotenen Videoströme geöffnet beziehungsweise geschlossen werden. Gleichzeitig stellt KinectSensor über sogenannte FrameSources **FrameReader** bereit. Letztere ermöglichen den Zugriff auf die gewünschten Frames beziehungsweise Frame-Typen. FrameSources stehen stellvertretend für die jeweiligen Videoströme. Für jeden Frame-Typ existieren daher auch eine zugehörige FrameSource sowie ein zugehöriger FrameReader. Soll beispielsweise ein InfraredFrame gelesen werden, wird zunächst auf die InfraredFrameSource zugegriffen. Diese gibt, wie in Codebeispiel 1 zu sehen, über die Klassenmethode **OpenReader()** ein Objekt der Klasse InfraredFrameReader zurück.

```

1 KinectSensor sensor = KinectSensor.GetDefault();
2 sensor.Open();
3 InfraredFrameReader reader = sensor.InfraredFrameSource.OpenReader();

```

Codebeispiel 1: Erstellen des KinectSensors- und InfraredFrameReader-Objekts

Der Zugriff auf das InfraredFrame erfolgt Event-basiert über eine FrameArrived-Methode. Diese wird an das InfraredFrameArrived-Event des InfraredFrameReaders gebunden und immer aufgerufen, sobald ein neues Frame zur Verfügung steht. Abhängig vom gewählten Frame-Typ und den Lichtverhältnissen erfolgt ein Aufruf entweder jede 1/15- oder jede 1/30-Sekunde. Siehe hierzu auch Codebeispiel 2.

```

1 reader.InfraredFrameArrived += this.Reader_InfraredFrameArrived;

```

Codebeispiel 2: (Infrared)FrameArrived-Methode wird an (Infrared)FrameArrived-Event des Readers gebunden

Wie in Codebeispiel 3 zu sehen ist, kann innerhalb der FrameArrived-Methode mittels **AcquireFrame()** auf das Frame beziehungsweise dessen **FrameReference** zugegriffen werden. Dies sollte allerdings unter Zuhilfenahme des **using-Patterns** erfolgen. Dadurch wird sichergestellt, dass für das Frame allokierte System-Ressourcen wieder freigegeben werden, sobald diese nicht mehr benötigt werden. Das angeforderte Frame sollte zudem auf Null getestet werden, da es passieren kann, dass dieses bereits abgelaufen ist, bevor die Methode aufgerufen wird [79]. Wurden ein oder mehrere Frames ausgelesen, können diese verwendet werden, um beispielsweise Nutzerrepräsentationen zu erzeugen. Wie dies in dieser Arbeit umgesetzt wurde, wird in Kapitel 3.3 erläutert.

```

1 void Reader_InfraredFrameArrived(InfraredFrameReader sender,
2 InfraredFrameArrivedEventArgs args)
3 {
4     using (InfraredFrame frame = args.FrameReference.AcquireFrame()){
5         if (frame != null){
6             //Weiterverarbeitung
7         }
8     }
9 }

```

Codebeispiel 3: Zugriff auf das gewünschte (Infrared)Frame mittels FrameArrived-Methode

3.3 Software-Prototyp

Wie bereits zu Beginn des Kapitels 3 beschrieben gliedert sich die Entwicklung des im Rahmen dieser Arbeit entstandenen Software-Prototyps in fünf Schritte. Im Folgenden wird auf diese und die damit verbundenen Herausforderungen genauer eingegangen.

Optische Erfassung der Nutzer und Generierung der Darstellungen

Um die verschiedenen Nutzerrepräsentationen darstellen zu können, mussten zunächst Wege gefunden werden, spätere Studienteilnehmer mittels Kinect optisch zu erfassen. Je nach Repräsentation wurden dabei unterschiedliche Ansätze gewählt. Die Vorgehensweise wird exemplarisch anhand der Repräsentationen **Quadrat** und **Silhouette** vorgestellt.

Zur Erinnerung: Das **Quadrat** ist eine Nutzerrepräsentation, welche aus der Top-Down-Perspektive, die absolute Position des Nutzers im Raum wiedergibt. Zur korrekten Platzierung des Quadrats auf einer vertikalen Fläche (dem Bildschirm) war es daher nötig, den X- und Z-Wert der räumlichen Position des Nutzers auf den X- und Y-Wert des Quadrats abzubilden. Das bedeutet, dass sich das Quadrat nach oben beziehungsweise nach unten bewegte, sofern der Nutzer einen Schritt nach vorne beziehungsweise hinten machte. Die Abbildung entlang der Horizontalen verhielt sich analog zur Bewegungsrichtung des Nutzers.

Konkret erfolgte die Abbildung beziehungsweise Positionsberechnung des Quadrats über die in Codebeispiel 4 beschriebenen Funktionen.

```

1 screenPositionX = (this.canvas.ActualWidth / 2) +
  → ((this.canvas.ActualWidth / 2f) * userPositionX / (MAX-SENSOR-RANGE /
  → 2f));
2 screenPositionY = canvas.ActualHeight * ((userPositionY -
  → MIN-SENSOR-RANGE) / (MAX-SENSOR-RANGE - MIN-SENSOR-RANGE));

```

Codebeispiel 4: Berechnung der X- und Y-Position des Quadrats. `canvas.ActualHeight` beziehungsweise `canvas.ActualWidth` geben Höhe und Breite der Zeichenfläche zurück. Bei `MIN-` und `MAX-SENSOR-RANGE` handelt es sich um im Programm festgelegte Werte, die angeben, wie groß der Abstand zur Kinect mindestens sein muss und maximal sein darf, damit diese den Nutzer zuverlässig erfassen kann.

Um die oben beschriebene Abbildung zu ermöglichen, war es zunächst notwendig, die räumliche Position des Nutzers zu bestimmen. Dazu wurde kontinuierlich auf das aktuellste

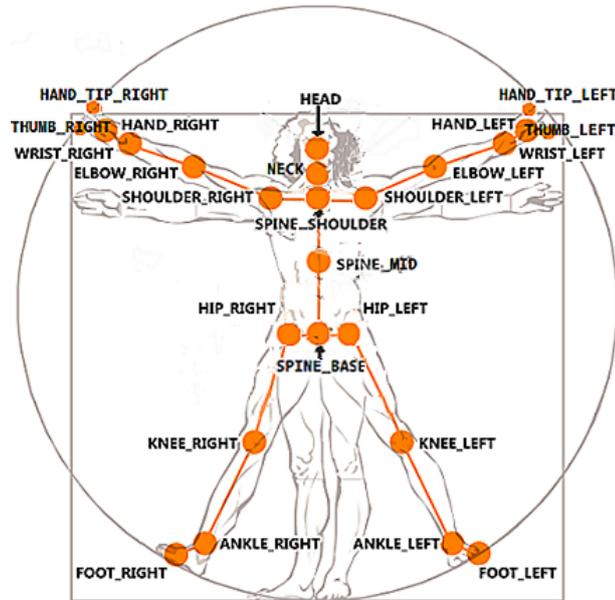


Abbildung 3.2: Die von Kinect der zweiten Generation erkannten Gelenke beziehungsweise Joins [89]

BodyFrame zugegriffen und das darin enthaltene Body-Array ausgelesen. Wie in Tabelle 3.1 zu sehen ist, beinhaltet dieses alle notwendigen Informationen der von Kinect momentan erfassten Skelette. Wichtig hierbei sind vor allem die aktuellen räumlichen Positionen der einzelnen Gelenke. Welche dieser von Kinect erkannt werden, zeigt Abbildung 3.2.

Da sich das Gelenk **SPINE BASE** (unteres Ende der Wirbelsäule) horizontal gesehen in der Körpermitte befindet und daher der räumlichen Position des Nutzers am nächsten kommt, wurde es ausgewählt und dessen Position mit jener des Nutzers gleichgesetzt. Dies war möglich, da für die obige Abbildung lediglich X- und Z-Wert der räumlichen Position des Nutzers relevant waren. Die Auswahl dieses Gelenks hatte aber noch einen weiteren Grund. Wie aus Kapitel 5 hervorgeht, bestanden zwei der Bedingungen der Hauptstudie darin, entweder ganz still zu stehen oder nur die Arme zu bewegen. Es war von großer Wichtigkeit, dass sich das Quadrat in diesen Fällen nicht bewegte, da die Unterscheidung zwischen Aufnahme und Live-Darstellung sonst vereinfacht worden wäre. Wären andere Gelenke (wie zum Beispiel **NECK** (Genick) oder **SPINE MID** (Mitte der Wirbelsäule)) gewählt worden, hätte dies jedoch zu entsprechenden Bewegungen des Quadrats geführt, da bei Armbewegungen diese Teile des Körpers oft unwillkürlich mitbewegt werden.

Im Fall der Nutzerrepräsentation **Silhouette** gestaltete sich die optische Erfassung der Teilnehmer etwas anders. Hier wurde das **BodyIndexFrame** genutzt, um einerseits entsprechende Nutzer erfassen und andererseits zwischen diesen und dem Hintergrund unterscheiden zu können. Konkret liefert das **BodyIndexFrame** Informationen darüber, ob ein bestimmter Pixel zu einem Nutzer/Skelett gehört oder nicht. Um die Silhouette zu erzeugen, mussten die Daten der **BodyIndexFrames** durchlaufen und die Pixel abhängig von ihrer Zugehörigkeit in einer bestimmten Farbe eingefärbt werden. Hierbei empfiehlt es sich, mittels **usage-Pattern** direkt auf den **KinectBuffer** des **BodyIndexFrames** zuzugreifen, um so die Abarbeitung zu beschleunigen.

Darstellung der festgelegten Nutzerrepräsentationen

Die Darstellung der Nutzerrepräsentationen erfolgte entweder als **Image** (Silhouette, Skelett und Spiegelbild) oder **Rectangle** (Quadrat). Entsprechende Objekte wurden der von **WPF**¹³ bereitgestellten .xaml-Datei hinzugefügt. Dadurch konnten Repräsentationen, die gerade nicht angezeigt werden sollten, durch ändern der Eigenschaft `Visibility` ausgeblendet werden.

Einen wichtigen Aspekt in Bezug auf die Darstellung stellte auch der sogenannte `Offset` dar. Damit wurde im Laufe der Entwicklung eine Funktion bezeichnet, welche es ermöglichte, die aktuelle Nutzerrepräsentation um einen zufälligen Wert auf der X-Achse zu verschieben. Diese Funktion war aufgrund der in Kapitel 4 beschriebenen `Positions-Strategie` notwendig geworden. Da es sich hierbei um eine besondere Herausforderung handelte, wird auf diese im Abschnitt `Herausforderungen` näher eingegangen.

Verarbeitung von Eingaben

Um im Rahmen der Hauptstudie ermitteln zu können, wie gut die Nutzerrepräsentationen erkannt werden, war es nötig, die Zeit bis zur Erkennung zu messen. Dazu wurde ein Xbox 360 Wireless Controller [87] genutzt. Wie in Kapitel 5 erklärt sollten die Teilnehmer einen Button des Controllers drücken, sobald sie die Repräsentation erkannten. Wie die Kommunikation mit dem Controller und der softwareseitige Zugriff auf eventuell gedrückte Buttons erfolgte, wird nun erklärt.

Zur Kommunikation mit dem Controller wurde die Bibliothek **XInputDotNet** [85] eingesetzt. Diese kapselt die von Microsoft angebotene `XInput`-Bibliothek [98] und kann - im Gegensatz zum Pendant von Microsoft - auch unter `WPF` eingesetzt werden, was für diese Arbeit von besonderer Relevanz war.

`XInputDotNet` ermöglicht den Zugriff auf den Controller über ein **GamePad**- sowie **GamePadState**-Objekt. Letzteres repräsentiert den aktuellen Status des Controllers, gibt also Auskunft darüber, welche Buttons gedrückt werden. Das `GamePad`-Objekt steht - ähnlich wie bei der Entwicklung mit `Kinect` - für ein physisches Gerät, in diesem Fall für den Controller. Wie bereits in Kapitel 5 beschrieben, konnten sich die Teilnehmer zu Beginn der Studie einen Button auf dem Controller aussuchen, den sie drückten, sobald sie die Repräsentation erkannten. Wäre dem nicht so gewesen, hätten die Teilnehmer den entsprechenden Button vor dem Drücken eventuell erst suchen müssen, wodurch die Messwerte möglicherweise verzerrt worden wären. Für die Implementierung bedeutete dies, dass geprüft werden musste, ob überhaupt ein Button auf dem Controller gedrückt wurde. Da `XInputDotNet` hierfür keine Schnittstelle bereitstellt, wurde das, wie in Codebeispiel 5 ersichtlich, über einen Vergleich der sogenannten `PacketNumbers` realisiert, welche angeben, ob sich der `GamePadState`, also der Status des Controllers verändert hat.

Eingabe der Bedingungen

Es folgte die Eingabe der zu testenden Bedingungen. Da in nächster Zeit keine Studien mit anderen Repräsentationen zu erwarten waren und diese sowieso eine Erweiterung und/oder Änderung der Anwendung zur Folge hätten, entschied man, die Bedingungen direkt im Code abzulegen. Dies erfolgte in Form einer Liste von `Tripeln`. Jedes `Tripel` beinhaltete einen `String`, `boolean` und `enum`.

¹³Windows Presentation Foundation. Grafik-Framework und Teil des von Microsoft entwickelten .NET-Frameworks.

```

1 GamePadState gamePadState = GamePad.GetState(PlayerIndex.One);
2 if (gamePadState.PacketNumber != this.previousGamePadState.PacketNumber){
3     //PacketNumbers sind unterschiedlich --> Button wurde gedrückt
4     //--> Messwert protokollieren
5 }

```

Codebeispiel 5: PacketNumbers zweier GamePadStates werden verglichen, um herauszufinden, ob ein Button gedrückt wurde beziehungsweise sich der GamePadState verändert hat.

Der String diente zur Codierung der vier Strategien. Er wurde einerseits genutzt, um die Strategien einfacher im Software-Prototypen hinterlegen zu können und andererseits, um die Auswertung der Daten zu vereinfachen. Konkret konnte er in jeder Bedingung einen von vier Werten annehmen:

- 00 (Strategie: *still stehen*)
- 01 (Strategie: *Arme bewegen, sonst still stehen*)
- 10 (Strategie: *herumlaufen, Arme nicht bewegen*)
- 11 (Strategie: *herumlaufen und Arme bewegen*)

Der Boolean gab Auskunft darüber, ob der Offset aktiv sein sollte (true) oder nicht (false). Das Ende des Tripels bestand immer aus dem bereits erwähnten enum. Dieser konnte ebenfalls in jeder Bedingung einen von vier Werten annehmen, welcher die vier Repräsentationsformen Spiegelbild, Silhouette, Skelett und abstraktes Objekt (Quadrat) beschrieb.

Protokollierung von Messwerten

Der letzte Schritt, die Protokollierung der Messwerte während der Hauptstudie, war wesentlich für die spätere Auswertung der Daten (siehe Kapitel 5.5.2).

Protokolliert wurden neben der Zeit, die zur Erkennung der Repräsentation benötigt wurde, auch die aktuelle Bedingung (Nutzerrepräsentation, Offset - aktiv oder inaktiv - und Strategie), die Teilnehmer-ID sowie die Angabe, ob die Repräsentation richtig erkannt wurde oder nicht.

Während der Hauptstudie wurden diese Angaben unter Zuhilfenahme der **StreamWriter**-Klasse in eine .csv-Datei geschrieben.

3.3.1 Herausforderungen

Die Herausforderungen während der Entwicklung bestanden zum einen in der Optimierung der Anwendung hinsichtlich Ressourcennutzung und zum anderen, wie bereits angesprochen, in der Realisierung des Offsets. Beide Punkte werden nun genauer erläutert.

Die Optimierung der Anwendung beziehungsweise des Programmcodes spielte während der Entwicklung eine wichtige Rolle, da jederzeit sichergestellt sein musste, dass die Nutzerrepräsentationen - entsprechende Hardware vorausgesetzt - flüssig dargestellt werden. Dies bedeutete einerseits, dass Systemressourcen, die von Kinect in Anspruch genommen wurden (wie beispielsweise Arbeitsspeicher), so schnell wie möglich wieder freigegeben werden mussten. Dazu wurden das bereits beschriebene usage-Pattern sowie der Methodenaufwurf **Dispose()** genutzt. Dieser dient dazu, nicht mehr benötigte Frames zu verwerfen. Insgesamt konnte so die Auslastung des Arbeitsspeichers durch den Software-Prototyp dauerhaft auf etwa 120 MB gehalten werden. Funktionen, die gleichzeitig zur Erfassung durch Kinect ablaufen sollten, wie beispielsweise die

Kommunikation mit dem Controller oder eine Stoppuhr zur Zeitmessung, wurden in eigenen Threads gekapselt.

Für die Hauptstudie wesentlich war auch die Umsetzung des Offsets. Um bei der Verschiebung der Live-Darstellung eine zusätzliche Auslastung der CPU zu vermeiden, erfolgte diese nicht auf Pixel- sondern auf Bildebene. Das heißt, das Image-Objekt der aktuellen Repräsentation wurde durch den Methodenaufruf **Canvas.Left()** verschoben. Wichtig hierbei war, innerhalb der Darstellungsgrenzen zu bleiben; die Repräsentation sollte nicht so weit nach rechts oder links verschoben werden, dass sie in Teilen oder im Ganzen nicht mehr sichtbar war. Um dies zu gewährleisten, wurden umfangreiche Tests durchgeführt und ein Regelset implementiert, welches entsprechende Parameter prüft, bevor eine Verschiebung durchgeführt wird.

Eine weitere technische Herausforderung bestand in der Aufnahme der Nutzerrepräsentationen. Trotz softwareseitiger Optimierungen gelang es zunächst nicht, die Nutzerrepräsentationen flüssig darzustellen und gleichzeitig in der höchstmöglichen Qualität aufzuzeichnen. Daher wurde ein sogenannter HDMI-Grabber genutzt. Hierbei handelt es sich um ein Gerät, welches Videomaterial mittels eines eingebauten H.264-Encoders in nahezu Echtzeit encodieren kann. Dies hatte zur Folge, dass die CPU entlastet und für andere Aufgaben genutzt werden konnte (wie beispielsweise die Darstellung von Nutzerrepräsentationen). Konkret wurde das Modell LGP Lite [78] des Herstellers AVerMedia eingesetzt. Damit wurden Aufnahmen mit einer Bitrate von 60 Mbit/s ermöglicht. Rein optisch war somit kein Unterschied zwischen Live-Darstellung und Aufnahme zu erkennen. Dies war vor allem im Kontext der Hauptstudie wichtig, deren Relevanz verloren gegangen wäre, hätten die Teilnehmer der Studie die Live-Darstellung anhand eines Qualitätsunterschieds identifizieren können.

4 Vorstudie

Bevor die in Kapitel 5 beschriebene Hauptstudie durchgeführt werden konnte, mussten in einer explorativen Vorstudie zunächst grundlegende Erkenntnisse im Umgang mit Nutzerrepräsentationen gewonnen werden. Welche Ziele die Vorstudie genau verfolgte, wird im folgenden Abschnitt beschrieben.

4.1 Ziele

Die Ziele der Vorstudie bestanden vor allem darin, herauszufinden, welche Strategien Nutzer anwenden, wenn sie ihre eigene Nutzerrepräsentation auf einem Public Display identifizieren müssen und inwiefern sich diese Strategien zur Identifikation unterschiedlicher Repräsentationen eignen. Des Weiteren sollte in Erfahrung gebracht werden, ob diese Strategien bewusst genutzt werden und ihr Einsatz womöglich geplant oder eher spontan erfolgt.

Von großer Bedeutung war außerdem die so bezeichnete Positions-Strategie. Es wurde angenommen, dass ein Nutzer seine Repräsentation erkennen kann, wenn er seine Position mit den Positionen der Repräsentationen auf einem Display vergleicht. Ein Zweck der Vorstudie bestand auch darin, diese Annahme zu bestätigen oder zu widerlegen.

4.2 Nutzerrepräsentationen

Im Rahmen dieser Arbeit wurden vier unterschiedliche Formen von Nutzerrepräsentationen zur genaueren Betrachtung herangezogen: das Spiegelbild, die Silhouette, das Skelett sowie das Quadrat.

Bevor genauer auf den Verlauf der Vorstudie eingegangen wird, werden zunächst die genannten Repräsentationsformen im Detail beschrieben und es wird begründet, weshalb gerade diese ausgewählt wurden.

Spiegelbild

Das Spiegelbild zeigt eine zweidimensionale, texturierte Farb-Darstellung des Nutzers, das heißt das Abbild des Nutzers wird so dargestellt, als wäre er mit einer herkömmlichen Videokamera aufgezeichnet worden. Details wie Kleidung, Gesichtszüge, Haare oder auch Accessoires wie Ohringe sind sichtbar. Der einzige Unterschied zur Aufnahme mit einer herkömmlichen Videokamera besteht darin, dass der Hintergrund entfernt und stattdessen eine gleichmäßig schwarze Fläche zu sehen ist. Dadurch wurde die Darstellung auf die Repräsentation des Nutzers reduziert und dieser konnte im Zuge der Studien nicht durch Objekte im Hintergrund abgelenkt werden.

Silhouette

Bei der Silhouette handelte es sich um eine schemenhafte Darstellung des Nutzers. Diese zeigte dieselben Konturen wie das Spiegelbild, eine Texturierung war jedoch nicht vorhanden. Stattdessen wurden alle Pixel, die der Repräsentation des Nutzers zugeordnet wurden, mit derselben Farbe eingefärbt. Auch hier wurde der Hintergrund entfernt.

Skelett

Charakteristische Eigenschaft der Repräsentationsform Skelett war, dass diese ausschließlich die Positionen der von Kinect erfassten Gelenke (mehr dazu in Kapitel 3.3) darstellte. Diese wurden durch Linien miteinander verbunden, was den Eindruck des Skeletts vermittelte.

Quadrat

Das Quadrat war die einzige Repräsentationsform, welche eine Ansicht von oben realisierte. Ein Quadrat, welches auf schwarzem Grund dargestellt wurde, zeigte die absolute Position des Nutzers im Raum. Bewegte sich dieser nach vorne, reagierte das Quadrat mit einer Bewegung nach oben. Bewegte sich der Nutzer nach hinten, bewegte sich das Quadrat nach unten. Die Bewegungen nach links/rechts verliefen analog. Im Gegensatz zu Spiegelbild, Silhouette und Skelett reagierte das Quadrat also ausschließlich auf Bewegungen des ganzen Körpers in eine Richtung; Bewegungen einzelner Gliedmaßen wurden nicht abgebildet.

Die genannten Repräsentationsformen wurden aus mehreren Gründen gewählt. Zum einen handelt es sich um die vielversprechendsten Repräsentationsformen (siehe dazu Kapitel 2.6). Zum anderen fand eine Untersuchung in dieser Kombination in der Literatur noch nicht statt. Hinzu kommt, dass sich die beschriebenen Repräsentationsformen stark in ihrem jeweiligen Abstraktionsgrad unterscheiden. Außerdem werden zwei der Repräsentationsformen - konkret das Spiegelbild und die Silhouette - in die Klasse appearance matching eingeordnet; die beiden übrigen in die Klasse kinesthetic matching. Durch die Wahl der genannten Repräsentationsformen konnte also eine gleichmäßige Verteilung auf die beiden Klassen vorgenommen werden. Dies war vor allem im Kontext der Studien wichtig, da hier kein Ungleichgewicht (beispielsweise in Bezug auf die anwendbaren Strategien) herrschen sollte.

4.3 Aufbau

Die Vorstudie bestand aus zwei Teilen. Während des ersten Teils wurden die Teilnehmer gebeten, aus einer Reihe jeweils gleichartiger Nutzerrepräsentationen ihre Repräsentation zu identifizieren. Während des zweiten Teils wurde ein semistrukturiertes Interview geführt. Der Zweck dieses Interviews bestand darin, herauszufinden, welche Strategien die Teilnehmer im ersten Teil der Studie anwandten.

Die Teilnahme erfolgte in 3 Gruppen von jeweils 4 Personen. Dies hatte mehrere Gründe. Zum einen wäre die Erkennung der eigenen Repräsentation trivial gewesen, hätte man die Vorstudie mit Einzelpersonen durchgeführt. Dadurch wäre es kaum möglich gewesen, Erkenntnisse über angewandte Strategien zu erlangen. Zum anderen erhoffte man sich durch Gruppen von Teilnehmern eine diversifizierte Menge an Strategien sammeln zu können. Außerdem wurde angenommen, dass manche Strategien auf dem Verhalten und der Bewegung anderer Nutzer basieren und daher nur zum Vorschein kommen, wenn sich mehrere Nutzer gleichzeitig vor dem Display aufhalten.

Die Studie fand in einem Raum der Arbeitsgruppen Medieninformatik und Mensch-Maschine-Interaktion statt, welcher im Vorfeld an die Bedürfnisse der Studie angepasst wurde. Dies umfasste das Entfernen von Objekten, die die Bewegungsfreiheit der Teilnehmer hätten einschränken können. Bereiche, in denen sich Forschungsprojekte befanden, wurden mittels Absperrbändern vom Rest des Raumes abgetrennt. Klebestreifen auf dem Boden markierten den späteren Interaktionsbereich beziehungsweise Erfassungsbereich von Kinect. An der Front des Raumes waren zwei quer gestellte Tische, Kinect und ein Kurzstanzprojektor aufgestellt. Dieser

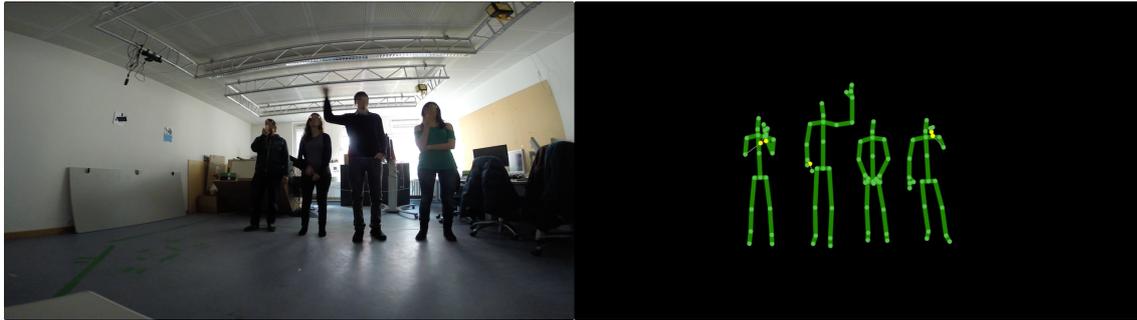


Abbildung 4.1: Ein Teilnehmer versucht durch Armbewegung seine Repräsentation zu identifizieren (eigene Grafik).

wurde während der Durchführung genutzt, um die genannten Nutzerrepräsentationen an die Wand zu projizieren. Die Projektion hatte eine Größe von etwa 2 m in der Diagonale. Sie erfolgte im Verhältnis 16:9. Die Auflösung der Projektion betrug 1280×720 Pixel. Während der Studie standen die Teilnehmer etwa 3,60 m von der Projektionsfläche entfernt.

Um Fehlern bei der Auswertung vorzubeugen, wurde der praktische Teil der Studie auf Video aufgezeichnet. Während des Interviews lief eine Audio-Aufzeichnung. Beides geschah mit dem Einverständnis der TeilnehmerInnen.

Wie gewünscht wurde die Teilnahme an der Studie entweder mit einem Amazon-Gutschein oder MMI-Punkt vergütet.

Während der Studie war außer den Teilnehmern nur der Studienleiter im Raum anwesend.

4.4 Ablauf

Zu Beginn der Studie wurden alle 4 Teilnehmer aus der jeweiligen von insgesamt 3 Gruppen vom Studienleiter begrüßt und in den dafür vorgesehenen Raum gebeten. Es folgte eine kurze Einführung. Den Teilnehmern wurde erklärt, dass auf der Projektionsfläche (Wand) sogleich vier gleichartige Nutzerrepräsentationen erscheinen würden. Ihre Aufgabe bestünde darin, ihre Repräsentation zu identifizieren. Hätten alle Teilnehmer ihre Repräsentation identifiziert, ginge man zur nächsten Repräsentationsform über. Insgesamt erfolgten 24 Durchläufe, das heißt jede Repräsentation wurde 6 Mal gezeigt. Abbildung 4.1 zeigt einen dieser Durchläufe (in diesem Fall wurde das Skelett dargestellt) mit einer der Projektionsfläche zugewandten Teilnehmergruppe.

Um die Teilnehmer im Hinblick auf potentiell anwendbare Strategien nicht zu beeinflussen, wurde ihnen nicht gesagt, wie sie ihre Repräsentation identifizieren sollten, allerdings wurden sie gebeten, sich ausschließlich in dem auf dem Boden markierten Bereich aufzuhalten. Dies war notwendig, da sie sich sonst möglicherweise aus dem Erfassungsbereich von Kinect entfernt hätten. Sie wurden außerdem darüber informiert, dass sich an den praktischen Teil der Studie ein semistrukturiertes Interview anschließen würde.

4.5 Ergebnisse

Im Folgenden werden die Ergebnisse der Vorstudie zusammengefasst. Dabei wird zunächst auf die Teilnehmer selbst und anschließend auf die von ihnen angewandten Strategien eingegangen. Schließlich erfolgt eine Auseinandersetzung mit der bereits angesprochenen Positions-Strategie und Anmerkungen der Teilnehmer.

4.5.1 Teilnehmer

Insgesamt nahmen 12 Personen an der Studie teil. Dabei handelte es sich ausschließlich um Studenten/-innen. Die Teilnehmer waren zwischen 21 und 33 Jahren alt; 7 von ihnen waren weiblich, 5 männlich.

Während des Interviews wurden die Teilnehmer auch gefragt, inwiefern sie bereits über Erfahrungen mit Bewegungssteuerungssystemen verfügen. Ein Teilnehmer gab an, durch die private Nutzung einer Xbox beziehungsweise Kinect bereits viele Erfahrungen mit einem solchen System gesammelt zu haben. Alle anderen Teilnehmer hatten nur wenige Erfahrungen damit gemacht oder waren ganz unerfahren.

4.5.2 Bewegungs-Strategien

Zunächst sei gesagt, dass die Darstellungsreihenfolge der Repräsentationen zufällig bestimmt wurde, um die Reaktion der Teilnehmer auf unterschiedliche Einstiegsrepräsentationen zu sehen. Interessant war auch herauszufinden, wie schnell die Teilnehmer bei unterschiedlichen Einstiegsrepräsentationen erkannten, dass diese auf Bewegung reagierten. Dadurch ergab sich folgende Verteilung: Die erste Gruppe begann mit dem Quadrat, die zweite mit der Silhouette und die dritte mit dem Skelett. Das Spiegelbild wurde als Einstiegsrepräsentation absichtlich außen vor gelassen, da Vorab-Tests gezeigt hatten, dass die Reaktion auf Bewegung in diesem Fall sofort ersichtlich war.

Die Beobachtungen zeigten, dass die Teilnehmer in allen drei Fällen innerhalb weniger Sekunden die Reaktion der Repräsentationen auf Bewegung erkannten. Im Fall des Quadrats dauerte dies allerdings etwas länger. Die Teilnehmer waren ratlos, bis einer von ihnen auf die Idee kam, einen Schritt nach vorn zu machen.

Bezüglich der Reaktionen auf die Einstiegsrepräsentationen kann festgehalten werden, dass diese sehr unterschiedlich ausfielen. Sie reichten von Ratlosigkeit (Quadrat) über Interesse (Silhouette) bis hin zu Erstaunen (Skelett). Das Spiegelbild wurde im Laufe der Studie zunächst ebenfalls mit Interesse wahrgenommen, dieses schwand aber recht schnell. Beim Skelett verhielt es sich umgekehrt, hier wurde regelrecht der Spieltrieb geweckt. Immer wieder wurde durch verschiedene Bewegungen damit herumexperimentiert.

In Bezug auf die Anwendung der Strategien konnten drei wesentliche Erkenntnisse gewonnen werden.

Erstens hat sich gezeigt, dass sich die Studienteilnehmer im Wesentlichen auf vier Strategien zur Identifikation ihrer Repräsentationen beschränkten. Diese sind folgende:

- *still stehen*, Identifikation durch betrachten/vergleichen von Aussehen, Statur und Haltung
- *winken/Hände bewegen*
- *herumlaufen*
- *herumlaufen und gleichzeitiges Winken/Hände bewegen*

Zweitens war die Entscheidung, welche dieser Strategien verwendet wird/werden, stark von der aktuell dargestellten Repräsentation abhängig. Beispielsweise wurde von den Teilnehmern relativ schnell bemerkt, dass das Quadrat nicht auf Arm- beziehungsweise Handbewegungen reagiert und stattdessen der ganze Körper in eine beliebige Richtung bewegt werden muss.

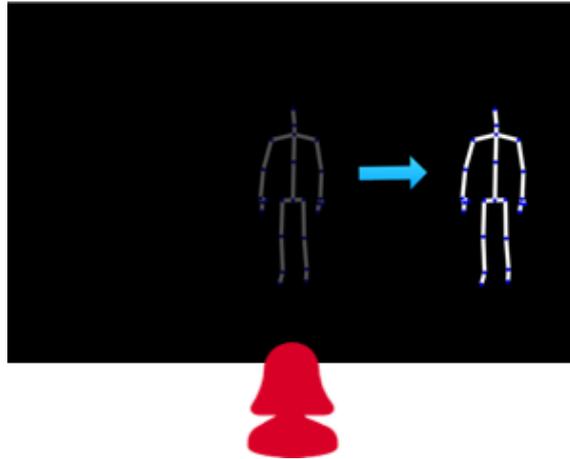


Abbildung 4.2: Funktionsweise des Offsets: Steht eine Person (rot im Bild) mittig vor dem Display wird deren Repräsentation auch mittig auf dem Display dargestellt (Skelett Mitte). Ist der Offset aktiv, wird die Repräsentation um einen Zufallswert nach links oder rechts verschoben (Skelett rechts).

Drittens gaben einige Teilnehmer an, das Verhalten der anderen Teilnehmer beobachtet und anschließend schnelle, zu diesen konträre Bewegungen gemacht zu haben. Bewegten sich beispielsweise mehrere Teilnehmer langsam nach vorne, ging ein Teilnehmer sehr schnell nach vorne oder nach hinten. Dadurch konnte dieser recht schnell herausfinden, welche Darstellung zu ihm gehörte.

Nach Aussage der meisten Teilnehmer erfolgte die Anwendungen die Strategien spontan/reflexartig. Lediglich ein Teilnehmer gab an bei der ersten Repräsentation (Quadrat) zunächst überlegt zu haben, welche Strategie man anwenden könnte.

4.5.3 Positions-Strategie

Die Vorstudie hat gezeigt, dass die in Kapitel 4.1 beschriebene Positions-Strategie tatsächlich angewandt wird. Nach etwa einem Drittel der Durchläufe begannen die Studienteilnehmer ihre Repräsentationen auf dem Display allein durch Abgleich mit ihrer physischen Position zu finden. Nach Aussagen einiger Teilnehmer während des Interviews wurden ab diesem Zeitpunkt kaum noch andere Strategien aktiv angewandt, höchstens zur Gegenprobe.

Die Anwendung der Positions-Strategie hatte zur Folge, dass die Teilnehmer ihre Repräsentationen sehr schnell erkennen konnten. Um die Anwendung der Positions-Strategie während der Hauptstudie zu vermeiden und gewährleisten zu können, dass die Teilnehmer nur jene Strategien anwandten, die ihnen mitgeteilt wurden, mussten die Repräsentationen der Teilnehmer, wie in Abbildung 4.2 ersichtlich, für bestimmte Bedingungen um einen Zufallswert verschoben werden (Offset).



Abbildung 5.1: Screenshot einer der Aufnahmen. Die Strategie lautete hier: *herumlaufen* (eigene Grafik).

5 Hauptstudie

Basierend auf den Erkenntnissen der Vorstudie wurde die Hauptstudie durchgeführt. Damit sollten die festgelegten Strategien näher untersucht und miteinander verglichen werden. Welche Ziele mit der Hauptstudie genau verfolgt wurden, wird im folgenden Abschnitt beschrieben.

5.1 Ziele

Das Ziel der Hauptstudie bestand vor allem darin, herauszufinden, wie die festgelegten Nutzerrepräsentationen hinsichtlich Erkennungsgeschwindigkeit und Fehlerquote im Vergleich zueinander und unter Anwendung der verschiedenen Strategien abschneiden.

Dazu wurden die Repräsentationen mit und ohne Offset sowie mit den durch die Vorstudie ermittelten Strategien getestet. Außerdem wurde versucht, neben der Repräsentation der Teilnehmer, möglichst viele Repräsentationen gleichzeitig darzustellen, um festzustellen, wie sich das auf die Leistung der Teilnehmer auswirkt. Da Kinect der zweiten Generation maximal sechs Skelette/Körper gleichzeitig erfassen kann, wurde diese Anzahl gewählt (siehe auch Kapitel 5.2).

Einige Teilnehmer der Vorstudie gaben an, Repräsentationen wie Silhouette oder Skelett mit der Zeit besser erkannt zu haben. Deshalb sollte zudem untersucht werden, wie sich die Erkennungsgeschwindigkeiten der Repräsentationen mit der Zeit verändern und ob sich mit der Zeit eine Art Lerneffekt einstellte.

5.2 Aufnahmen

Bevor näher auf den Ablauf und die Ergebnisse der Hauptstudie eingegangen wird, muss gesagt werden, dass diese, im Gegensatz zur Vorstudie, aus zwei Gründen nicht mit Teilnehmergruppen sondern einzelnen Teilnehmern durchgeführt wurde. Zum einen wäre, hätte man eine Gleichverteilung der Bedingungen bei physischen Teilnehmern gewährleisten wollen, eine unrealistisch hohe Zahl an Teilnehmern erforderlich gewesen. Zum anderen hat sich in der Vorstudie gezeigt (siehe Abschnitt Strategien in Kapitel 4.5.2), dass einige Teilnehmer die übrigen Teilnehmer aus der Gruppe beobachteten und zu deren Bewegungen schnelle, konträre Bewegungen machten.

Stattdessen wurde ein alternativer Ansatz gewählt, der auf die Projektion aufgenommener Repräsentationen setzte. Um die Hauptstudie durchführen zu können, war es daher zunächst notwendig, entsprechende Aufnahmen zu erstellen. Wie das Erstellen der Aufnahmen ablief, wie diese strukturiert waren und wie genau sie in der Hauptstudie genutzt wurden, wird in den nächsten Abschnitten beschrieben.

Das Erstellen der Aufnahmen gestaltete sich wie folgt: 12 Personen wurden in 2 Gruppen mit je 6 Personen eingeteilt. Im Wechsel wurden die Gruppen gebeten, sich in einen Raum zu begeben, in dem, ähnlich wie in der Vorstudie, Kinect und ein Kurzdistanzprojektor aufgebaut waren. Basierend auf den 4 Repräsentationsformen und den aus der Vorstudie ermittelten 4 Strategien wurden 16 Bedingungen gebildet. Jede Bedingung enthielt die Kombination aus einer Repräsentationsform und einer Strategie. Abhängig von der jeweils aktuellen Bedingung wurden insgesamt 6 Repräsentationen derselben Repräsentationsform auf einer Projektionsfläche dargestellt, beispielsweise 6 Skelette oder 6 Silhouetten. Die Teilnehmer der sich im Raum befindlichen Gruppe wurden gebeten, die in der Bedingung festgelegte Strategie auszuführen, zum Beispiel zu *winken/die Arme zu bewegen* oder *herumzulaufen*. Jede Bedingung wurde in zwei 2-minütigen Videos festgehalten (siehe Abbildung 5.1). Diese hatten eine Auflösung von 1280×720 Pixel. Bei der Aufnahme wurde nur der Inhalt der Projektionsfläche aufgezeichnet, nicht die Teilnehmer. Waren die ersten 2 Minuten um, fand die zweite Aufnahme derselben Bedingung statt, diesmal allerdings mit den Personen, deren Repräsentationen im ersten Video nicht zu sehen waren. Die Gruppenzusammensetzungen wurden außerdem nach jeder Bedingung verändert. Es wurde sichergestellt, dass jede Person mindestens ein Mal in jeder Bedingung vorkam. Dadurch sollte es den Teilnehmern der Hauptstudie, vor allem im Fall des Spiegelbilds, erschwert werden, sich einzelne Personen auf den Aufnahmen zu merken, um ihr Spiegelbild durch Ausschluss eben jener, bereits bekannter Personen schneller identifizieren zu können. Insgesamt wurden 32 Aufnahmen ($4 \text{ Repräsentationen} \times 4 \text{ Strategien} \times 2 \text{ Videos}$) erstellt, welche während der Hauptstudie parallel zur Live-Repräsentation des Teilnehmers dargestellt wurden.

Wie in Kapitel 3.3.1 beschrieben wurden die Aufnahmen in sehr hoher Qualität erstellt. Dadurch konnte gewährleistet werden, dass von den Teilnehmern der Hauptstudie keine Unterschiede zwischen Aufnahme und Live-Repräsentation in Punkto Bildqualität wahrgenommen werden konnten. Andernfalls wären die Messergebnisse möglicherweise beeinträchtigt worden.

5.3 Aufbau

Die Hauptstudie bestand, ebenso wie die Vorstudie, aus zwei großen Teilen. Den ersten Teil stellten die bereits in Kapitel 5.2 beschriebenen Aufnahmen dar. Sie dienten gewissermaßen als Vorarbeit und waren notwendig um den zweiten Teil - die eigentliche Studie - durchführen zu können. Im Folgenden wird genauer auf deren Aufbau eingegangen.

Wie im Fall der Vorstudie fand auch die Hauptstudie in einem Raum der Arbeitsgruppen Medieninformatik und Mensch-Maschine-Interaktion statt. Dieser war auch hier mit Kinect und Kurzdistanzprojektor ausgestattet. Letzterer war notwendig, um im Projektionsbild keine Schatten zu werfen. Wie in der Vorstudie betrug die Größe der Projektion etwa 2 m in der Diagonale (Seitenverhältnis: 16:9). Sie hatte eine Auflösung von 1280×720 Pixel. Um zu verhindern, dass sich Teilnehmer aus dem Sichtfeld von Kinect hinausbewegten, wurden auf dem Boden mittels Klebeband entsprechende Markierungen angebracht. Der Abstand vom Mittelpunkt des markierten Bereichs bis zur Projektionsfläche betrug auch hier etwa 3,60 m.

Es wurde außerdem sichergestellt, dass keine Personen an der Studie teilnahmen, die bereits an der Vorstudie teilgenommen hatten. Diese Einschränkung war notwendig, da die Ergebnisse sonst durch das Vorwissen der Teilnehmer hätten verzerrt werden können.

Wie auch bei der Vorstudie wurde die Teilnahme an der Hauptstudie je nach Wunsch mit einem Amazon-Gutschein oder MMI-Punkt vergütet. Während der Studie wurden Video- und Fotoaufnahmen erstellt. Alle Teilnehmer erklärten sich mit der Erstellung der Aufnahmen sowie der Nutzung dieser für wissenschaftliche Zwecke einver-

standen und unterzeichneten daher eine entsprechende Erklärung. Insgesamt nahm die Studie pro Teilnehmer etwa 45 Minuten in Anspruch.

5.4 Ablauf

Zu Beginn der Studie wurde der jeweilige Teilnehmer vom Studienleiter begrüßt und über den Ablauf der Studie informiert. Zunächst wurde er über die Notwendigkeit der Videoaufzeichnungen aufgeklärt.

Es folgte die Erläuterung der Aufgabe. Dabei wurde dem Teilnehmer gesagt, dass sogleich 7 gleichartige Repräsentationen auf die Wand projiziert würden. Seine Aufgabe bestünde darin, so schnell und genau wie möglich herauszufinden, welche dieser Repräsentationen zu ihm gehöre. Um dies herauszufinden, könne er jeweils eine von vier verschiedenen Strategien anwenden. Welche das sei, würde vor jedem Durchgang vom Studienleiter festgelegt. Die erste Strategie bestünde darin *still zu stehen*, die zweite *zu winken/die Arme zu bewegen*, die dritte *im Raum herumzulaufen* und die vierte *die Arme zu bewegen und gleichzeitig im Raum herumzulaufen*. Der Teilnehmer wurde außerdem darauf hingewiesen, sich immer so zu bewegen, dass er mit dem Gesicht zu Kinect ausgerichtet sei, sich also nicht um die eigene Achse zu drehen. Des Weiteren wurde er gebeten, sich immer in dem mit Klebestreifen markierten Bereich aufzuhalten.

Anschließend wurde dem Teilnehmer ein Xbox 360 Wireless Controller gezeigt. Er wurde gebeten, sich einen Button auf dem Controller auszusuchen und diesen zu drücken, sobald er seine Repräsentation erkannt hatte. Dadurch wurde ermittelt, wie lange der Teilnehmer gebraucht hatte, um die jeweilige Repräsentation zu erkennen. Anschließend sollte er dem Studienleiter gegenüber seinen Verdacht äußern, diesem also mitteilen, welche der Repräsentationen er für seine hielt. Die Aussage des Teilnehmers wurde daraufhin überprüft und vermerkt. Dadurch wurde die Fehlerquote der jeweiligen Repräsentation gemessen.

Um die Erkennung der Repräsentation nicht künstlich zu vereinfachen und dadurch die Ergebnisse zu verzerren, wurde der Teilnehmer außerdem gebeten, den Controller während der ganzen Studie nicht in beiden Händen, sondern nur in einer Hand zu halten.

Es folgte die eigentliche Studie. Diese bestand aus insgesamt 96 Durchläufen, wobei in jedem Durchlauf eine Bedingung getestet wurde. Jede dieser Bedingungen setzte sich aus einer der 4 Repräsentation, der jeweiligen Offset-Einstellung (Offset aktiv oder inaktiv) sowie einer der 4 Strategien zusammen. Um Aussagen über einen potentiellen Lerneffekt treffen zu können, wurde jede der 32 Bedingungen 3 Mal getestet (daher 96 Durchläufe).

Zu Beginn jedes Durchgangs wurde jeder Teilnehmer gebeten, seine Augen zu schließen. Daraufhin wurde der Software-Prototyp gestartet. Dieser legte für den Teilnehmer zunächst die Startbedingung fest. Welche Startbedingung dabei ausgewählt wurde, hing, um eine Gleichverteilung der Bedingungen unter den Teilnehmern zu ermöglichen, von der jeweiligen Teilnehmer-ID ab. Anschließend startete die Erfassung mittels Kinect. Die in der aktuellen Bedingung festgelegte Repräsentationsform wurde genutzt, um die Repräsentation des Nutzers darzustellen. Zudem wurde aus einer der beiden Aufnahmen der jeweiligen Bedingung zufällig ein 30-sekündiger Clip ausgewählt und gleichzeitig mit der Repräsentation des Teilnehmers angezeigt (siehe Abbildung 5.2).

Ein wichtiger Aspekt hierbei waren auch die Lichtverhältnisse. Da die Aufnahmen zu einer anderen Tageszeit und damit unter anderen Lichtverhältnissen stattfanden als die Studie, mussten vor jedem Studientermin Tests durchgeführt und die Lichtverhältnisse durch Absenken/Hochziehen der Jalousien gegebenenfalls angepasst werden. Wäre dies nicht geschehen, hätten sich beim Spiegelbild hinsichtlich Live-Darstellung und Aufnahme Unterschiede in den Lichtverhältnissen bemerkbar gemacht.

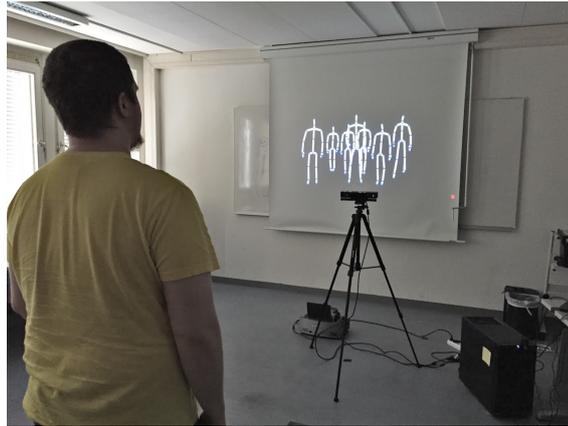


Abbildung 5.2: Ein Teilnehmer der Studie versucht, seine Repräsentation unter Zuhilfenahme der Strategie *still stehen* zu erkennen (eigene Grafik).

Waren Aufnahme und Repräsentation des Teilnehmers sichtbar, wurde der Teilnehmer gebeten, die in der geltenden Bedingung festgelegte Strategie auszuführen. Währenddessen zählte der Studienleiter von 3 auf 1 herunter. War er bei 1 angekommen, sollte der Teilnehmer seine Augen öffnen, die Strategie weiterhin anwenden und so versuchen, seine Repräsentation zu identifizieren. Das Schließen und Öffnen der Augen war notwendig, da der Teilnehmer zunächst von Kinect erfasst werden musste und sich die Darstellung seiner Repräsentation daher, im Vergleich zur Aufnahme, immer um wenige Sekunden verzögerte. Hätte er die Augen nicht geschlossen, hätte er bemerkt, dass plötzlich eine weitere Repräsentation angezeigt wurde und daher womöglich daraus gefolgert, dass dies seine ist.

Fragebögen

Waren alle Bedingungen 3 Mal getestet worden, wurde der Teilnehmer gebeten, zwei Fragebögen auszufüllen. Laut erstem Fragebogen waren zwecks Zustellung der Vergütung (Amazon-Gutschein oder MMI-Punkt) Angaben zu E-Mail Adresse und Matrikelnummer zu machen. Auf den zweiten Fragebogen (siehe Abbildungen A7, A8 und A9 im Anhang) wird im Folgenden näher eingegangen.

Neben der Erhebung quantitativer Daten war man auch daran interessiert zu erfahren, wie die Teilnehmer die jeweiligen Repräsentationsarten hinsichtlich Erkennungszeit einschätzen, ob sie bestimmte Repräsentationsformen präferieren und wenn ja weshalb. Um diese Informationen in Erfahrung zu bringen, wurde ein zweiter Fragebogen gestaltet. Dieser enthielt insgesamt 9 Fragen. Da sich die Fragen 1 bis 4 und 5 bis 8 jeweils nur in einem Punkt unterschieden, werden an dieser Stelle nur die jeweils erste Frage dieser Fragenblöcke und die letzte Frage genannt:

- Wie lange denken Sie, hat es gedauert, bis Sie Ihre Darstellung des Quadrats erkannt haben?
Antwortmöglichkeit: Likert-Skala von 1: Es ging sehr schnell bis 5: Es dauerte sehr lange.

- Stellen Sie sich vor, Sie wollen mit einem Display im öffentlichen Raum mittels Bewegung interagieren. Dabei sollen Nutzer ihre Darstellung möglichst schnell und präzise erkennen können. Welche der vier Darstellungen eignet sich Ihrer Meinung nach am besten dafür?
Antwortmöglichkeit: Drop-Down-Liste mit den Punkten 1. Quadrat, 2. Silhouette, 3. Skelett und 4. Spiegelbild
- Weshalb haben Sie sich in den letzten Fragen für genau diese Reihenfolge entschieden?
Antwortmöglichkeit: Freitext

Der zweite Fragebogen diente außerdem der Erhebung demographischer Angaben. Neben Alter, Geschlecht und Beruf/Studiengang wurde auch die Größe abgefragt, da vermutet wurde, dass besonders große Teilnehmer es womöglich einfacher hatten, ihre Repräsentation zu identifizieren.

5.5 Auswertung

An die in den vorangegangenen Unterkapiteln beschriebene Hauptstudie schloss sich die Auswertung der Protokoll-Daten an. Deren Vorgehensweise sowie die damit verbundenen Ergebnisse werden im nun folgenden Kapitel näher beschrieben.

5.5.1 Vorgehensweise

Die Auswertung der Daten erfolgte in vier Schritten. Der erste Schritt bestand aus einer grundlegenden Analyse der Daten. Diese sollte sicherstellen, dass die Daten vollständig sind und keine Fehler enthalten. Dabei stellte sich jedoch heraus, dass 116 der insgesamt 1824 Zeitwerte (das entspricht 6,36 %) doppelt vorkamen, also dass jeweils zwei Mal direkt hintereinander derselbe Wert abgespeichert wurde.

Die Erklärung hierfür liegt in der Art und Weise, wie die Zeit gemessen wurde, die die Teilnehmer bis zur Erkennung ihrer Repräsentationen benötigten. Wie in Kapitel 5.4 bereits beschrieben wurden die Studienteilnehmer gebeten, ihre Augen zu schließen, bevor die Aufnahme und Live-Repräsentation dargestellt wurden. Anschließend zählte der Studienleiter von 3 auf 1 herunter, woraufhin die Teilnehmer ihre Augen öffneten und versuchten ihre Repräsentation zu identifizieren. Mit dem Herunterzählen auf 1 betätigte der Studienleiter auch eine Taste auf der Tastatur, was dazu führte, dass durch den Software-Prototyp eine Stoppuhr gestartet wurde. Durch drücken des Buttons auf dem Controller seitens des Teilnehmers wurde die Stoppuhr wieder angehalten.

Es gab Fälle in denen der Controller betätigt wurde, bevor die Stoppuhr überhaupt gestartet war. Dies führte dazu, dass der zuletzt gemessene Wert erneut protokolliert wurde. Gleichzeitig bedeutete es jedoch auch, dass der jeweils erste Wert des Paares verwendet werden konnte. Die 116 als Duplikat identifizierte Werte wurden daraufhin entfernt.

Es folgte die Erkennung von Ausreißern¹⁴. Hierbei wurden besonders niedrige beziehungsweise hohe Messwerte entfernt, da man davon ausgehen musste, dass diese durch äußere Begleitumstände entstanden sind. Beispielsweise kam es vor, dass Teilnehmer bei der Strategie *stehen bleiben*, sehr lange versuchten, ihre Repräsentation zu identifizieren, was vor allem im Fall des Quadrats so gut wie unmöglich war da sie dazu keinerlei Anhaltspunkte hatten.

Das Erkennen von Ausreißern lief in zwei Phasen ab. In der ersten Phase wurden ausschließlich jene Werte entfernt, die kleiner oder gleich 500 ms waren. Der Grund hierfür war folgender: Laut Kieras [36] kann davon ausgegangen werden, dass ein Mensch für einen Mausklick¹⁵ mindestens 700 ms benötigt. Dies schließt sowohl die Zeit ein, die benötigt wird, um die entsprechende Entscheidung zu treffen, als auch die Zeit, die benötigt wird, um den Mausklick tatsächlich auszuführen. Da die Entfernung aller Werte kleiner 700 ms jedoch zur Folge gehabt hätte, dass

¹⁴engl. outlier detection

¹⁵Entsprechende Angaben für Spiele-Controller konnten nicht gefunden werden. Daher wurden jene für Mausklicks als Referenz herangezogen.

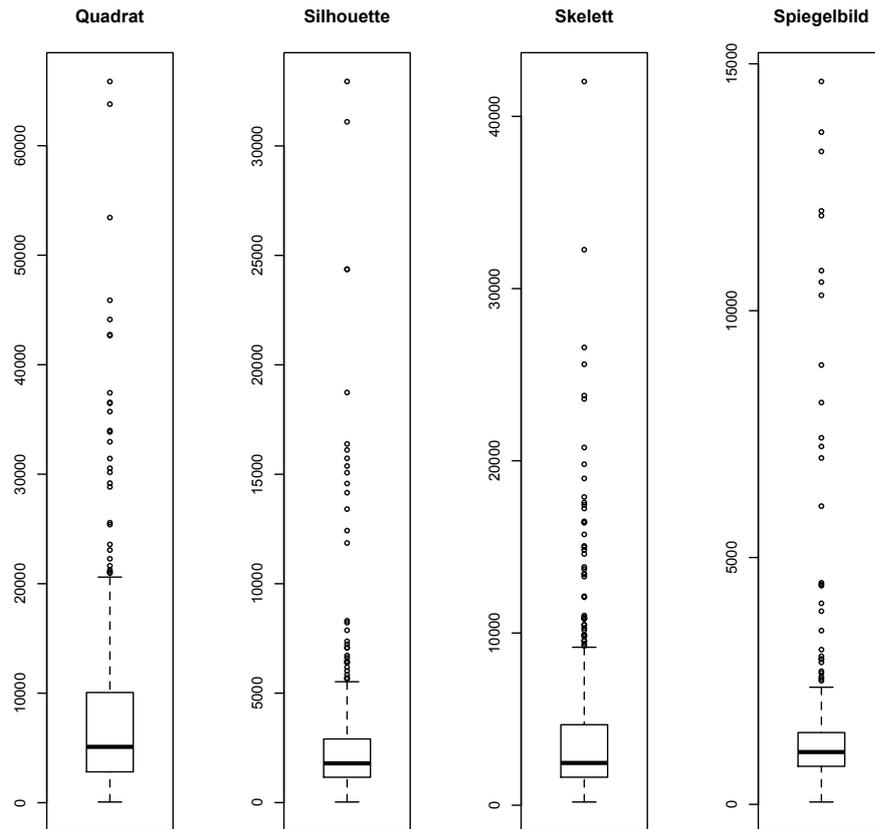


Abbildung 5.3: Die Boxplots der Repräsentationsformen Quadrat, Silhouette, Skelett und Spiegelbild. Die y-Achse stellt die Erkennungszeit in Millisekunden dar, die Kreise stehen für Ausreißer. Die Whisker/Antennen hatten in allen erstellten Boxplots eine maximale Länge vom 1,5-fachen des Interquartilabstands.

unter anderem 89 Werte der Repräsentation Spiegelbild entfernt worden wären (dies entspricht fast 20 %), entschied man sich dafür, 500 ms als Grenzwert festzulegen. Dadurch wurden im Fall des Spiegelbilds nur 35 Werte entfernt. Insgesamt hatte der erste Schritt der Outlier-Detection die Entfernung von 53 Werten zur Folge.

In der zweiten Phase der Ausreißerererkennung wurden Boxplots erstellt, um sehr große Werte identifizieren und entfernen zu können. Die Datenbasis der Boxplots war abhängig von den Berechnungen, die im Rahmen der deskriptiven Statistik durchgeführt werden sollten. Es konnte kein Boxplot über die gesamte Datenbasis erstellt werden, da sonst die meisten Messwerte der Repräsentationsform Quadrat durch die im Vergleich zu den übrigen Repräsentationsformen deutlich längere Erkennungszeit entfernt worden wären. Stattdessen wurden für alle erforderlichen Berechnungen (beispielsweise durchschnittliche Erkennungszeiten für alle vier Repräsentationsformen oder einzelne Bedingungen) Boxplots auf Grundlage der jeweiligen Datenbasis erstellt. Die Abbildung 5.3 zeigt die erstellten Boxplots für die vier Repräsentationen Quadrat, Silhouette, Skelett und Spiegelbild. Sie wurden herangezogen, um die durchschnittlichen Erkennungszeiten der jeweiligen Repräsentationen berechnen zu können. Wie in der Abbildung zu sehen ist, gab es große Unterschiede in Bezug auf Verteilung und Maximalwerte der Ausreißer. Insgesamt wurden für die vier Repräsentationsformen folgende Mengen an Werten als Ausreißer identifiziert: Quadrat: 26 von 456 Werten (5,7%), Silhouette: 33 von 456 Werten (7,24%), Skelett: 43 von 456 Werten (9,43%), Spiegelbild: 30 von 456 Werten (6,58%).

Es sei außerdem gesagt, dass im Allgemeinen keine Entscheidungswerte entfernt wurden, auch dann nicht, wenn es sich bei den dazugehörigen Zeitwerten um Ausreißer handelte. Die Gründe hierfür sind folgende: War der zu einem Entscheidungswert gehörende Zeitwert kleiner als 500 ms, ging man davon aus, dass der entsprechende Teilnehmer sich sehr sicher war, die jeweilige Repräsentation unter der aktuellen Bedingung erkennen zu können und daher sehr schnell drückte.

Handelte es sich beim Zeitwert um einen durch die Boxplots identifizierten Ausreißer, also einen sehr großen Wert, nahm man an, dass sich der jeweilige Teilnehmer viel Zeit ließ, seine jeweilige Repräsentation zu erkennen, obwohl dies unter der aktuell geltenden Bedingung sehr schwierig war. Dieses Verhalten wurde während der Studie mehrfach beobachtet (vor allem wenn es sich bei der Repräsentationsform um das Quadrat und bei der Strategie um *still stehen* handelte).

An das Erkennen und Entfernen der Ausreißer schloss sich der dritte Schritt an: das Erstellen einer deskriptiven Statistik. Im Rahmen dieser wurden Mittelwerte über die einzelnen Repräsentationsformen, Bedingungen und Durchläufe berechnet. Zudem fanden Berechnungen der Fehlerquote statt, sowohl auf der Ebene der Repräsentationsformen wie auch auf der Ebene der 32 Bedingungen.

Die Auswertung endete mit dem vierten Schritt, der Durchführung zweier statistischer Tests sowie Anwendung eines statistischen Modells. Dadurch sollte die Signifikanz der gemessenen Daten festgestellt werden, das heißt, es sollte ermittelt werden, ob die in der Hauptstudie gemessenen Werte und in der deskriptiven Statistik ermittelten Zusammenhänge zufällig entstanden sind oder tatsächlich durch die verschiedenen Bedingungen der Hauptstudie verursacht wurden. Konkret wurde zunächst ein lineares gemischtes Modell auf Basis der Daten erzeugt. Dieses wurde gewählt, da es sich, im Gegensatz zum ursprünglich gewählten ANOVA Test, durch seine Robustheit eher für Fälle eignet, in denen einzelne Messwerte fehlen, was hier, durch Entfernen der Duplikate und Ausreißer der Fall war. Das Modell wurde zunächst auf solche Zeitwerte angewandt, für die galt, dass die entsprechende Repräsentation korrekt erkannt wurde. Für jede Bedingung wurden jeweils alle drei Durchgänge durch Bildung des Mittelwerts zu einem Durchgang zusammengefasst. Die Datenmenge, die dem linearen gemischten Modell unterworfen wurde, enthielt daher für jeden der 19 Teilnehmer Werte für 32 Bedingungen. Da hierbei signifikante Effekte festgestellt wurden, erfolgten mit dem Modell ein zweiter und dritter Durchlauf. Man war insbesondere daran interessiert herauszufinden, ob ein aktivierter beziehungsweise deaktivierter Offset signifikante Auswirkungen auf die Erkennungszeiten hatte. Daher erfolgten die zweite und dritte Anwendung des Modells nur auf jeweils 16 Bedingungen.

Die Untersuchung der Entscheidungswerte in Bezug auf Signifikanz erfolgte durch den Friedman-Test sowie einen anschließenden Wilcoxon-Test.

Da für die Durchführung des Friedman-Tests Zahlen notwendig waren, wurden die ermittelten und als true beziehungsweise false vorliegenden Entscheidungswerte aller drei Durchläufe zusammengefasst, auf die Zahlen 0 bis 3 abgebildet und wie folgt in eine Tabelle eingetragen. Traten bei einem Studienteilnehmer über alle 3 Durchläufe einer Bedingung hinweg keine Fehler auf, wurde an der entsprechenden Stelle in der Tabelle eine 0, bei 1 Fehler eine 1, bei 2 Fehlern eine 2 und bei 3 Fehlern eine 3 eingefügt.

Die Durchführung des Friedman-Tests erfolgte ein Mal mit der gesamten eben beschriebenen Tabelle (32 Bedingungen), ein Mal mit jenen Bedingungen bei denen der Offset aktiv (16 Bedingungen) und ein Mal mit jenen Bedingungen bei denen der Offset inaktiv (16 Bedingungen) war. In allen drei Fällen lag die Asymptotische Signifikanz deutlich unter 0,05. Somit konnte davon ausgegangen werden, dass bei einzelnen Bedingungen signifikante Unterschiede zwischen deren Mittlerer Ränge bestanden.

Da der Friedman-Test jedoch keine Auskunft darüber gibt, bei welchen Bedingungen dies der

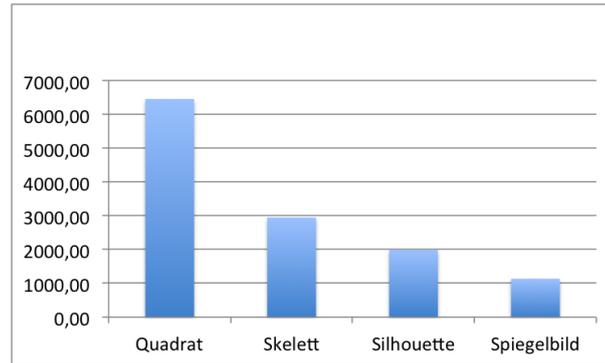


Abbildung 5.4: Die durchschnittlichen Erkennungszeiten der vier Repräsentationsformen in Millisekunden.

Fall ist, wurde ein weiterer Test durchgeführt, der sogenannte Wilcoxon-Test. Hierbei wurden die Bedingungen paarweise auf einen signifikanten Unterschied getestet. Im ersten Schritt erfolgte dies mit allen 32 Bedingungen. Daher wurden zunächst 496 paarweise Tests durchgeführt. Die Bonferroni-Anpassung führte in diesem Fall zu einem neuen $p = 0,05/496 = 0,0001008064516$. Im zweiten und dritten Schritt wurden mit den ersten 16 beziehungsweise letzten 16 der 32 Bedingungen jeweils 120 paarweise Tests durchgeführt. In beiden Fällen galt durch die Bonferroni-Anpassung: $p = 0,05/120 = 0,0004166666667$.

5.5.2 Ergebnisse

Es folgt nun eine Darstellung der Ergebnisse der Hauptstudie. Hierbei wird zunächst auf die Teilnehmer und anschließend auf weitere Aspekte wie Repräsentationen, Lerneffekt oder auch Fragebögen eingegangen. Das Kapitel schließt mit einem Abschnitt, der sich mit einigen Einschränkungen der Hauptstudie auseinandersetzt.

Teilnehmer

An der Hauptstudie nahmen insgesamt 19 Personen teil. Diese waren zwischen 18 und 36 Jahren alt. Das Durchschnittsalter lag bei 26,5 Jahren. 11 Teilnehmer waren männlich, 8 weiblich. 10 Teilnehmer waren Studenten der Studiengänge Medieninformatik oder Mensch-Maschine-Interaktion und 3 Softwareentwickler. Die Betätigungsfelder der übrigen 6 Teilnehmer verteilten sich auf Physik, Chemie, Pharmazie, Romanistik, Statistik und Nordamerikastudien.

Repräsentationen

In Bezug auf die Repräsentationen lassen sich folgende Ergebnisse festhalten: Wie in Abbildung 5.4 zu sehen ist, stellt das Spiegelbild jene Repräsentationsart dar, die, wie im Vorfeld erwartet, am schnellsten erkannt wurde. In Kontrast dazu steht das Quadrat, welches durchschnittlich am langsamsten erkannt wurde. Silhouette und Skelett liegen diesbezüglich an zweiter beziehungsweise dritter Position.

Vergleicht man die konkreten Erkennungszeiten, fällt auf, dass der Unterschied zwischen Spiegelbild (1137 ms) und Silhouette (1993 ms) beziehungsweise Silhouette und Skelett (2947 ms) jeweils etwa 1 Sekunde beträgt. Im Gegensatz dazu fällt der Unterschied zwischen Skelett und Quadrat (6456 ms) mit 3,51 Sekunden deutlich größer aus.

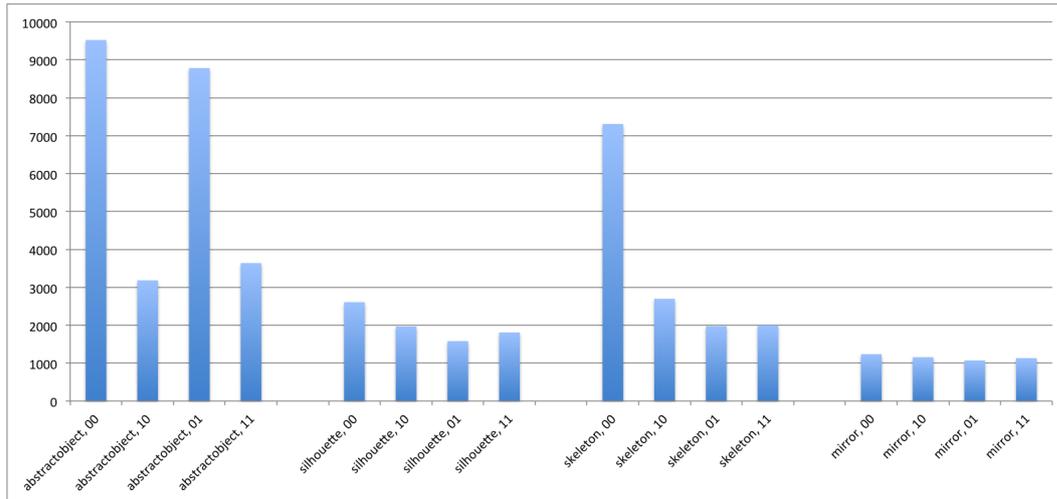


Abbildung 5.5: Durchschnittliche Erkennungszeiten in Abhängigkeit der Repräsentationsformen und Strategien. Die Zeiten sind in Millisekunden angegeben.

Hinsichtlich der Fehlerquote verhält es sich ähnlich. Auch hier wurde das beste Ergebnis durch das Spiegelbild erzielt; es gelang den Teilnehmern in allen Fällen ihre Repräsentation zu erkennen. Da jedoch davon ausgegangen werden kann, dass sich jeder Mensch - sofern er keinen Einschränkungen unterliegt - im Spiegel erkennen kann, ist dieses Ergebnis nicht verwunderlich. Im Falle der übrigen Repräsentationsformen verteilte sich die Fehlerquote wie folgt: Das Quadrat wurde in 136 von 409 Fällen (33,25 %) nicht korrekt erkannt. Damit stellt es die Repräsentationsform mit der höchsten Fehlerquote dar. Skelett und Silhouette schnitten mit 32 von 424 (7,5 %) beziehungsweise 16 von 423 (3,78 %) falsch erkannten Repräsentationen auch hier deutlich besser als das Quadrat ab.

Strategien

Betrachtet man die Erkennungszeiten der Repräsentationsformen in Verbindung mit den Strategien, fällt auf, dass diese (wie in Abbildung 5.5 zu sehen), je nach angewandter Strategie, stark variieren.

Allen Repräsentationsformen ist hierbei gemein, dass die Erkennung unter Anwendung der Strategie *still stehen* im Vergleich zu den anderen Strategien am längsten dauerte. Im Fall des Spiegelbilds fiel der Unterschied zu den anderen Strategien mit 36 - 220 ms jedoch deutlich geringer aus als bei den übrigen Repräsentationsformen - Quadrat: 525 - 6565 ms, Silhouette: 607 - 1.066 ms, Skelett 4094 - 5.995 ms).

Es zeigte sich außerdem, dass für die Repräsentationen Quadrat und Skelett neben *still stehen* jeweils eine weitere Strategie zu deutlich längeren Erkennungszeiten führte. Im Fall des Quadrats handelte es sich um die Strategie *Arme bewegen* beziehungsweise *winken/Hände bewegen*, im Fall des Skeletts um die Strategie *herumlaufen/Arme nicht bewegen*. Wurden diese Strategien angewandt, ergaben sich im Vergleich zu den anderen Strategien Zeitunterschiede von 4.587 - 5.950 ms beim Quadrat und 560 - 888 ms beim Skelett. Unter Anwendung welcher Strategien die vier Repräsentationen am besten abschnitten zeigt Abbildung 5.5. Die Abbildungen A5 und A6 (beide im Anhang) stellen denselben Sachverhalt dar, unterscheiden jedoch zwischen den Bedingungen mit inaktivem und jenen mit aktivem Offset.

Beim Blick auf die Verteilung der Fehlerquote zeigte sich auch hier wieder ein ähnliches Bild. Abgesehen vom Spiegelbild zeigten alle Repräsentationsformen deutlich erhöhte Fehlerraten, sobald die Strategie *still stehen* zum Einsatz kam. Im Falle des Quadrats stiegen die Fehlerraten

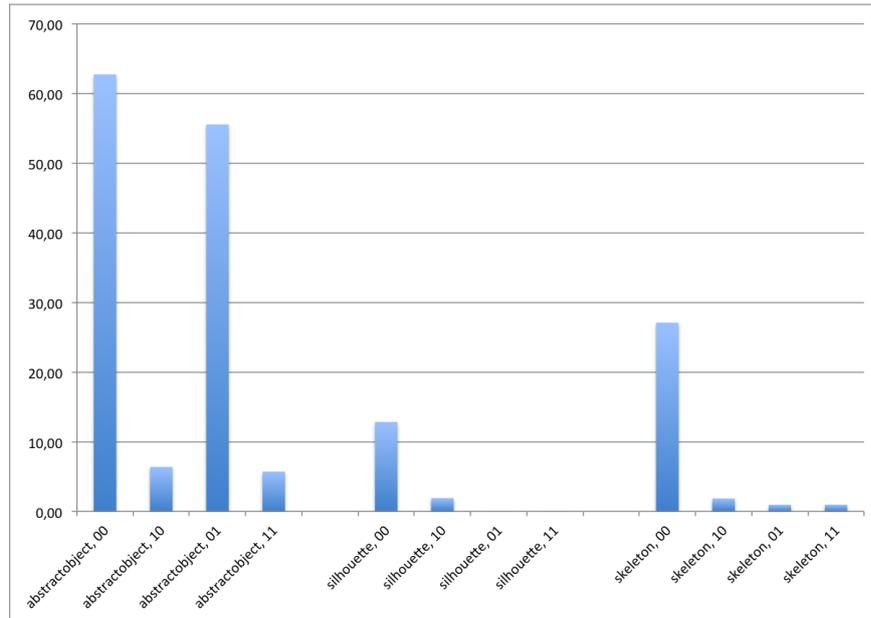


Abbildung 5.6: Die Fehlerquoten der Repräsentationsformen Silhouette, Skelett und Quadrat in Verbindung mit den Strategien *still stehen* (00), *herumlaufen* (10), *Arme bewegen/Winken* (01), *herumlaufen und Arme bewegen/Winken* (11). Dargestellt werden die Werte aller Bedingungen (Offset aktiv und Offset inaktiv).

zudem auch dann stark an, wenn *Arme bewegen* angewandt wurde. Dies lässt sich damit erklären, dass sowohl *still stehen* als auch *Arme bewegen* keinerlei Auswirkungen auf die Repräsentationsform Quadrat haben. Werden diese Strategien angewandt, bewegt sich das Quadrat nicht. Wie in Abbildung 5.6 ersichtlich lag der Anteil der falsch identifizierten Repräsentationen für die Strategie *still stehen* bei 12,84 % (Silhouette), 27,10% (Skelett) und 62,75 % (Quadrat).

Lerneffekt

Im Rahmen der Auswertung war man auch daran interessiert herauszufinden, ob eine Art Lerneffekt aufgetreten ist, also einzelne Repräsentationsformen unter bestimmten Bedingungen mit der Zeit schneller erkannt werden konnten. Insbesondere galt die Aufmerksamkeit hier der Repräsentationsform Silhouette, da einige Teilnehmer während der Studie aussagten, ihre Silhouette zu Beginn kaum, mit der Zeit aber immer schneller erkannt zu haben.

Um Aussagen diesbezüglich treffen zu können, wurden für jede der 32 Bedingungen die Durchschnittswerte aller 3 Durchläufe berechnet. Die daraus gewonnen Erkenntnisse werden im Folgenden näher erläutert.

Zunächst muss gesagt werden, dass ein Lerneffekt, wenn überhaupt, nur bei 10 Bedingungen aufgetreten ist. Wie in Tabelle A3 (Anhang) zu sehen zeigen diese 10 (3 × Quadrat, 1 × Silhouette, 2 × Skelett und 4 × Spiegelbild) der insgesamt 32 Bedingungen über alle drei Durchläufe hinweg Verkürzungen der durchschnittlichen Erkennungszeiten. Bei 8 der 10 Bedingungen war der Offset deaktiviert, bei 2 aktiviert. Betrachtet man die durchschnittlichen Erkennungszeiten dieser Bedingungen genauer, stellt man fest, dass sich diese vor allem im 2. Durchlauf verkürzten (siehe Spalte 'Verkürzung Durchgänge 1 und 2' der Tabelle A3 im Anhang) und vor allem bei den Bedingungen verkürzten, die die Strategie *still stehen* enthielten. Mögliche Gründe hierfür werden in Kapitel 6 besprochen.

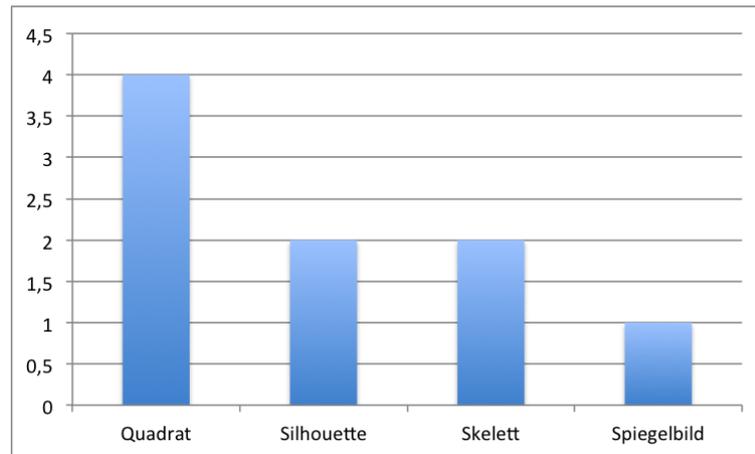


Abbildung 5.7: Einordnung der vier Strategien in Bezug auf die Erkennungszeit nach Meinung der Studienteilnehmer (Fragebogen, je kleiner die Zahl, desto schneller wurde die Repräsentation nach Meinung der Teilnehmer erkannt). Die Werte zeigen Mediane.

Offset

Neben den bereits erwähnten Aspekten sollte auch untersucht werden, inwiefern ein inaktiver Offset die Erkennungszeiten verringert beziehungsweise die Identifikation der eigenen Repräsentation erleichtert (Stichwort Positions-Strategie). Um dies untersuchen zu können, wurde die Reihenfolge der Bedingungen während der Hauptstudie derart gestaltet, dass in jedem Durchlauf alle 16 Bedingungen mit inaktivem und alle 16 mit aktivem Offset aufeinander folgten.

Es wurde erwartet, dass die Teilnehmer nach einigen getesteten Bedingungen mit inaktivem Offset erkennen würden, dass keine Verschiebung stattgefunden hat und sie ihre Repräsentationen, ähnlich wie in der Vorstudie, bis zur erneuten Aktivierung des Offsets allein durch Vergleich ihrer Position mit den Positionen der Repräsentationen identifizieren und die Erkennungszeiten dadurch sinken würden. Die Analyse der Daten zeigt, dass dies bei 7 der 19 Teilnehmer der Fall war. Hier weisen in mindestens zwei der drei Durchläufe jene Bedingungen mit inaktivem Offset durchschnittlich bessere Erkennungszeiten auf. Bei den übrigen 12 Teilnehmern weisen entweder keiner (2 Teilnehmer) oder einer der Durchläufe (10 Teilnehmer) kürzere Erkennungszeiten bei Bedingungen mit inaktivem Offset auf. Es stellt sich die Frage, weshalb die Deaktivierung des Offsets nicht bei mehr Teilnehmern zu Verkürzungen der Erkennungszeiten führte. Das Kapitel 6 befasst sich auch hiermit genauer und versucht Erklärungen zu finden.

Fragebogen

Wie bereits in Kapitel 5.5.1 beschrieben, wurden die Studienteilnehmer im Anschluss an den praktischen Teil der Studie gebeten, zwei Fragebögen auszufüllen. Der zweite dieser Fragebögen enthielt verschiedene Fragen zu den Repräsentationsformen und Strategien. Im nun folgenden Abschnitt werden die Antworten auf diese Fragen zusammengefasst und besprochen.

Die ersten vier Fragen des Fragebogens bezogen sich auf die vier Repräsentationen und die damit verbundenen Erkennungsgeschwindigkeiten. Die Teilnehmer sollten auf einer Likert-Skala mit den Werten 1 - 5 angeben, wie schnell sie ihrer Meinung nach die einzelnen Repräsentationsformen erkannten. 1 stand hierbei für: Es ging sehr schnell; 5 für: Es dauerte sehr lange. Dadurch sollte herausgefunden werden, ob und wenn ja inwiefern sich die Wahrnehmung beziehungsweise

das gefühlte Ergebnis vom gemessenen unterscheidet.

Wie in Abbildung 5.7 zu sehen ist, deckt sich das Ergebnis des Fragebogens nicht ganz mit dem der Messwerte. Alle Teilnehmer waren der Meinung, dass sich das Spiegelbild am schnellsten erkennen ließ (Median = 1, Standard-Abweichung = 0). Am zweitschnellsten ließen sich nach Meinung der Teilnehmer die Silhouette (Median = 2, Standardabweichung = 0,89) und das Skelett (Median = 2, Standardabweichung = 0,70) erkennen. Das Quadrat wurde von den Teilnehmern als die Repräsentationsform eingeschätzt, deren Erkennung am längsten dauerte (Median = 4, Standardabweichung = 0,85).

Die nächsten vier Fragen des Fragebogens beschäftigten sich mit der generellen Präferenz bezüglich der vier Repräsentationsformen. Diese sollten von den Teilnehmern absteigend nach deren Präferenz beziehungsweise Eignung für die Nutzung im öffentlichen Raum sortiert werden. In einer sich daran anschließenden Frage wurden die Teilnehmer gebeten zu begründen, weshalb sie genau diese Sortierung vorgenommen hatten.

Bis auf zwei Teilnehmer waren alle der Meinung, dass das Spiegelbild für die Nutzung im öffentlichen Raum am besten geeignet sei, da dieses neben Gesichtszügen auch Kleidung zeige und körperliche Merkmale wie beispielsweise die Form und Größe beinhalte. Ein Teilnehmer äußerte jedoch auch Datenschutzbedenken und den Wunsch nach Anonymisierung.

Nach Ansicht von 17 der 19 Teilnehmer stellt die Silhouette die für den beschriebenen Zweck am zweitbesten geeignete Repräsentationsform dar. Dies ist insofern überraschend, als die Erkennungsgeschwindigkeit offensichtlich nicht das alleinige Bewertungskriterium darstellte. Die Eignung der Silhouette als zweitbeste Repräsentationsform wurde vor allem mit der Erkennbarkeit einzelner Körpermerkmale begründet. Einzelne Teilnehmer argumentierten auch mit Kleidungsstücken, deren Konturen erkannt werden könnten. Die Identifikation könne zudem durch das Erkennen der Formen einzelner Gegenstände wie beispielsweise Handtaschen erleichtert werden. Dies hat sich auch in der Studie selbst bemerkbar gemacht, da in wenigen Fällen der Controller in der Silhouette sichtbar war. Nachteile wurden in der möglichen Verdeckung gesehen, die in der Studie durch die Nutzung von Aufnahmen in Kombination mit der Live-Darstellung zum Teil auftraten.

16 der 19 Teilnehmer waren der Meinung, dass das Skelett am drittbesten dazu geeignet sei, potentielle Nutzer an ein Display im öffentlichen Raum heranzuführen. Die meisten der 16 Teilnehmer argumentierten damit, dass beim Skelett, im Gegensatz zu Silhouette und Spiegelbild, eine Bewegung notwendig sei, um die entsprechende Repräsentation zu erkennen. Des Weiteren seien hier weniger Unterscheidungsmerkmale vorhanden als bei Silhouette und Spiegelbild.

Das Quadrat wurde, im Vergleich zu den übrigen Repräsentationsformen, als am wenigsten geeignet betrachtet. Die Teilnehmer sahen die Tatsache, dass dieses nur auf Bewegungen des ganzen Körpers in eine Richtung reagierte, als großen Nachteil an. Lediglich ein Teilnehmer zog das Quadrat dem Skelett vor. Letzteres wurde in diesem Fall als zu große Ablenkung empfunden.

Statistische Tests

Wie in Kapitel 5.5.1 bereits beschrieben, wurde ein lineares gemischtes Modell auf die Erkennungszeiten sowie ein Friedman- und Wilcoxon-Test auf die Entscheidungswerte angewandt. Dadurch sollte herausgefunden werden, inwiefern ein statistisch signifikanter Zusammenhang zwischen Fehlerrate und unabhängigen Variablen (Strategie, Repräsentation und Offset) beziehungsweise Erkennungszeit und unabhängigen Variablen besteht.

Das lineare gemischte Modell diente der Erkennung eines Zusammenhangs zwischen unabhängigen Variablen (Repräsentation, Strategie und Offset) und Erkennungszeit. Wie in Abbildung A1 (Anhang) zu sehen offenbarte die erste Anwendung des Modells signifikante Auswirkungen der unabhängigen Variablen Repräsentation und Strategie auf die Erkennungszeit. Zudem zeigt die Abbildung, dass die Auswirkungen der unabhängigen Variable Repräsentation

Tests auf feste Effekte, Typ II^a

Quelle	Zähler-Freiheitsgrade	Nenner-Freiheitsgrade	F-Wert	Signifikanz
Konstanter Term	1	44,208	164,216	,000
Representation	3	52,468	20,061	,000
Movement_Type	3	44,406	2,742	,054
Representation * Movement_Type	9	43,993	4,187	,001

a. Abhängige Variable: Decision_Time.

Abbildung 5.8: Ergebnisse des linearen gemischten Modells (feste Effekte, 16 Bedingungen pro Teilnehmer, Offset aktiv, Representation entspricht Repräsentation, Movement_Type entspricht Strategie).

auf die Erkennungszeit auch signifikant von der gewählten Strategie abhängt.

Um Aussagen darüber treffen zu können, wodurch diese Signifikanzen hervorgerufen werden, wurden jeweils 16 der Bedingungen in zwei weiteren Anwendungen des Modells genauer untersucht; zuerst die Bedingungen mit aktivem Offset, anschließend jene mit inaktivem Offset. Dabei wurden jeweils alle Repräsentationsformen sowie Strategien paarweise miteinander verglichen. Betrachtet man jene 16 Bedingungen, bei denen der Offset aktiv war, ergibt sich folgendes Bild: Wie in Abbildung 5.8 ersichtlich, besteht, wie auch bei der Analyse aller 32 Bedingungen, eine Signifikanz in Bezug auf die unabhängige Variable Repräsentation. Ebenso hängen auch hier die Auswirkungen der Repräsentation auf die Erkennungszeit statistisch signifikant von der gewählten Strategie ab. Im Gegensatz zur Analyse aller 32 Bedingungen kann in Bezug auf die Strategie allein jedoch keine Signifikanz festgestellt werden. Weshalb die Variable Repräsentation signifikante Auswirkungen auf die Erkennungszeiten hatte, zeigt Abbildung A2 (Anhang). Diese enthält alle paarweisen Vergleiche der vier Repräsentationsformen. Bis auf den Vergleich Quadrat (abstractobject) - Skelett (skeleton) zeigen hier alle Vergleiche Signifikanzen bei einem Signifikanzniveau von 5 % auf.

Bei Betrachten der übrigen 16 Bedingungen (Offset inaktiv) zeigten sich, wie in Abbildung 5.9 zu sehen, Signifikanzen sowohl bei Repräsentation als auch Strategie. Ebenso ist die Verbindung von Repräsentation und Strategie signifikant bei einem Signifikanzniveau von 5 %. Dies wird durch die paarweisen Vergleiche der Repräsentationsformen (siehe Abbildung A3 im Anhang) und Strategien (siehe Abbildung A4 im Anhang) bestätigt. In beiden Fällen finden sich Paare, die eine statistische Signifikanz aufweisen. Dies zeigt, dass auch hier die Nullhypothese widerlegt wurde, wonach davon ausgegangen wird, dass die unabhängigen Variablen keinerlei Einfluss auf die Erkennungszeiten hatten.

Der Wilcoxon-Test, dem alle 32 Bedingungen kombiniert unterworfen wurden, zeigte bei einem Signifikanzniveau von 5 % keinerlei signifikante Ergebnisse. Die Durchführung des Tests mit den 16 Bedingungen bei denen der Offset deaktiviert war, resultierte in signifikanten Ergebnissen bei insgesamt 10 von 120 Paaren. Wie in Abbildung 5.10 und Tabelle A1 (Anhang) zu sehen ist, zeigen zehn Bedingungen einen signifikanten Unterschied (Signifikanz-Wert < 5 %) in Bezug auf die Fehlerrate zur Bedingung abstractobject_00, also zum Quadrat mit der Strategie *stehen bleiben*.

Tests auf feste Effekte, Typ III^a

Quelle	Zähler-Freiheitsgrade	Nenner-Freiheitsgrade	F-Wert	Signifikanz
Konstanter Term	1	44,243	164,997	,000
Representation	3	59,044	23,099	,000
Movement_Type	3	29,353	7,564	,001
Representation * Movement_Type	9	48,437	2,982	,007

a. Abhängige Variable: Decision_Time.

Abbildung 5.9: Ergebnisse des linearen gemischten Modells (feste Effekte, 16 Bedingungen pro Teilnehmer, Offset inaktiv, Representation entspricht Repräsentation, Movement_Type entspricht Strategie).

Im Fall der übrigen 16 Bedingungen (Offset aktiviert) traten bei insgesamt 20 von 120 paarweisen Tests (siehe Abbildung 5.11 und Tabelle A2 im Anhang) signifikante Ergebnisse auf.

Zusammenfassend lässt sich Folgendes sagen: Es konnten signifikante Haupteffekte zwischen der unabhängigen Variablen Repräsentation und den abhängigen Variablen Erkennungszeit und Fehlerrate festgestellt werden. Außerdem konnten signifikante Haupteffekte zwischen der unabhängigen Variablen Strategie und den abhängigen Variablen Erkennungszeit und Fehlerrate nachgewiesen werden. Paare, die signifikante Unterschiede zeigten, sind in den Abbildungen 5.11 und 5.10 beziehungsweise den Tabellen A2 und A1 (beide im Anhang) zu sehen.

Dies bedeutet, dass die unabhängigen Variablen Strategie und Repräsentation tatsächlich Einfluss auf die in der Hauptstudie gemessenen Werte hatten und letztere nicht zufällig entstanden sind.



Abbildung 5.10: Paare (Linien) bei denen der Wilcoxon-Test (Offset deaktiviert) signifikante Unterschiede hinsichtlich der Fehlerrate zeigte. Die Grafik ist von links nach rechts zu lesen.

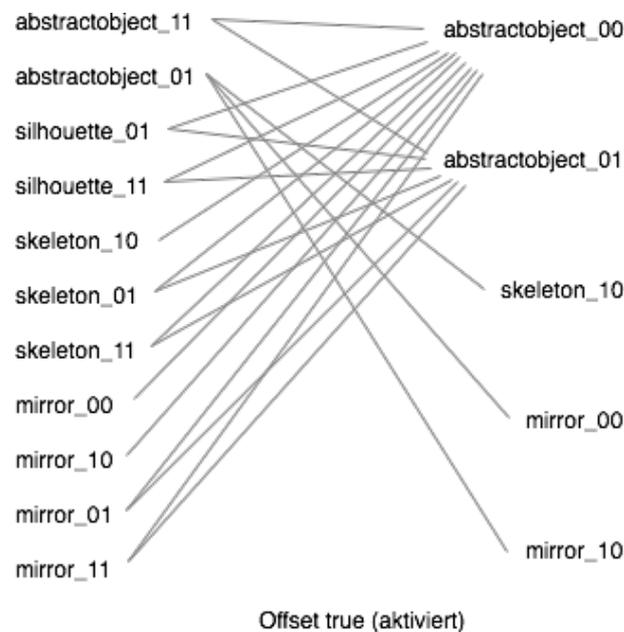


Abbildung 5.11: Paare (Linien) bei denen der Wilcoxon-Test (Offset aktiviert) signifikante Unterschiede hinsichtlich der Fehlerrate zeigte. Die Grafik ist von links nach rechts zu lesen.

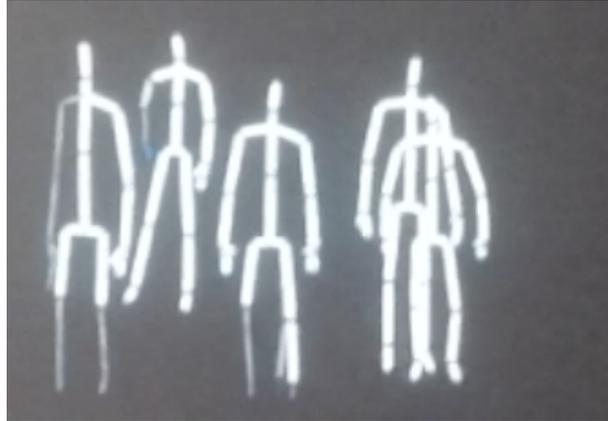


Abbildung 5.12: Zwei Skelett-Repräsentationen überlagern sich während der Hauptstudie (rechts im Bild). Die kleinere der beiden Darstellungen gehört zum Studienteilnehmer, die andere ist Teil der Aufnahme.

Einschränkungen

Sowohl durch technische als auch durch gruppenimmanente Umstände unterlag die Hauptstudie mehreren Einschränkungen. Da diese möglicherweise Einfluss auf die Ergebnisse hatten, wird nun genauer darauf eingegangen.

Eine dieser Einschränkungen bestand in der Entfernung der insgesamt 116 doppelten Werte. Hierdurch wurde die Datenbasis um 6,4 % verkleinert. Durch eine entsprechende Abfrage im Programmcode des Software-Prototyps hätte die Entstehung dieser Duplikate verhindert werden können.

Weitere Einschränkungen bestanden hinsichtlich der Aufnahmen und des Offsets.

Die Aufnahmen wurden vor allem deswegen an Stelle zusätzlicher echter Teilnehmer genutzt, da untersucht werden sollte, inwiefern sich Nutzer bei der Erkennung ihrer Repräsentationen durch die Repräsentationen anderer Nutzer beeinflussen lassen, beziehungsweise inwiefern die Erkennung dadurch erschwert wird. Dies wäre mit echten Teilnehmern nur sehr schwer möglich gewesen, da sehr viele Teilnehmer nötig gewesen wären, um den Teilnehmer, dessen Erkennungszeit gemessen werden sollte, 'abzulenken'. Ein Problem der Aufnahmen lag darin, dass die Teilnehmer während der Aufnahmen bei bestimmten Bedingungen *herumlaufen* sollten, was immer wieder dazu führte, dass manche Teilnehmer von anderen verdeckt wurden. Dadurch verschwanden die Repräsentationen der verdeckten Teilnehmer so lange, bis letztere für Kinect wieder sichtbar waren. Dies hatte zur Folge, dass die Teilnehmer der Hauptstudie nicht immer 6 Repräsentationen in der Aufnahme sahen, was die Erkennung der eigenen Repräsentation möglicherweise vereinfachte. Außerdem lassen sich die in der Hauptstudie ermittelten Messwerte dadurch nicht auf alle Interaktionen mit 6 weiteren Personen, die gleichzeitig mit einem Public Display interagieren, übertragen. Zudem lässt sich aus den gemessenen Werten nicht ableiten, wie sich diese verändern würden, wären weniger als 7 Personen gleichzeitig an der Interaktion mit einem Public Display beteiligt beziehungsweise weniger als 7 Repräsentationen gleichzeitig auf diesem zu sehen. Um Aussagen diesbezüglich treffen zu können, müssen weitere Studien durchgeführt werden.

Da die Live-Darstellung des Teilnehmers während der Hauptstudie aus technischen Gründen¹⁶ immer über der Aufnahme liegen musste, führte das in einigen Fällen dazu, dass Teile der

¹⁶Da nur der Hintergrund der Live-Darstellung, nicht aber der der Aufnahme transparent waren, war es nicht möglich, die Live-Darstellung unter die Aufnahme zu legen. Hätte man dies getan, wäre sie von der Aufnahme komplett verdeckt worden und daher nicht sichtbar gewesen.

Aufnahme durch diese verdeckt beziehungsweise überlagert wurden (siehe hierzu auch Abbildung 5.12). Möglicherweise sorgte dies dafür, dass die Erkennung der eigenen Repräsentation in solchen Fällen erschwert wurde.

Eine weitere Einschränkung betraf den Offset. Diese Funktion sorgte dafür, dass die Live-Darstellung um einen zufälligen Wert entweder nach rechts oder links verschoben wurde. Eine Skalierung der Größe fand jedoch nicht statt. Dadurch konnte es, je nach Bewegung der Teilnehmer beziehungsweise Repräsentationen in der Aufnahme dazu kommen, dass die Repräsentation des Teilnehmers größer war als die Repräsentationen in der Aufnahme. Gegebenenfalls erleichterte das den Studienteilnehmern ebenfalls die Erkennung ihrer Repräsentationen.

6 Diskussion

Im nun folgenden Kapitel werden die Ergebnisse der Hauptstudie diskutiert. Dabei wird auf einige, die Ergebnisse betreffende Fragestellungen, genauer eingegangen und es wird versucht Antworten auf diese zu finden.

Zusammenhang zwischen Erkennungszeiten und Repräsentationsformen/Strategien

Wie aus Abbildung 5.5 hervorgeht, können die Erkennungszeiten von Nutzerrepräsentationen - je nach eingesetzter Strategie - sehr stark schwanken. In der Abbildung zeigt sich, dass die Teilnehmer, vor allem beim Quadrat, zum Teil deutlich mehr Zeit benötigten als bei den übrigen Repräsentationsformen, um die eigene Repräsentation zu erkennen. Weshalb ist das so?

In Kapitel 4.2 wird erklärt, dass die in dieser Arbeit verwendete Implementierung des Quadrats die absolute Position des Nutzers im Raum von oben zeigt. Sie reagiert daher ausschließlich auf Bewegungen des Körpers in eine bestimmte Richtung. Bewegungen einzelner Gliedmaßen (wie beispielsweise der Arme oder Hände) werden nicht abgebildet. Daher können die Strategien *still stehen* und *Arme bewegen/winken* beim Quadrat nicht effektiv angewandt werden, was sich in der entsprechenden Zahl widerspiegelt.

In Bezug auf die Ergebnisse der übrigen Repräsentationsformen lässt sich Folgendes feststellen:

Die Erkennungszeiten der Repräsentationsform Silhouette ähneln sich - über alle Strategien hinweg - deutlich stärker als jene des Quadrats. Eine leicht erhöhte durchschnittliche Erkennungszeit zeigte sich hier lediglich unter Anwendung der Strategie *Stehen bleiben*. Dies zeigt, dass auch Repräsentationsformen, die der Klasse 'appearance matching' zugeordnet werden, etwas leichter erkannt werden können, wenn sich der Nutzer - in welcher Form auch immer - bewegen kann.

Das Skelett zeigt, wird die Strategie *stehen bleiben* angewandt, einen deutlichen Anstieg der Erkennungszeit. Dieser ist - im Vergleich zur Silhouette - deutlich stärker ausgeprägt was die Klassifizierung dieser Repräsentationsform ('kinesthetic matching') bestätigt und zeigt, dass eine schnelle Erkennung hier stark von der Möglichkeit der Bewegung profitiert. Die Strategie *Herumlaufen* zeigt hier ebenfalls einen, wenn auch sehr leichten, Anstieg im Vergleich zu den übrigen beiden Strategien. Grund hierfür sind möglicherweise die bereits beschriebenen Überlappungen von Live-Darstellung und Aufnahme, die die Erkennung eventuell etwas erschwerten.

Wie in Abbildung 5.5 zu sehen benötigten die Studienteilnehmer beim Spiegelbild - im Gegensatz zu allen anderen Repräsentationsformen - unter Anwendung der verschiedenen Strategien immer in etwa gleich viel Zeit, um ihre Repräsentationsform zu erkennen. Grund hierfür könnte sein, dass die Teilnehmer unterbewusst stets eher auf das Aussehen der Spiegelbilder achteten und weniger auf deren Bewegungen. Es liegt also die Vermutung nahe, dass die 'appearance matching'-Eigenschaft der Repräsentation Spiegelbild über alle Strategien hinweg dominant ist.

Offset

Wie bereits erwähnt wurden im Rahmen der Hauptstudie in jedem Durchgang zunächst die 16 Bedingungen (zur Erinnerung: 4 Repräsentationsformen \times 4 Strategien) mit aktiviertem und anschließend jene 16 mit deaktiviertem Offset getestet. War der Offset deaktiviert, wurde erwartet, dass die Teilnehmer dies bemerken würden und bis zu dessen Reaktivierung die bereits beschriebene Positions-Strategie anwenden würden, wodurch sich die Erkennungszeiten verkürzen hätten müssen. Wie in Kapitel 5.5.2 beschrieben, war dies bei 7 von 19 Teilnehmern

der Fall. Hier zeigten in mindestens 2 von 3 Durchgängen jene Bedingungen mit deaktiviertem Offset durchschnittlich kürzere Erkennungszeiten als jene mit aktiviertem Offset. Die Frage ist, weshalb dieser Effekt nicht häufiger in Erscheinung trat. Eine Antwort darauf könnte sein, dass der Großteil der Teilnehmer (14 von 19) mit Bedingungen startete, bei denen der Offset aktiviert war. Dadurch vermuteten sie möglicherweise, dass es nur Bedingungen mit aktiviertem Offset geben würde und bemerkten die mit deaktiviertem Offset nicht. Im Gegensatz dazu sahen die Teilnehmer der Vorstudie ausschließlich Bedingungen mit deaktiviertem Offset, da das Ziel der Vorstudie ja unter anderem darin bestand herauszufinden, ob die Positions-Strategie überhaupt angewandt wird.

Ebenso ist denkbar, dass 16 Bedingungen nicht genug waren, um zu verstehen, dass kein Offset aktiv ist. Im Rahmen der Vorstudie wurden insgesamt 24 Durchläufe pro Gruppe durchgeführt; im Rahmen der Hauptstudie 16 mit aktivem/inaktivem Offset pro Durchlauf.

Auch könnte es daran gelegen haben, dass die Teilnehmer der Hauptstudie die Repräsentationen immer nur solange betrachten durften, bis sie ihre Repräsentation erkannt hatten. Anschließend wurden sie wieder gebeten, ihre Augen zu schließen. Die Teilnehmer der Vorstudie konnten, im Gegensatz dazu, ihre Augen offen lassen und die Repräsentationen so lange betrachten, bis alle Teilnehmer ihre Repräsentationen erkannt hatten.

Lerneffekt

Wie in Kapitel 5.5.2 bereits erwähnt wurde, verkürzten sich die durchschnittlichen Erkennungszeiten von 10 der 32 Bedingungen sowohl im 2. als auch im 3. Durchgang. Auffällig hierbei ist, dass die Repräsentationen bei 6 dieser 10 Bedingungen im 2. Durchgang zwischen 22 und 33% schneller erkannt werden konnten. Außerdem fällt auf, dass der Offset bei allen der 6 Bedingungen deaktiviert war und in 5 der 6 Fälle die Strategie *stehen bleiben* Anwendung fand. Es liegt die Vermutung nahe, dass es sich hierbei tatsächlich um einen Lerneffekt handelte. Die Teilnehmer sahen die jeweiligen Repräsentationen zum zweiten Mal, kannten sie also bereits und konnten sie daher schneller erkennen.

Im 3. Durchgang fanden bei 7 der 10 Bedingungen, verglichen mit dem 2. Durchgang, kleinere Verkürzungen statt. In 3 Fällen war die Verkürzung im 3. Durchgang jedoch deutlich größer als die im 2. Durchgang. Da hierbei jedoch drei unterschiedliche Strategien angewandt wurden, lässt sich abschließend nicht klären, wodurch diese Werte hervorgerufen wurden.

7 Leitfaden zur Implementierung von Anwendungen

Basierend auf den Ergebnissen der Vor- und Hauptstudie sowie den bereits existierender Arbeiten wurde ein Leitfaden entwickelt, welcher Anwendungsentwicklern und -designern dabei behilflich sein soll, die passende Repräsentationsform für das jeweilige Anwendungsszenario zu finden. Der Leitfaden spricht Empfehlungen für Repräsentationsformen in Bezug auf Strategien, Privatsphäre, Erkennungsdauer, Interaktionsdauer und Präzision der Navigation aus. Im folgenden Kapitel wird dieser in Form einer Tabelle dargestellt.

Kriterium	Empfehlung
Strategien	Wie die Ergebnisse der Hauptstudie gezeigt haben, ist die Zeit, die benötigt wird, um die jeweiligen Repräsentationsformen zu erkennen, abhängig von der gewählten Strategie. Dies gilt jedoch auch umgekehrt. Bietet ein Anwendungsszenario potentiellen Nutzern kaum Bewegungsfreiheit, sollte die Repräsentationsform Quadrat nicht gewählt werden, da diese nur durch Bewegung des Körpers erkannt werden kann. Dasselbe gilt für das Skelett . Im Fall der Silhouette reicht es aus, wenn die <i>Bewegung der Arme</i> beziehungsweise das <i>Winken</i> ermöglicht werden. Um eine zuverlässige Erkennung der Repräsentationsform Spiegelbild zu ermöglichen, bedarf es keinerlei spezieller Strategien. Wie die Hauptstudie gezeigt hat, ist dies bereits zuverlässig möglich, wenn potentielle Nutzer einfach <i>still stehen</i> .
Privatsphäre	Im Rahmen der Hauptstudie äußerte ein Teilnehmer beim Spiegelbild Bedenken hinsichtlich des Datenschutzes. Soll der Datenschutz gewahrt bleiben, wird daher empfohlen - sofern es im Rahmen des jeweiligen Anwendungsszenarios möglich ist - alternativ auf die Silhouette zurückzugreifen. Diese Repräsentationsform zeigt keine Details wie Gesichtszüge oder Kleidung und kann daher bedenkenlos eingesetzt werden. Zu beachten ist jedoch, dass hier, zumindest eine Bewegung der Arme wenn nicht des ganzen Körpers möglich sein muss.
Erkennungsdauer	Steht die Erkennungsdauer im Fokus, empfiehlt sich die Nutzung der Repräsentationsformen Spiegelbild oder Silhouette. Erste konnte im Rahmen der Studie durchschnittlich in 1,14 Sekunden, letztere in 1,99 Sekunden erkannt werden. Kommen diese Repräsentationsformen nicht in Betracht, eignet sich alternativ auch das Skelett. Die Erkennungszeit betrug hier durchschnittlich 2,95 Sekunden. Fehler traten vor allem bei Anwendung der Strategie <i>still stehen</i> auf. Da diese in der Praxis vermutlich jedoch vor allem beim Spiegelbild angewandt wird, sind sie vernachlässigbar.

Interaktionsdauer	<p>Beobachtungen, die während der Vorstudie und auch im Rahmen anderer Arbeiten gemacht wurden, legen die Vermutung nahe, dass die Interaktion mit einem Public Display von längerer Dauer ist, wird die Repräsentationsform Skelett genutzt. Zudem zeigte sich, dass die Interaktion auf eher spielerische Art erfolgt, was möglicherweise auch dazu beiträgt, dass Nutzer länger motiviert sind, mit dem Display zu interagieren [1, 71]. Außerdem gibt es Hinweise, dass es sich bei der Silhouette um eine eher sachliche Interaktion handelt [2]. Es muss jedoch betont werden, dass es sich hierbei lediglich um Vermutungen handelt die sich auf andere Arbeiten stützen. Diese Aspekte wurden im Rahmen dieser Arbeit nicht untersucht.</p>
Präzision der Navigation	<p>Obwohl dies im Rahmen dieser Arbeit nicht untersucht wurde, besteht der Verdacht, dass das Quadrat eine präzisere Navigation zu einem bestimmten Punkt ermöglicht, da es, im Gegensatz zu den übrigen Repräsentationsformen, die absolute Position des Nutzers (von oben) darstellt. Sollte dem so sein, könnte es sich womöglich gut dafür eignen, potentielle Nutzer in den Sweetspot zu führen.</p>

8 Fazit

Das nun folgende Kapitel stellt den Abschluss dieser Arbeit dar. Es stellt die wichtigsten Punkte heraus und gibt einen Ausblick auf weitere mögliche Forschungsaspekte.

8.1 Zusammenfassung

Ziel dieser Arbeit war es herauszufinden, welche Strategien Nutzer anwenden, um unterschiedliche Formen von Nutzerrepräsentationen zu erkennen und wie gut diese (im Hinblick auf Erkennungsgeschwindigkeit und Fehlerrate) unter Zuhilfenahme jener Strategien erkannt werden können.

Dazu wurden zwei Studien durchgeführt. Der Durchführung der Studien ging die Entwicklung eines Software-Prototyps voraus, welcher es ermöglichte vier Nutzerrepräsentationen unter Anwendung der jeweiligen Strategien zu testen. Dieser nutzte Kinect der zweiten Generation, um die Teilnehmer der Studien optisch zu erfassen und die entsprechenden Repräsentationen zu generieren.

Bei der ersten Studie handelte es sich um eine explorative Vorstudie. Ihr Ziel bestand einerseits darin, herauszufinden welche Strategien Nutzer anwenden, um ihre eigene Nutzerrepräsentation zu erkennen. Es stellte sich heraus, dass die meisten Teilnehmer eine, mehrere oder alle der fünf folgenden Strategien nutzten:

- *Still stehen*, Identifikation durch betrachten/vergleichen von Aussehen, Statur und Haltung
- *Winken/Bewegen der Hände*
- *Herumlaufen*
- *Herumlaufen und gleichzeitiges Winken/Bewegen der Hände*
- *Positions-Strategie*

Des Weiteren konnte in der Vorstudie beobachtet werden, dass die Teilnehmer mit den jeweiligen Repräsentationsformen unterschiedlich interagierten. Vor allem beim Skelett war die Interaktion eher spielerischer Natur.

Im Anschluss an die Vorstudie wurde die Hauptstudie durchgeführt. Sie hatte zum Ziel herauszufinden, wie schnell und mit welcher Fehlerrate vier unterschiedliche Nutzerrepräsentationen (Quadrat, Silhouette, Skelett und Spiegelbild) unter Anwendung der bereits beschriebenen Strategien erkannt werden.

Es stellte sich heraus, dass die vier Repräsentationsformen deutliche Unterschiede in Bezug auf die durchschnittliche Erkennungszeit und Fehlerrate aufweisen. Zudem konnte festgestellt werden, dass bei den Repräsentationsformen Quadrat, Silhouette und Skelett Erkennungszeit und Fehlerrate von der eingesetzten Strategie und damit von der Funktionsweise beziehungsweise von den Eigenschaften der jeweiligen Repräsentationsform abhängig sind. Beispielsweise konnten mit dem Quadrat deutlich kürzere Erkennungszeiten erzielt werden, wenn die Strategie die Bewegung durch den Raum erforderte. Ebenso erzielte die Silhouette kürzere Erkennungszeiten als das Skelett sofern die Strategie *still stehen* Anwendung fand.

Schließlich wurde, basierend auf den Ergebnissen der Hauptstudie, ein Leitfaden erstellt. Dieser soll Entwicklern und Designern, die Anwendungen für Public Displays entwerfen, dabei helfen, die richtige Repräsentationsform für das jeweilige Anwendungsszenario zu finden. Um dies

zu ermöglichen, werden in Bezug auf fünf Kriterien, die im Kontext Nutzerrepräsentationen eine wichtige Rolle spielen, Handlungsempfehlungen ausgesprochen.

8.2 Ausblick

Wie in Kapitel 2 beschrieben wurde, stellen Nutzerrepräsentationen eine vielversprechende Technik dar, um potentielle Nutzer auf Public Displays aufmerksam zu machen, Interaktivität zu vermitteln und potentielle Nutzer an entsprechende Bildschirme heranzuführen. Daher wurde im Rahmen dieser Arbeit untersucht, wie gut diese Art von Repräsentationen überhaupt von Nutzern erkannt werden, vor allem unter Zuhilfenahme unterschiedlicher Strategien.

Nicht untersucht wurden jedoch längerfristige Auswirkungen dieser Repräsentationsformen auf die Motivation der Nutzer. Beispielsweise wäre es interessant zu erfahren, ob die Repräsentationsform Skelett tatsächlich länger andauernde Interaktions-Sitzungen zur Folge hat, wie Beobachtungen während der Vorstudie vermuten lassen. Einige Teilnehmer der Vorstudie waren von dieser Repräsentationsform sehr angetan, da diese spielerische Interaktionen ermöglicht.

Sollte dem so sein, wäre es womöglich lohnenswert, zweistufige Nutzerrepräsentationen genauer zu untersuchen. In entsprechenden Implementierungen könnte zunächst eine Repräsentationsform dargestellt werden, die von potentiellen Nutzern sehr schnell zu erkennen ist und die Aufmerksamkeit dieser auf sich zieht (beispielsweise das Spiegelbild). Erfolgt daraufhin eine Interaktion, könnte zu einer zweiten Repräsentationsform wie dem Skelett gewechselt werden, um eine spielerische Interaktion zu ermöglichen und so die Motivation potentieller Nutzer zu steigern. Die Frage ist jedoch, ob ein Wechsel von Repräsentationen nicht eher zu Verwirrungen der Nutzer führen würde.

Auch wurde bisher noch nicht erforscht, wie verschiedene Alters-/Zielgruppen auf die jeweiligen Repräsentationsformen reagieren und wie diese aussehen müssen, um deren Aufmerksamkeit erregen zu können. Beispielsweise ist es denkbar, dass Kinder eher mit Public Displays interagieren, auf denen den Spieltrieb fördernde Repräsentationsformen dargestellt werden. Ebenfalls könnte es in diesem Zusammenhang sehr hilfreich sein, an die jeweilige Alters-/Zielgruppe angepasste Avatar-basierte Repräsentationsformen näher zu untersuchen.

Inhalt der beigelegten DVD

- Zwei Versionen des Software-Prototyps in Form von VisualStudio-Projekten (inkl. verwendeter Frameworks).
Der Unterschied der beiden Versionen besteht darin, dass die erste Version noch nicht die finale Funktion enthält, um Nutzerrepräsentationen entlang der X-Achse zu verschieben (Offset). Sie fand in der Vorstudie Verwendung. In der zweiten Version, welche während der Hauptstudie zum Einsatz kam, ist besagte Funktion bereits fertig.
- Verwendete Quellen in Form von PDF-Dateien
- Digitale Version dieser Arbeit in Form einer .tex-Datei (inkl. aller verwendeter Abbildungen)
- Alle Ergebnisse der im Rahmen der Arbeit ausgewerteten Daten (Outlier Detection, Deskriptive Statistik, Friedman- und Wilcoxon-Test sowie gemischtes Lineares Modell)
- Aufnahmen, die im Vorfeld der Hauptstudie erstellt und im Rahmen dieser genutzt wurden. Alle weiteren Aufnahmen konnten aufgrund der zu großen Datenmenge (> 113 GB) nicht auf der DVD gespeichert werden. Sie wurden stattdessen auf einer Festplatte abgelegt und liegen dem Betreuer vor.
- Präsentation der Arbeit als .key- und .pdf-Datei

Anhang

Signifikant unterschiedliche Paare		p-Wert
abstractobject_10	abstractobject_00	$3,60 \times 10^{-04}$
silhouette_10	abstractobject_00	$3,69 \times 10^{-04}$
silhouette_01	abstractobject_00	$3,60 \times 10^{-04}$
silhouette_11	abstractobject_00	$3,60 \times 10^{-04}$
skeleton_10	abstractobject_00	$3,45 \times 10^{-04}$
skeleton_01	abstractobject_00	$3,60 \times 10^{-04}$
mirror_00	abstractobject_00	$3,60 \times 10^{-04}$
mirror_10	abstractobject_00	$3,60 \times 10^{-04}$
mirror_01	abstractobject_00	$3,60 \times 10^{-04}$
mirror_11	abstractobject_00	$3,60 \times 10^{-04}$

Tabelle A1: Ergebnis des Wilcoxon-Tests (Offset deaktiviert).

Signifikant unterschiedliche Paare		p-Wert
abstractobject_11	abstractobject_00	$3,60 \times 10^{-04}$
abstractobject_11	abstractobject_01	$3,60 \times 10^{-04}$
abstractobject_01	mirror_00	$3,60 \times 10^{-04}$
abstractobject_01	skeleton_10	$3,60 \times 10^{-04}$
abstractobject_01	mirror_10	$3,60 \times 10^{-04}$
silhouette_01	abstractobject_00	$3,42 \times 10^{-04}$
silhouette_01	abstractobject_01	$3,60 \times 10^{-04}$
silhouette_11	abstractobject_00	$3,42 \times 10^{-04}$
silhouette_11	abstractobject_01	$3,60 \times 10^{-04}$
skeleton_10	abstractobject_00	$3,42 \times 10^{-04}$
skeleton_01	abstractobject_00	$3,52 \times 10^{-04}$
skeleton_01	abstractobject_01	$3,65 \times 10^{-04}$
skeleton_11	abstractobject_00	$3,42 \times 10^{-04}$
skeleton_11	abstractobject_01	$3,60 \times 10^{-04}$
mirror_00	abstractobject_00	$3,42 \times 10^{-04}$
mirror_10	abstractobject_00	$3,42 \times 10^{-04}$
mirror_01	abstractobject_00	$3,42 \times 10^{-04}$
mirror_01	abstractobject_01	$3,60 \times 10^{-04}$
mirror_11	abstractobject_00	$3,42 \times 10^{-04}$
mirror_11	abstractobject_01	$3,60 \times 10^{-04}$

Tabelle A2: Ergebnis des Wilcoxon-Tests (Offset aktiviert).

Bedingung	1. Durchgang	2. Durchgang	Verkürzung Durchgang 2 (gegenüber 1)	3. Durchgang	Verkürzung Durchgang 3 (gegenüber 2)
abstractobject, 11, true	4181,43	4151,13	0,72%	3430,31	17,36%
abstractobject, 00, false	11743,78	8207,82	30,11%	7599,43	7,41%
abstractobject, 10, false	3355,25	3062,19	8,73%	2034,00	33,58%
silhouette, 00, false	3283,00	2545,89	22,45%	2333,89	8,33%
skeleton, 00, false	9237,94	7036,06	23,84%	6522,69	7,30%
skeleton, 10, false	3896,59	2820,00	27,63%	2342,71	16,93%
mirror, 01, true	1175,82	1033,94	12,07%	991,94	4,06%
mirror, 00, false	1707,67	1132,94	33,66%	1081,27	4,56%
mirror, 10, false	1257,33	1173,87	6,64%	1073,94	8,51%
mirror, 01, false	1122,61	992,57	11,58%	935,35	5,76%

Tabelle A3: Bedingungen, bei denen eine Verkürzung der durchschnittlichen Erkennungszeiten zwischen erstem und zweitem sowie zweitem und drittem Durchgang festgestellt wurde. **true** im Feld Bedingung bedeutet, dass der Offset hier aktiv war, **false** bedeutet, dass er inaktiv war.

Tests auf feste Effekte, Typ III^a

Quelle	Zähler-Freiheitsgrade	Nenner-Freiheitsgrade	F-Wert	Signifikanz
Konstanter Term	1	82,767	297,595	,000
Representation	3	107,086	43,146	,000
Movement_Type	3	69,980	9,441	,000
Offset	1	73,619	1,321	,254
Representation * Movement_Type	9	105,287	6,081	,000
Representation * Offset	3	107,081	,414	,743
Movement_Type * Offset	3	69,955	1,239	,302
Representation * Movement_Type * Offset	9	105,277	,687	,719

a. Abhängige Variable: Decision_Time.

Abbildung A1: Ergebnisse des linearen gemischten Modells (feste Effekte, 32 Bedingungen pro Teilnehmer, Representation entspricht Repräsentation, Movement_Type entspricht Strategie).

Paarweise Vergleiche^a

(I) Representation	(J) Representation	Signifikanz ^f	95% Konfidenzintervall für Differenz ^c	
			Untergrenze	Obergrenze
abstractob	mirror	,000	2033,145	6803,810
	silhouette	,003	1002,855	5896,429
	skeleton	,254	-646,625	4546,318
mirror	abstractobject	,000	-6803,810	-2033,145
	silhouette	,013	-1796,303	-141,368
	skeleton	,000	-3793,290	-1143,973
silhouette	abstractobject	,003	-5896,429	-1002,855
	mirror	,013	141,368	1796,303
	skeleton	,039	-2950,575	-49,017
skeleton	abstractobject	,254	-4546,318	646,625
	mirror	,000	1143,973	3793,290
	silhouette	,039	49,017	2950,575

Basiert auf geschätzten Randmitteln

*. Die Differenz der Mittelwerte ist in Stufe ,05 signifikant.

a. Abhängige Variable: Decision_Time.

c. Anpassung für Mehrfachvergleiche: Bonferroni.

Abbildung A2: Ergebnisse des linearen gemischten Modells (paarweise Vergleiche der Repräsentationen, 16 Bedingungen pro Teilnehmer, Offset aktiv, Representation entspricht Repräsentation, Movement_Type entspricht Strategie).

Paarweise Vergleiche^a

(I) Representation	(J) Representation	Signifikanz ^f	95% Konfidenzintervall für Differenz ^c	
			Untergrenze	Obergrenze
abstractobject	mirror	,000	2629,320	8467,033
	silhouette	,002	1302,702	7223,191
	skeleton	,040	90,415	6107,855
mirror	abstractobject	,000	-8467,033	-2629,320
	silhouette	,001	-2144,886	-425,574
	skeleton	,000	-3512,569	-1385,514
silhouette	abstractobject	,002	-7223,191	-1302,702
	mirror	,001	425,574	2144,886
	skeleton	,059	-2353,551	25,928
skeleton	abstractobject	,040	-6107,855	-90,415
	mirror	,000	1385,514	3512,569
	silhouette	,059	-25,928	2353,551

Basiert auf geschätzten Randmitteln

*. Die Differenz der Mittelwerte ist in Stufe ,05 signifikant.

a. Abhängige Variable: Decision_Time.

c. Anpassung für Mehrfachvergleiche: Bonferroni.

Abbildung A3: Ergebnisse des linearen gemischten Modells (paarweise Vergleiche der Repräsentationen, 16 Bedingungen pro Teilnehmer, Offset inaktiv, Representation entspricht Repräsentation, Movement_Type entspricht Strategie).

Paarweise Vergleiche^a

(I) Movement_Type	(J) Movement_Type	Signifikanz ^b	95% Konfidenzintervall für Differenz ^c	
			Untergrenze	Obergrenze
00	01	1,000	-2406,529	3498,096
	10	,035	108,710	4346,660
	11	,004	810,755	4917,102
01	00	1,000	-3498,096	2406,529
	10	,405	-860,876	4224,680
	11	,078	-174,062	4810,352
10	00	,035	-4346,660	-108,710
	01	,405	-4224,680	860,876
	11	,683	-438,250	1710,736
11	00	,004	-4917,102	-810,755
	01	,078	-4810,352	174,062
	10	,683	-1710,736	438,250

Basiert auf geschätzten Randmitteln

*. Die Differenz der Mittelwerte ist in Stufe ,05 signifikant.

a. Abhängige Variable: Decision_Time.

c. Anpassung für Mehrfachvergleiche: Bonferroni.

Abbildung A4: Ergebnisse des linearen gemischten Modells (paarweise Vergleiche der Strategien, 16 Bedingungen pro Teilnehmer, Offset inaktiv, Representation entspricht Repräsentation, Movement_Type entspricht Strategie).

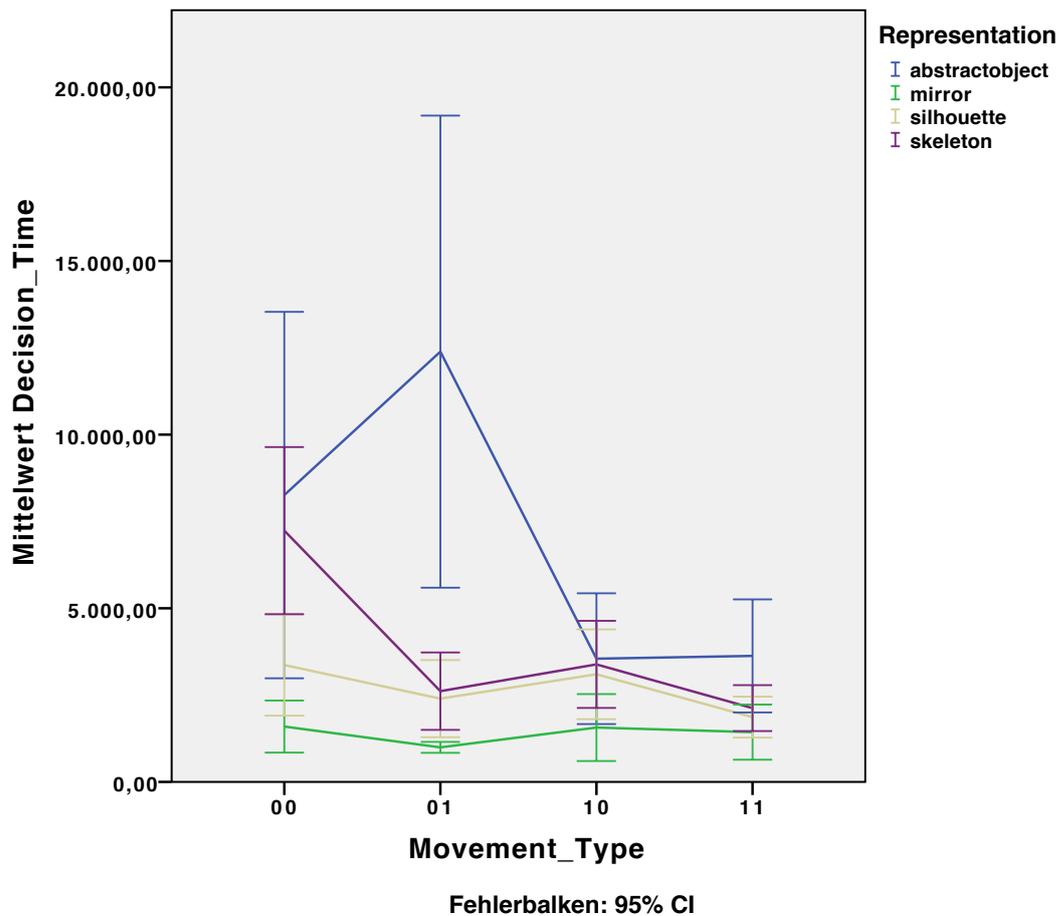


Abbildung A5: Verteilungen der durchschnittliche Erkennungszeiten über alle Teilnehmer hinweg abhängig von der gewählten Strategie. Die Balken stellen jeweils 95 % der Werte dar, für die die Entscheidungen der Teilnehmer in der Hauptstudie korrekt waren. Enthalten sind ausschließlich jene Bedingungen ohne aktiven Offset.

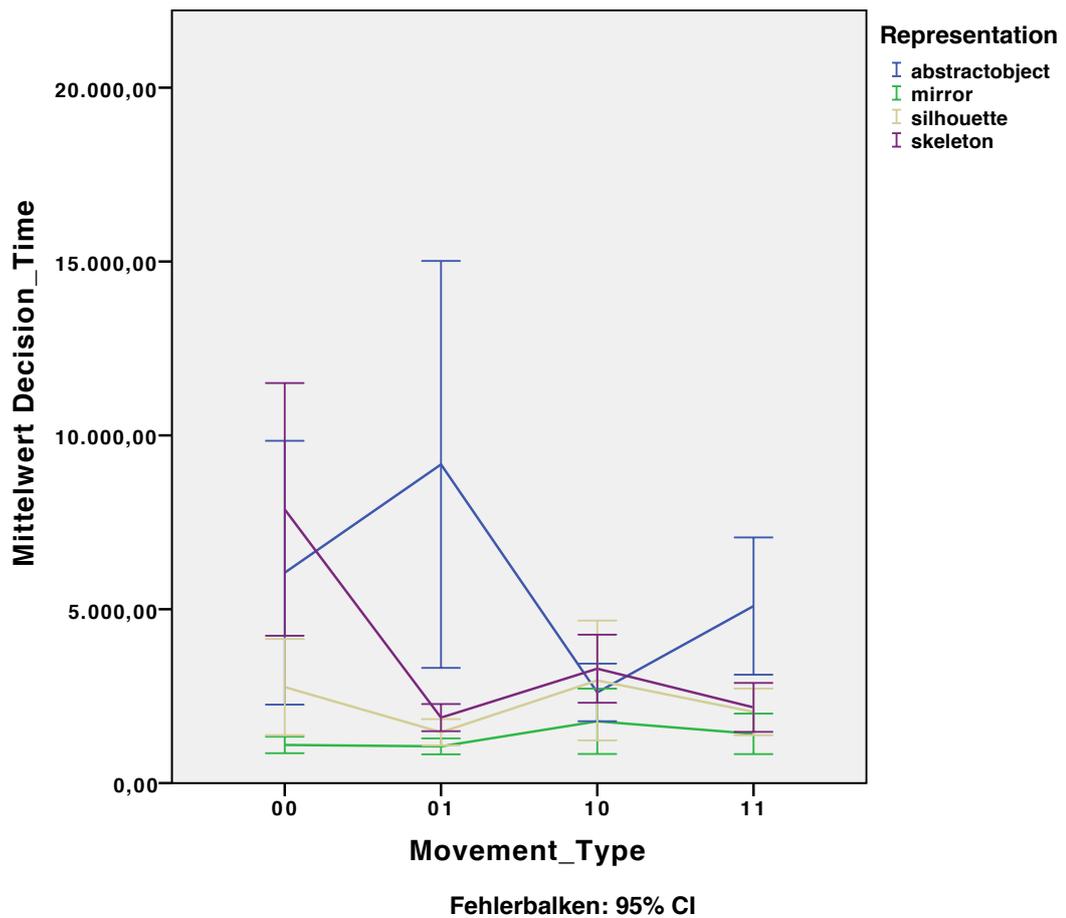


Abbildung A6: Verteilungen der durchschnittliche Erkennungszeiten über alle Teilnehmer hinweg abhängig von der gewählten Strategie. Die Balken stellen jeweils 95 % der Werte dar, für die die Entscheidungen der Teilnehmer in der Hauptstudie korrekt waren. Enthalten sind ausschließlich jene Bedingungen mit aktivem Offset.

Guiding Multiple People in Front of Public Displays - Feedback

* Erforderlich

1. **Wie lange denken Sie hat es gedauert bis Sie Ihre Darstellung des Quadrats erkannt haben? ***

Markieren Sie nur ein Oval.

1	2	3	4	5	
Es ging sehr schnell	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Es dauerte sehr lange

2. **Wie lange denken Sie hat es gedauert bis Sie Ihre Darstellung der Silhouette erkannt haben? ***

Markieren Sie nur ein Oval.

1	2	3	4	5	
Es ging sehr schnell	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Es dauerte sehr lange

3. **Wie lange denken Sie hat es gedauert bis Sie Ihre Darstellung des Skeletts erkannt haben? ***

Markieren Sie nur ein Oval.

1	2	3	4	5	
Es ging sehr schnell	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Es dauerte sehr lange

4. **Wie lange denken Sie hat es gedauert bis Sie Ihre Darstellung des Spiegelbildes erkannt haben? ***

Markieren Sie nur ein Oval.

1	2	3	4	5	
Es ging sehr schnell	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Es dauerte sehr lange

5. **Stellen Sie sich vor Sie wollen mit einem Display im öffentlichen Raum mittels Bewegung interagieren. Dabei sollen Nutzer ihre Darstellung möglichst schnell und präzise erkennen können. Welche der vier Darstellungen eignet sich Ihrer Meinung nach AM BESTEN dafür? ***

Markieren Sie nur ein Oval.

- Quadrat
- Silhouette
- Skelett
- Spiegelbild

6. Stellen Sie sich vor Sie wollen mit einem Display im öffentlichen Raum mittels Bewegung interagieren. Dabei sollen Nutzer ihre Darstellung möglichst schnell und präzise erkennen können. Welche der vier Darstellungen eignet sich Ihrer Meinung nach AM ZWEITBESTEN dafür? *

Markieren Sie nur ein Oval.

- Quadrat
- Silhouette
- Skelett
- Spiegelbild

7. Stellen Sie sich vor Sie wollen mit einem Display im öffentlichen Raum mittels Bewegung interagieren. Dabei sollen Nutzer ihre Darstellung möglichst schnell und präzise erkennen können. Welche der vier Darstellungen eignet sich Ihrer Meinung nach AM DRITTBESTEN dafür? *

Markieren Sie nur ein Oval.

- Quadrat
- Silhouette
- Skelett
- Spiegelbild

8. Stellen Sie sich vor Sie wollen mit einem Display im öffentlichen Raum mittels Bewegung interagieren. Dabei sollen Nutzer ihre Darstellung möglichst schnell und präzise erkennen können. Welche der vier Darstellungen eignet sich Ihrer Meinung nach AM VIERTBESTEN dafür? *

Markieren Sie nur ein Oval.

- Quadrat
- Silhouette
- Skelett
- Spiegelbild

9. Weshalb haben Sie sich in den letzten Fragen für genau diese Reihenfolge entschieden? *

.....

.....

.....

.....

.....

Guiding Multiple People in Front of Public Displays - Demographische Angaben

Angaben werden nicht mit anderen Daten in Verbindung gebracht

10. Größe *

.....

11. Alter *

.....

Abbildung A8: Zweite Seite des Feedback-Fragebogens der Hauptstudie.

1.7.2016

Guiding Multiple People in Front of Public Displays - Feedback

12. Geschlecht *

Wählen Sie alle zutreffenden Antworten aus.

Männlich

Weiblich

Sonstiges:

13. Beruf/Studiengang *

.....

Bereitgestellt von



Abbildung A9: Dritte Seite des Feedback-Fragebogens der Hauptstudie.

Literatur

- [1] Christopher Ackad, Andrew Clayphan, Martin Tomitsch und Judy Kay. »An In-the-wild Study of Learning Mid-air Gestures to Browse Hierarchical Information at a Large Interactive Public Display«. In: *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. Osaka, Japan: ACM, 2015, S. 1227–1238.
- [2] Christopher Ackad, Martin Tomitsch und Judy Kay. »Skeletons and Silhouettes: Comparing User Representations at a Gesture-based Large Display«. In: *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. San Jose, CA, USA: ACM, 2016, S. 2343–2347.
- [3] Florian Alt, Andreas Bulling, Gino Gravanis und Daniel Buschek. »GravitySpot : Guiding Users in Front of Public Displays Using On-Screen Visual Cues«. In: *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology*. Charlotte, NC, USA: ACM, 2015, S. 47–56.
- [4] Stavros Antifakos und Bernt Schiele. »LaughingLily: Using a flower as a real-world information display«. In: *Proceedings of the 5th International Conference on Ubiquitous Computing*. Seattle, WA, USA: Springer, 2003, S. 161–162.
- [5] Rafael Ballagas, Michael Rohs und Jennifer G Sheridan. »Sweep and Point & Shoot: Phocam-Based Interactions for Large Public Displays«. In: *CHI '05 Extended Abstracts on Human Factors in Computing Systems*. Portland, OR, USA: ACM, 2005, S. 1200–1203.
- [6] Thomas Baudel und Michel Beaudouin-Lafon. »Charade: Remote Control of Objects Using Free-hand Gestures«. In: *Communications of the ACM* 36.7 (1993), S. 28–35.
- [7] Gilbert Beyer, Florian Alt, Jörg Müller, Albrecht Schmidt, Karsten Isakovic, Stefan Klose, Manuel Schiewe und Ivo Haußen. »Audience Behavior Around Large Interactive Cylindrical Screens«. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11)*. Vancouver, BC, Canada: ACM, 2011, S. 1021–1030.
- [8] Gilbert Beyer, Vincent Binder, Nina Jäger und Andreas Butz. »The Puppeteer Display : Attracting and Actively Shaping the Audience with an Interactive Public Banner Display«. In: *Proceedings of the International Conference on Designing Interactive System (DIS'14)*. Vancouver, BC, Canada: ACM, 2014, S. 935–944.
- [9] Marc Böhlen und Michael Mateas. »Office Plant No. 1: Intimate Space and Contemplative Entertainment«. In: *Leonardo* 31.5 (1998), S. 345–348.
- [10] Sebastian Boring, Dominikus Baur, Andreas Butz, Sean Gustafson und Patrick Baudisch. »Touch Projector: Mobile Interaction Through Video«. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11)*. Atlanta, GA, USA: ACM, 2010, S. 2287–2296.
- [11] Sebastian Boring, Marko Jurmu und Andreas Butz. »Scroll, tilt or move it: using mobile phones to continuously control pointers on large public displays«. In: *Proceedings of the 21st Annual Conference of the Australian Computer-Human Interaction Special Interest Group: Design: Open 24/7 (OZCHI '09)*. Launceston, Tasmania: ACM, 2009.
- [12] Richard Daniel Borovoy. »Folk computing: designing technology to support face-to-face community building«. Diss. Massachusetts Institute of Technology, 2002.
- [13] Richard Daniel Borovoy, Fred Martin, Mitchel Resnick und Brian Silverman. »GroupWear: Nametags That Tell About Relationships«. In: *Conference Summary on Human Factors in Computing Systems (CHI '98)*. Los Angeles, CA, USA: ACM, 1998, S. 329–330.

- [14] Richard Daniel Borovoy, Fred Martin, Sunil Vemuri, Mitchel Resnick, Brian Silverman und Chris Hancock. »Meme Tags and Community Mirrors: Moving from Conferences to Collaboration«. In: *Proceedings of the 1998 ACM Conference on Computer Supported Cooperative Work (CSCW '98)*. Seattle, WA, USA: ACM, 1998, S. 159–168.
- [15] Richard Daniel Borovoy, M. McDonald, F. Martin und M. Resnick. »Things that blink: Computationally augmented name tags«. In: *IBM Systems Journal* 35.3-4 (1996), S. 488–495.
- [16] Harry Brignull und Yvonne Rogers. »Enticing people to interact with large public displays in public spaces«. In: *Ninth IFIP TC13 International Conference on Human-Computer Interaction (INTERACT '03)*. Zurich, Switzerland: IOS Press, 2003, S. 17–24.
- [17] Gregor Broll, Wolfgang Reithmeier, Paul Holleis und Matthias Wagner. »Design and Evaluation of Techniques for Mobile Interaction with Dynamic NFC-displays«. In: *Proceedings of the Fifth International Conference on Tangible, Embedded, and Embodied Interaction (TEI '11)*. Funchal, Portugal: ACM, 2011, S. 205–212.
- [18] Keith Cheverst, Faisal Taher, Matthew Fisher, Daniel Fitton und Nick Taylor. »The design, deployment and evaluation of situated display-based systems to support coordination and community«. In: *Ubiquitous Display Environments*. Heidelberg, New York, Dordrecht, London: Springer, 2012, S. 105–124.
- [19] Elizabeth F. Churchill, Les Nelson, Laurent Denoue und Andreas Girgensohn. »The Plasma Poster Network: Posting Multimedia Content in Public Places«. In: *Proceedings of the 9th IFIP TC13 International Conference on Human-Computer Interaction (INTERACT'03)*. Zurich, Switzerland: ACM, 2003, S. 599–606.
- [20] Nigel Davies, Sarah Clinch und Florian Alt. »Pervasive Displays: Understanding the Future of Digital Signage«. In: *Synthesis Lectures on Mobile and Pervasive Computing* 8.1 (2014), S. 1–128.
- [21] Tilman Dingler, Tobias Bagg, Yves Grau und Niels Henze. »uCanvas: A Web Framework for Spontaneous Smartphone Interaction with Ubiquitous Displays«. In: *Proceedings of the 15th IFIP TC 13 International Conference (INTERACT '15)*. Bamberg, Germany: Springer, 2015, S. 402–409.
- [22] Douglas Easterly. »Bio-Fi: Inverse Biotelemetry Projects«. In: *Proceedings of the 12th Annual ACM International Conference on Multimedia (MM '04)*. New York, NY, USA: ACM, 2004, S. 182–183.
- [23] Martin J. Eppler und Jeanne Mengis. »The Concept of Information Overload - A Review of Literature from Organization Science, Accounting, Marketing, MIS, and Related Disciplines«. In: *Kommunikationsmanagement im Wandel: Beiträge aus 10 Jahren*. Hrsg. von Miriam Meckel und Beat F. Schmid. Wiesbaden: Gabler, 2008, S. 271–305.
- [24] Joe Finney und Nigel Davies. »FLUMP: The FLEXible Ubiquitous Monitor Project«. In: *Proceedings of the the 3rd Cabernet Radicals Workshop*. CaberNet, 1996.
- [25] Robert S. Fish, Robert E. Kraut und Barbara L. Chalfonte. »The VideoWindow System in Informal Communication«. In: *Proceedings of the 1990 ACM Conference on Computer-supported Cooperative Work (CSCW '90)*. Los Angeles, California, USA: ACM, 1990, S. 1–11.
- [26] Dustin Freeman, Hrvoje Benko, Meredith Ringel Morris und Daniel Wigdor. »Shadow-Guides: Visualizations for In-situ Learning of Multi-touch and Whole-hand Gestures«. In: *Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces (ITS '09)*. Banff, Alberta, Canada: ACM, 2009, S. 165–172.

- [27] Adrian Friday, Nigel Davies und Christos Efstratiou. »Reflections on long-term experiments with public displays«. In: *Computer* 45.5 (2012), S. 34–41.
- [28] Jefferson Y. Han. »Low-cost multi-touch sensing through frustrated total internal reflection«. In: *Proceedings of the 18th annual ACM symposium on User interface software and technology (UIST '05)* (2005), S. 115–118.
- [29] David Holstius, John Kembel, Amy Hurst, Peng-Hui Wan und Jodi Forlizzi. »Infotropism: living and robotic plants as interactive displays«. In: *Proceedings of the 5th conference on Designing interactive systems: processes, practices, methods, and techniques (DIS '04)*. Cambridge, MA, USA: ACM, 2004, S. 215–221.
- [30] Elaine M. Huang, Anna Koster und Jan Borchers. »Overcoming assumptions and uncovering practices: When does the public really look at public displays?«. In: *Proceedings of the 6th International Conference on Pervasive Computing (Pervasive '08)*. Sydney, Australia: Springer, 2008, S. 228–243.
- [31] Giulio Jacucci, Ann Morrison, Gabriela T. Richard, Jari Kleimola, Peter Peltonen, Lorenza Parisi und Toni Laitinen. »Worlds of Information: Designing for Engagement at a Public Multi-touch Display«. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '10)*. Atlanta, GA, USA: ACM, 2010, S. 2267–2276.
- [32] Nassim Jafarinaini, Jodi Forlizzi, Amy Hurst und John Zimmerman. »Breakaway: An Ambient Display Designed to Change Human Behavior«. In: *CHI '05 Extended Abstracts on Human Factors in Computing Systems*. Portland, OR, USA: ACM, 2005, S. 1945–1948.
- [33] Hao Jiang, Eyal Ofek, Neema Moraveji und Yuanchun Shi. »Direct Pointer - Direct Manipulation for Large-Display Interaction using Handheld Cameras«. In: *Proceedings of the International Conference on Human Factors in Computing Systems (CHI'06)*. Montreal, Quebec, Canada: ACM, 2006, S. 1107–1110.
- [34] Wendy Ju und David Sirkin. »Animate Objects: How Physical Motion Encourages Public Interaction«. In: *Proceedings of the 5th international conference on Persuasive Technology (PERSUASIVE '10)*. Copenhagen, Denmark: Springer, 2010, S. 40–51.
- [35] Mohamed Khamis, Florian Alt und Andreas Bulling. »A field study on spontaneous gaze-based interaction with a public display using pursuits«. In: *Proceedings of the ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '15)*. Tokyo, Japan: ACM, 2015, S. 863–872.
- [36] David Kieras. »Using the keystroke-level model to estimate execution times«. In: *University of Michigan* (2001).
- [37] Bill Kules, Hyunmo Kang, Catherine Plaisant, Anne Rose und Ben Shneiderman. »Immediate usability: A case study of public access design for a community photo library«. In: *Interacting with Computers* 16.6 (2004), S. 1171–1193.
- [38] Ekaterina Kurdyukova, Mohammad Obaid und Elisabeth André. »Direct, Bodily or Mobile Interaction? Comparing Interaction Techniques for Personalized Public Displays«. In: *Proceedings of the 11th International Conference on Mobile and Ubiquitous Multimedia (MUM'12)* (2012), S. 44–52.
- [39] Satoshi Kuribayashi und Akira Wakita. »PlantDisplay: Turning Houseplants into Ambient Display«. In: *Proceedings of the 2006 ACM SIGCHI International Conference on Advances in Computer Entertainment Technology (ACE '06)*. Los Angeles, California, USA: ACM, 2006.
- [40] Gordon Kurtenbach und Eric A Hulteen. »Gestures in human-computer communication«. In: *The art of human-computer interface design* (1990), S. 309–317.

- [41] Christian Lander, Sven Gehring, Antonio Krüger, Sebastian Boring und Andreas Bulling. »GazeProjector: Accurate Gaze Estimation and Seamless Gaze Interaction Across Multiple Displays«. In: *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology (UIST '15)*. 395–404. Charlotte, NC, USA: ACM, 2015.
- [42] Anil Madhavapeddy, David Scott, Richard Sharp und Eben Upton. »Using camera-phones to enhance human-computer interaction«. In: *Proceedings of the 6th International Conference on Ubiquitous Computing (UbiComp '04)*. Nottingham, UK: Springer, 2004.
- [43] Jennifer Mankoff und Bill Schilit. »Supporting Knowledge Workers Beyond the Desktop with Palplates«. In: *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems (CHI '97)*. Atlanta, GA, USA: ACM, 1997, S. 550–551.
- [44] Paul Marshall, Richard Morris, Yvonne Rogers, Stefan Kreitmayer und Matthew Davies. »Rethinking 'Multi-User' - An In-The-Wild Study of How Groups Approach a Walk-Up-and-Use Tabletop Interface«. In: *Proceedings of ACM CHI 2011 Conference on Human Factors in Computing Systems (CHI '11)*. Vancouver, BC, Canada: ACM, 2011, S. 3033–3042.
- [45] Daniel Michelis und Jörg Müller. »The audience funnel: Observations of gesture based interaction with multiple large displays in a city center«. In: *Intl. Journal of Human-Computer Interaction* 27.6 (2011), S. 562–579.
- [46] Stanley Milgram. »The Experience of Living in Cities«. In: *Science* 167.3924 (1970), S. 1461–1468.
- [47] Jörg Müller, Gilles Bailly, Thor Bossuyt und Niklas Hillgren. »MirrorTouch: Combining Touch and Mid-air Gestures for Public Displays«. In: *Proceedings of the 16th international conference on Human-computer interaction with mobile devices & services (MobileHCI '14)*. Toronto, ON, Canada: ACM, 2014, S. 319–328.
- [48] Jörg Müller, Dieter Eberle und Konrad Tollmar. »Communiplay: a field study of a public display mediaspace«. In: *Proceedings of the 32nd annual ACM conference on Human factors in computing systems (CHI '14)*. Toronto, ON, Canada: ACM, 2014, S. 1415–1424.
- [49] Jörg Müller, Oliver Paczkowski und Antonio Krüger. *Situated Public News and Reminder Displays*. Bd. 4794/2007. Ambient Intelligence. 2007, S. 248–265.
- [50] Jörg Müller, Robert Walter, Gilles Bailly, Michael Nischt und Florian Alt. »Looking Glass: A Field Study on Noticing Interactivity of a Shop Window«. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '12)*. Austin, TX, USA: ACM, 2012, S. 297–306.
- [51] Jörg Müller, Dennis Wilmsmann, Juliane Exeler, Markus Buzbeck, Albrecht Schmidt, Tim Jay und Antonio Krüger. »Display blindness: The effect of expectations on attention towards digital signage«. In: *Proceedings of the 7th International Conference on Pervasive Computing (Pervasive '09)*. Galveston, TX, USA: Springer, 2009, S. 1–8.
- [52] Mathieu Nancel, Julie Wagner, Emmanuel Pietriga, Olivier Chapuis und Wendy Mackay. »Mid-air pan-and-zoom on wall-sized displays«. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11)*. Vancouver, BC, Canada: ACM, 2011, S. 177–186.
- [53] David H Nguyen, Joseph Tullio, Tom Drewes und Elizabeth D Mynatt. *Dynamic door displays*. Techn. Ber. Georgia Institute of Technology, 2000.

- [54] Tongyan Ning, Jörg Müller, Robert Walter, Gilles Bailly, Chachatvan Wacharamanotham, Jan Borchers und Florian Alt. »No Need To Stop: Menu Techniques for Passing by Public Displays«. In: *Workshop on Large Displays in Urban Life at International Conference on Human Factors in Computing Systems (CHI '11)*. Vancouver, BC, Canada: ACM, 2011.
- [55] T. Ojala, V. Kostakos, H. Kukka, T. Heikkinen, T. Linden, M. Jurmu, S. Hosio, F. Kruger und D. Zanni. »Multipurpose Interactive Public Displays in the Wild: Three Years Later«. In: *Computer* 45.5 (2012), S. 42–49.
- [56] Gonzalo Parra, Joris Klerkx und Erik Duval. »Understanding Engagement with Interactive Public Displays: An Awareness Campaign in the Wild«. In: *Proceedings of The International Symposium on Pervasive Displays (PerDis '14)*. Copenhagen, Denmark: ACM, 2014, S. 180–185.
- [57] Peter Peltonen, Esko Kurvinen, Antti Salovaara, Giulio Jacucci, Tommi Ilmonen, John Evans, Antti Oulasvirta und Petri Saarikko. »It's Mine, Don't Touch!: Interactions at a Large Multi-Touch Display in a City Centre«. In: *Proceedings of the 26th annual CHI conference on Human factors in computing systems (CHI '08)*. Florence, Italy: ACM, 2008, S. 1285–1294.
- [58] Stuart Reeves, Steve Benford, Claire O'Malley und Mike Fraser. »Designing the Spectator Experience«. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '05)*. Portland, OR, USA: ACM, 2005, S. 741–750.
- [59] Roy Rodenstein. »Employing the periphery: the window as interface«. In: *CHI '99 Extended Abstracts on Human Factors in Computing Systems*. Pittsburgh, PA, USA: ACM, 1999, S. 204.
- [60] Yvonne Rogers, William R. Hazlewood, Paul Marshall, Nick Dalton und Susanna Hertrich. »Ambient Influence: Can Twinkly Lights Lure and Abstract Representations Trigger Behavioral Change?«. In: *Proceedings of the 12th ACM International Conference on Ubiquitous Computing (UbiComp '10)*. Copenhagen, Denmark: ACM, 2010, S. 261–270.
- [61] Michael Rohs. »Real-world Interaction with Camera Phones«. In: *Proceedings of the Second International Conference on Ubiquitous Computing Systems (UCS '04)*. Tokyo, Japan: Springer, 2005, S. 74–89.
- [62] Elisa Rubegni, Nemanja Memarovic und Marc Langheinrich. »Talking to Strangers: Using Large Public Displays to Facilitate Social Interaction«. In: *Design, User Experience, and Usability. Theory, Methods, Tools and Practice: First International Conference, DUXU 2011, Held as Part of HCI International 2011, Orlando, FL, USA, July 9-14, 2011, Proceedings, Part II*. Hrsg. von Aaron Marcus. Berlin, Heidelberg: Springer, 2011, S. 195–204.
- [63] Enrico Rukzio, Gregor Broll, Karin Leichtenstern und Albrecht Schmidt. »Mobile Interaction with the Real World: An Evaluation and Comparison of Physical Mobile Interaction Techniques«. In: *Proceedings of the 2007 European Conference on Ambient Intelligence (AmI '07)*. Darmstadt, Germany: Springer, 2007, S. 1–18.
- [64] Enrico Rukzio, Karin Leichtenstern, Vic Callaghan, Paul Holleis, Albrecht Schmidt und Jeannette Chin. »An Experimental Comparison of Physical Mobile Interaction Techniques: Touching, Pointing and Scanning«. In: *Proceedings of the 8th International Conference on Ubiquitous Computing (UbiComp '06)*. Orange County, CA, USA: Springer, 2006, S. 87–104.

- [65] Khoovirajsingh Seewoonauth, Enrico Rukzio, Robert Hardy und Paul Holleis. »Touch & Connect and Touch & Select: Interacting with a Computer by Touching It with a Mobile Phone«. In: *Proceedings of the 11th International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI '09)*. Bonn, Germany: ACM, 2009, S. 36–45.
- [66] Sophie Stellmach und Raimund Dachsel. »Investigating gaze-supported multimodal pan and zoom«. In: *Proceedings of the Symposium on Eye Tracking Research and Applications (ETRA '12)*. Santa Barbara, CA, USA: ACM, 2012, S. 357–360.
- [67] Nick Taylor und Keith Cheverst. »Social interaction around a rural community photo display«. In: *International Journal of Human Computer Studies* 67.12 (2009), S. 1037–1047.
- [68] Nick Taylor, Keith Cheverst, Dan Fitton, Nicholas J. P. Race, Mark Rouncefield und Connor Graham. »Probing Communities: Study of a Village Photo Display«. In: *Proceedings of the 19th Australasian Conference on Computer-Human Interaction: Entertaining User Interfaces (OZCHI '07)*. Adelaide, Australia: ACM, 2007, S. 17–24.
- [69] Tobii® Technology. *Tobii Eye Tracking - An introduction to eye tracking and Tobii Eye Trackers*. Techn. Ber. 2010. URL: <http://acuity-ets.com/downloads/Tobii%20Eye%20Tracking%20Introduction%20Whitepaper.pdf>.
- [70] Maurice Ten Koppel, Gilles Bailly, Jörg Müller und Robert Walter. »Chained displays: Configurations of Public Displays Can Be Used to Influence Actor-, Audience-, and Passer-By Behavior«. In: *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems (CHI '12)*. Austin, TX, USA: ACM, 2012, S. 317–326.
- [71] Martin Tomitsch, Christopher James Ackad, Oliver Dawson, Luke Hespanhol und Judy Kay. »Who Cares About the Content? An Analysis of Playful Behaviour at a Public Display«. In: *Proceedings of The International Symposium on Pervasive Displays (PerDis '14)*. Copenhagen, Denmark: ACM, 2014, S. 160–165.
- [72] Robert Walter, Gilles Bailly und Jörg Müller. »StrikeAPose: revealing mid-air gestures on public displays«. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. Paris, France: ACM, 2013, S. 841–850.
- [73] Robert Walter, Andreas Bulling, David Lindlbauer, Martin Schuessler und Jörg Müller. »Analyzing visual attention during whole body interaction with public displays«. In: *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '15)*. Toyko, Japan: ACM, 2015, S. 1263–1267.
- [74] R Want, A Hopper, V Falcao und J Gibbons. »The active badge location system«. In: *ACM Transactions on Information Systems (TOIS)* 10.1 (1992), S. 91–102.
- [75] Daniel Wigdor und Dennis Wixon. *Brave NUI World: Designing Natural User Interfaces for Touch and Gesture*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2011.
- [76] Yanxia Zhang, Andreas Bulling und Hans Gellersen. »SideWays: A Gaze Interface for Spontaneous Interaction with Situated Displays«. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13)*. Paris, France: ACM, 2013, S. 851–860.
- [77] Yanxia Zhang, Jörg Müller, Ming Ki Chong, Andreas Bulling und Hans Gellersen. »Gaze-Horizon: Enabling Passers-by to Interact with Public Displays by Gaze«. In: *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '14)*. ACM, 2014, S. 559–563.

Web-Referenzen

- [78] AVerMedia. *LGPLite*. URL: http://avermedia.com/gaming/product/game_capture/lgp_lite (besucht am 04.06.2016).
- [79] Channel 9. *Programming Kinect for Windows v2: (02) Kinect Data Sources and Programming Model*. URL: <https://channel9.msdn.com/Series/Programming-Kinect-for-Windows-v2/02> (besucht am 02.06.2016).
- [80] Daily Mail Online. *The amazing telectroscope that lets you see New York from London's Tower Bridge*. URL: <http://www.dailymail.co.uk/sciencetech/article-1021258/The-amazing-telescope-lets-New-York-Londons-Tower-Bridge.html> (besucht am 16.06.2016).
- [81] ecafe.com. *Hole-In-Space, 1980*. URL: <http://www.ecafe.com/getty/HIS/index.html> (besucht am 16.06.2016).
- [82] Elo Touch Solutions. *Compare All Resistive Touch Technologies (4-, 5-, 6-, 7-, and 8-Wire Explained)*. URL: http://www.elotouch.com/Technologies/compare_resist.asp (besucht am 10.06.2016).
- [83] Elo Touch Solutions. *How Surface Capacitive Touch Technology Works*. URL: <http://www.elotouch.com/Technologies/SurfaceCapacitive/howitworks.asp> (besucht am 10.06.2016).
- [84] European Resuscitation Council. *ERC European Resuscitation Council*. URL: <https://www.erc.edu/> (besucht am 20.06.2016).
- [85] Github (speps). *XInputDotNet*. URL: <https://github.com/speps/XInputDotNet> (besucht am 04.06.2016).
- [86] Infoscreen GmbH. *Infoscreen GmbH*. URL: <https://www.infoscreen.de/> (besucht am 16.06.2016).
- [87] Microsoft. *Xbox 360 Wireless Controller für Windows*. URL: <https://www.microsoft.com/hardware/de-de/p/xbox-360-wireless-controller-for-windows> (besucht am 04.06.2016).
- [88] Microsoft Developer. *Kinect With Me Part 1 - Setting Up Your Machine*. URL: <https://blogs.msdn.microsoft.com/cndevs/2014/10/23/kinect-with-me-part-1-setting-up-your-machine/> (besucht am 02.06.2016).
- [89] Microsoft Developer Network. *JointType Enumeration*. URL: <https://msdn.microsoft.com/en-us/library/microsoft.kinect.jointtype.aspx> (besucht am 03.06.2016).
- [90] Microsoft Developer Network. *Kinect for Windows Sensor Components and Specifications*. URL: <https://msdn.microsoft.com/en-us/library/jj131033.aspx> (besucht am 02.06.2016).
- [91] Microsoft Developer Network. *Kinect for Windows v2 Runtime API Reference*. URL: <https://msdn.microsoft.com/en-us/library/dn758675.aspx> (besucht am 02.06.2016).
- [92] Microsoft Developer Network. *Microsoft.Kinect.Face Namespace*. URL: <https://msdn.microsoft.com/en-us/library/microsoft.kinect.face.aspx> (besucht am 10.07.2016).
- [93] Microsoft Developer Network. *WindowsPreview.Kinect Namespace*. URL: <https://msdn.microsoft.com/en-us/library/windowspreview.kinect.aspx> (besucht am 10.07.2016).

- [94] Microsoft Developer Network. *WindowsPreview.Kinect.Input Namespace*. URL: <https://msdn.microsoft.com/en-us/library/windowspreview.kinect.input.aspx> (besucht am 10.07.2016).
- [95] Microsoft Developer resources. *Kinect hardware*. URL: <https://developer.microsoft.com/en-us/windows/kinect/hardware> (besucht am 02.06.2016).
- [96] Microsoft Download Center. *Download Kinect for Windows SDK 1.8 from Official Microsoft Download Center*. URL: <https://www.microsoft.com/en-us/download/details.aspx?id=40278> (besucht am 02.06.2016).
- [97] Microsoft Download Center. *Download Kinect for Windows SDK 2.0 from Official Microsoft Download Center*. URL: <https://www.microsoft.com/en-us/download/details.aspx?id=44561> (besucht am 02.06.2016).
- [98] Microsoft Entwickler Ressourcen. *XInput Game Controller APIs*. URL: [https://msdn.microsoft.com/de-de/library/windows/desktop/hh405053\(v=vs.85\).aspx](https://msdn.microsoft.com/de-de/library/windows/desktop/hh405053(v=vs.85).aspx) (besucht am 04.06.2016).
- [99] Nintendo. *Wii-Fernbedienung Plus*. URL: <https://www.nintendo.de/Wii/Wii-mini/Steuerung/Steuerung-726560.html> (besucht am 02.06.2016).
- [100] playstation.com. *PlayStation® Move*. URL: <https://www.playstation.com/en-us/explore/accessories/playstation-move/> (besucht am 02.06.2016).
- [101] Twitch. *Broadcasting, Chat, and More Coming to Xbox One*. URL: <https://blog.twitch.tv/broadcasting-chat-and-more-coming-to-xbox-one-a65621bbee3a#.5lknf3fj3> (besucht am 02.06.2016).
- [102] ubicomp.oulu.fi. *Center for Ubiquitous Computing*. URL: <http://ubicomp.oulu.fi/studying-communities-through-public-displays/> (besucht am 16.06.2016).