

LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN
Institut für Informatik
Lehrstuhl für Mensch-Maschine-Interaktion
Prof. Dr. Andreas Butz

Masterarbeit

Modelling Multimodal Interaction in the Car

Carina Rothe
carina.saliger@live.de

Bearbeitungszeitraum: 16. 09. 2016 bis 17. 03. 2017
Betreuer: Dr. Bastian Pfleging
Externer Betreuer von BMW: Dipl. Ing. Florian Roider
Verantw. Hochschullehrer: Prof. Dr. Butz

Zusammenfassung

Die Informationssysteme im Auto werden immer komplexer und können teils schon mit multimodalen Eingaben wie Sprache, Touch oder Gestik bedient werden. Diese multimodalen Interaktionen eröffnen den Nutzern neue Möglichkeiten, die passende Modalität je nach Situation und Eignung zu wählen. Jedoch muss darauf geachtet werden, die Ablenkung des Fahrers gering zu halten und die Modalitäten in geeigneten Kombinationen zu unterstützen.

Dadurch motiviert haben wir in dieser Arbeit ein multimodales Modell erstellt, dass Designern multimodaler Schnittstellen helfen soll, die Interaktionszeiten verschiedener Modalitäten vorherzusagen und zu vergleichen. Dazu entwickelten wir einen Prototypen, der mit Touch, Geste und Sprache bedient werden kann. In einem Within-Subject-Design testeten 22 Probanden in vier Anwendungsbeispielen alle Kombinationen von Modalitäten mit einem Moduswechsel, sowie die unimodalen Varianten. Aus diesen Interaktionszeiten leiteten wir ein Modell mit verschiedenen Aktionszeiten für die Modalitäten Touch, Sprache und Geste ab, sowie deren Wechselkosten untereinander. Unser Modell lehnt sich an das Konzept des Keystroke-Level Modells und Erweiterungen an, jedoch konzentrieren wir uns nicht auf einzelne Operatoren, sondern auf etwas gröbere Aktionen und deren Wechselkosten.

Zur Validierung des multimodalen Modells werden in einer weiteren Studie fünf verschiedene Aufgaben mit bis zu vier Moduswechseln untersucht und mit unserer Vorhersage des Modells verglichen. Es zeigte sich, dass unser Modell multimodale Interaktionen mit einem durchschnittlichen Vorhersagefehler von 14,746% vorhersagen konnte. Diese Abschätzung von multimodalen Interaktionszeiten im Auto ermöglicht es Designern bereits in einem frühen Stadium der Entwicklung zu unterstützen und hilft ihnen die Ablenkung des Fahrers zu minimieren.

Abstract

In-Vehicle-Information-Systems getting more and more complex and the possibility for users using multimodal interactions like speech, touch and gestures is growing. A big advantage of multimodal interactions is that a user can decide in which situation he wants to use the best choice of interaction. But such a multimodal system should minimize the distraction of the driver and support the best combinations of modalities.

Motivated by this we designed in this thesis a multimodal model to support designers of multimodal interfaces by predicting the interaction time of such systems. By comparing interaction times the best combinations can be detected and a better design is possible. Our model is similar to the concept of the keystroke-level-model and its extensions, but we focus on actions for each modality and the resulting change costs, not on operator level. With the measured interaction times of our first study we developed a model with different times for each modality (touch, speech and gestures) and we also modeled the changing costs between one modality to another.

To validate our multimodal model we investigate in a second study five tasks and compared the total task time with our prediction of the model. We showed that the prediction of our model is a good match compared to the observed times. The average RMSE was 14,746%. Such predictions can support designer in an early stage of implementation and help them to reduce driver distraction.

Aufgabenstellung

In modernen Autos wird die Funktionalität von Informationssystemen im Auto immer komplexer. Es ist wichtig dem Nutzer eine einfache Eingabe zu ermöglichen, ohne ihn zu sehr visuell oder mental zu belasten. Durch die steigende Anzahl an eingebauter Technik in Autos wachsen auch die Eingabemöglichkeiten des Fahrers. Das Infotainmentsystem (IVIS) kann nicht nur durch haptische Knöpfe und Schalter oder Touch bedient werden, sondern auch per Sprache, Gesten und Blick. Die meisten Eingabeoptionen sind schon in aktuelle Autos integriert, jedoch ist es schwer für Entwickler der Schnittstellen, die Benutzbarkeit aller Eingabeoptionen sowie deren Kombination einzuschätzen, zum Beispiel bzgl. der Aufgabendauer.

Im Zuge dieser Masterarbeit soll daher die Bedienung im Fahrzeug beobachtet werden mit dem Ziel, die multimodale Interaktion für künftige Systeme modellieren zu können. Ein Beispiel dazu ist eine Erweiterung des Keystroke-Level Modells auf Bedienaufgaben eines multimodalen IVIS mit verschiedenen Inputmodalitäten, um damit Bedienzeiten vorhersagen zu können. Primäre Aufgabe wird dabei die Analyse und Berechnung von Interaktionszeiten. Dies erfordert einerseits eine aktuelle Bestandsaufnahme bzgl. existierender Modelle wie zum Beispiel erweiterten KLMs und deren Operatoren. Neue Bedienmöglichkeiten erfordern zudem die Ermittlung neuer Operatoren, um die Interaktion mit heutigen IVIS (zum Beispiel per Touch, Sprache und Geste) darstellen zu können. Ein besonderer Aspekt ist dabei die multimodale Bedienung, bei der zum Beispiel Handlungen nacheinander oder alternativ über verschiedene Modalitäten abgewickelt werden, die dementsprechend unterschiedlich lang dauern können (auch: Wechselzeiten / -kosten).

Basierend auf den empirisch zu ermittelnden Operatoren und Operatorzeiten soll dann ein erweitertes Modell erstellt werden, das die Vorhersage für eine multimodale Interaktion im Fahrzeug ermöglicht. Ein solches Modell kann dazu dienen, verschiedene Interaktionsmöglichkeiten zu vergleichen und erlaubt insbesondere in frühen Designphasen eine schnelle Abschätzung bzgl. der Interaktionsdauer im Fahrzeug.

Zu diesem Zweck ist es sowohl unter den Aspekten der Bedienbarkeit, der User Experience also auch der Fahrsicherheit / Fahrerablenkung geplant, verschiedene Untersuchungen (zum Beispiel Umfragen, Fokusgruppen, Laborstudien) durchzuführen, um Operatoren und Interaktionszeiten zu messen und zu validieren. Im Speziellen sind die folgenden Schritte angedacht:

1. Recherche: Finden verwandter Arbeiten und Analyse des aktuellen Stands von IVIS & KLM
2. Evtl. Nutzerbefragung/Brainstorming-Runde, um herauszufinden, welche Gesten und Sprachbefehle auf die Anwendungsbeispiele von Touch am besten geeignet sind.
3. Planung und Durchführung einer Studie zur Bestimmung von Interaktionszeiten multimodaler Interaktion im Fahrzeug.

4. Ableitung eines geeigneten Modells.
5. Validierung des Modells in einer weiteren Studie.
6. Ggf. Gestaltung eines Tools zur vereinfachten Berechnung von Interaktionen für neue Prototypen

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, 12. März 2017

.....

Inhaltsverzeichnis

1 Einleitung	11
2 Verwandte Arbeiten	15
2.1 Interaktionsraum beim Fahren	15
2.1.1 Design von Informationssystemen im Auto	16
2.1.2 Evaluation von Informationssystemen im Auto	16
2.2 Modelle zur Vorhersage von Interaktionszeiten	19
2.2.1 Fitts' Law	19
2.2.2 GOMS	20
2.2.3 Keystroke-Level Modell (KLM)	20
2.2.4 Erweiterung des KLM für tragbare Geräte	21
2.2.5 Erweiterung des KLM für Toucheingaben	23
2.2.6 Erweiterung des KLM im automobilen Kontext	23
2.3 Multimodale Interaktion	25
2.4 Multimodale Interaktion im automobilen Kontext	26
3 Workshop Connected Minds	29
3.1 Organisation und Ablauf des Workshops	29
3.2 Ergebnisse des Workshops Connected Minds	30
3.3 Abgeleitete Idee für ein multimodales Modell	33
4 Multimodaler Prototyp	35
4.1 Beschreibung und Umsetzung der Anwendungsbeispiele	35
4.2 Implementation des multimodalen Prototypen	38
4.2.1 Realisierung der Toucheingabe	39
4.2.2 Realisierung der Gesteneingabe	39
4.2.3 Realisierung der Spracheingabe	40
4.2.4 Protokollierung der relevanten Events	41
5 Studie und Auswertung	43
5.1 Studiendesign zur Erhebung multimodaler Interaktionszeiten	43
5.1.1 Permutation der Anwendungsbeispiele	43
5.1.2 Fragebögen der Studie	44
5.2 Durchführung der Studie	45
5.3 Quantitative Auswertung der Studienergebnisse	48
5.3.1 Studienteilnehmer	48
5.3.2 User Experience	48
5.3.3 Ermittlung aller Zeiten der Aktionen	50

5.3.4	Statistische Tests der Aktionszeiten	55
5.4	Ableitung des multimodalen Modells aus den gewonnenen Interaktionszeiten	56
5.5	Multimodales Modell mit Wechselkosten	61
5.6	Qualitative Auswertung der Studie	63
5.7	Diskussion zur Studie und Erstellung des Modells	64
6	Validierung	69
6.1	Studiendesign zur Validierung	69
6.1.1	Übungsbeispiele der Validierungsstudie	70
6.1.2	Anwendungsbeispiele der Validierungsstudie	70
6.2	Durchführung der Studie	74
6.3	Auswertung der Studienergebnissen	74
6.3.1	Teilnehmer	75
6.3.2	User Experience	75
6.3.3	Ergebnisse des Modells	77
6.4	Diskussion	82
6.4.1	Multimodales Modell	82
6.4.2	Qualitative Erkenntnisse	84
7	Zusammenfassung und Ausblick	85
8	Danksagung	87
9	Verzeichnisse	89
	Literaturverzeichnis	89
	Abbildungsverzeichnis	93
	Tabellenverzeichnis	97
10	Anhang	101

Kapitel 1

Einleitung

Schon längst besteht Autofahren aus mehr als lediglich dem Bedürfnis, von A nach B zu kommen. Ein Fahrer will im Auto vielmehr auch die Möglichkeit haben Dinge zu nutzen, welche nichts mit der Hauptaufgabe des Fahrens (Primäre Aufgabe) zu tun haben. Zum Beispiel navigiert sich der Autofahrer zu seinem Ziel, hört Musik, findet die nächste Tankstelle in der Umgebung, nimmt ein Telefongespräch an, sowie vieles mehr. Somit werden in aktuellen Autos die Funktionalität von Informationssystemen im Auto (IVIS von „In-Vehicle-Information System“) komplexer und die Inhalte umfangreicher ([Kern and Schmidt, 2009](#)). Hierbei besteht die Aufgabe und Herausforderung für Designer diese Masse an Information in geeigneter Weise für den Fahrer darzustellen und nutzbar zu machen. Dauert eine Interaktion mit einem IVIS zu lange oder ist sie schwer zu verstehen, sind dies häufig Gründe für zu große Ablenkungen beim Fahren, was Autounfälle mit sich zieht ([Neale et al., 2005](#)). Daher sollten diese potenziellen Ablenkungen so minimal wie möglich gehalten werden. Zusätzlich ist es wichtig, dem Nutzer eine einfache und intuitive Eingabe des Informationssystems im Auto zu ermöglichen, ohne ihn gleichzeitig zu sehr visuell oder mental zu belasten.

Hinzu kommt, dass die Bedienung von jeglichen Funktionen im Auto längst nicht mehr unimodal sind. Unimodal bezeichnet hierbei die Interaktion zwischen Mensch und Auto, welche lediglich eine Modalität verwendet, wie zum Beispiel haptische Bedienelemente. Dagegen ist es dem Fahrer möglich verschiedenste Funktionen multimodal auszuführen. Zu den üblichen haptischen Bedienelementen kommen somit zusätzlich Touchdisplays oder Möglichkeiten der Sprachbedienung, die in Kombination betätigt werden können.

Eine weit verbreitete Modalität ist die Spracheingabe, die zum Beispiel oft genutzt wird, um ein Telefonanruf zu starten oder anzunehmen. Dazu wird meist ein haptischer „Push to Talk“ Button gedrückt und anschließend der Sprachbefehl gesagt.

Ein wesentlicher Vorteil der Sprachsteuerung ist, dass die Hand nicht vom Lenkrad genommen werden muss und auch der Blick auf der Straße verbleiben kann. [Maciej and Vollrath \(2009\)](#) sind der Überzeugung, dass Sprachbedienung eine Unerlässlichkeit in zukünftigen Autos sein wird. Allerdings ist bei einer Autofahrt mit mehreren Mitfahrern, bei der geredet und Musik gehört wird, die Sprachsteuerung nicht immer optimal.

Auch Gestensteuerungen sind in einigen Autos bereits vertreten. Zum Beispiel wird im BMW 7er eine Kreisbewegung mit dem Finger nach rechts beziehungsweise links erkannt, um die Laut-

stärke zu erhöhen beziehungsweise zu verringern. Es ist allerdings meist nicht möglich mit einer dieser neuen Modalitäten alle vorhandenen Funktionen auszuführen. Ein Interface heißt seriell-redundant, wenn Modalitäten wie zum Beispiel Sprache, Gestik oder Touch beliebig gewechselt werden können (Neuss, 2001).

Mit einem seriell-redundanten multimodalen Informationssystemen im Auto könnte der Fahrer selbst entscheiden in welchen Situationen welcher Modus für den jeweiligen Schritt am einfachsten und sichersten ist. Dies stellt nach Muller and Weinberg (2011) einen wesentlichen Vorteil im Auto dar. Natürlich gibt es auch Kombinationen, welche mit bestimmten Modalitäten weniger sinnvoll sind als andere. Unsere Motivation ist deshalb, solche multimodalen Interaktionen in IVIS zukünftig besser zu verstehen, um diese optimal umsetzen zu können. Es ist wichtig die schnellsten und einfachsten Kombinationen verschiedener Modalitäten zu kennen, um dem Fahrer auch die besten Möglichkeiten in verständlicher Weise anzuzeigen. Damit soll die visuelle und mentale Beanspruchung für den Fahrer so gering wie möglich gehalten werden.

Um Interfaces bereits in einem frühen Stadium der Entwicklung zu testen wurde in der Vergangenheit häufig das Keystroke-Level Modell in der ursprünglichen oder den erweiterten Varianten verwendet. Damit können bereits vor der Implementierung eines Prototyps Interfacekonzepte auf ihre Bediendauer getestet werden, indem die Dauer einer gewählten Aufgabe vorhergesagt wird. Somit wird es möglich, auf einfache Art verschiedene Varianten zu vergleichen. Das Keystroke-Level Modell bezieht sich in der ursprünglichen Variante nur auf Desktop-basierte Anwendungen, die sich hauptsächlich auf Textverarbeitungsprogramme bezogen. In den letzten Jahrzehnten wurde dieses vereinfachte Konzept jedoch auch auf Geräte und Anwendungen übertragen und erwies sich hierbei als geeignete Methode zur Einschätzung der Interaktionsdauer.

Das Ziel dieser Arbeit ist es, dieses Konzept für seriell redundante multimodale IVIS zu adaptieren und anzupassen. Dazu entwickeln wir ein Modell, mit dem Interaktionszeiten eines multimodalen IVIS vorhergesagt und verglichen werden können. Wir konzentrieren uns auf die Modalitäten Touch, Sprache und Geste und auf die entstehenden Kosten bei einem Wechsel zwischen zwei Modalitäten. Jede Modalität besitzt ihre Vorteile und Nachteile und kann in geeigneter Kombination den Fahrer optimal unterstützen.

Im Zuge dieser Masterarbeit soll daher die Bedienung im Fahrzeug beobachtet werden, mit dem Ziel die multimodale Interaktion für künftige Systeme modellieren zu können. Ein besonderer Aspekt liegt dabei auf der multimodalen Bedienung, bei der zum Beispiel Handlungen nacheinander oder alternativ über verschiedene Modalitäten abgewickelt werden. Die Aktionen können dementsprechend unterschiedlich lang dauern und Wechselkosten enthalten. Basierend auf den empirisch zu ermittelnden Operatoren und Operatorzeiten soll dann ein erweitertes Modell erstellt werden, welches die Vorhersage für eine multimodale Interaktion im Fahrzeug ermöglicht. Zu diesem Zweck ist es sowohl unter den Aspekten der Bedienbarkeit und User Experience, als auch der Fahrsicherheit und Fahrerablenkung geplant, verschiedene Untersuchungen durchzuführen. Hierbei sollen Operatoren und Interaktionszeiten gemessen und validiert werden.

Gliederung

Im Zuge dieser Arbeit wurde folgendermaßen vorgegangen. Zu Beginn werden in **Kapitel 2** die Grundlagen und Richtlinien zu Informationssystemen im Auto, deren Design und Evaluationsmöglichkeiten erläutert. Anschließend wird das ursprüngliche Keystroke-Level Modell und deren

KAPITEL 1. EINLEITUNG

Erweiterungen auf tragbare Geräte, für Touch und dem KLM im automobilen Kontext erläutert. Außerdem befassen wir uns mit multimodalen Interaktionen im generellen, sowie mit multimodalen Interaktionen im automobilen Kontext.

In **Kapitel 3** erhalten wir in einer Brainstorming Runde einen Überblick über multimodale Interaktionen im Auto und deren bereits vorhandenen und möglichen Umsetzungsvarianten. Dort sammeln wir gemeinsam Ideen und gruppieren sie Schritt für Schritt, bis wir auf deren Grundlage unser Konzept aufbauen können. Aus den gesammelten Ideen und Ergebnissen leiten wir unsere Aktionen her und konstruieren geeignete Anwendungsbeispiele.

Kapitel 4 beschreibt die Idee und Implementierung des multimodalen Prototyps, der per Touch, Geste und Sprache, sowie je aus einer Kombination aus zwei Modalitäten bedient werden soll. Wir gehen auf die Anwendungsbeispiele und die Umsetzungen ein.

Das Studiendesign, die Durchführung der Studie und die anschließende Auswertung wird in **Kapitel 5** erläutert. Die 22 Probanden führen vier verschiedene Anwendungsbeispiele mit allen Moduskombinationen durch. Dies resultiert in 33 verschiedene Kombinationen, welche von jedem Probanden durchgeführt werden, um für alle Aktionen unseres Modells genügend Daten für jede Moduskombination und deren Wechselkosten zu erhalten. Aus den Studienergebnissen werden die Zeiten der Aktionen erhoben und ein Modell mit Durchschnittszeiten erstellt. Des weiteren untersuchen wir, welchen Einfluss ein Moduswechsel auf die Interaktionszeiten hat und in welchem Maße. Anschließend werden die Ergebnisse dieser Studie diskutiert und mit anderen Arbeiten verglichen.

Um die ermittelten Aktionszeiten des erstellten Modells zur Vorhersage/Abschätzung von Interaktionszeiten zu validieren, wird in **Kapitel 6** der Prototyp angepasst und in einer zweiten Studie an zehn weiteren Probanden getestet. Dazu werden die Daten ausgewertet und mit unserem erstellen Modell verglichen. Anschließend diskutieren wir über die Erkenntnisse.

Kapitel 7 fasst die Erkenntnisse beider Studien und der gesamten Arbeit zusammen und gibt einen Ausblick auf weitere Forschungsideen.

Kapitel 2

Verwandte Arbeiten

Im folgenden Kapitel gehen wir auf verschiedene Hintergründe und verwandten Arbeiten ein. Zur besseren Einordnung beginnen wir mit einer Einführung des Interaktionsraumes im Fahrzeug und geben einen kurzen Überblick zu den Bereichen Design und Evaluation von Informationssystemen im Auto.

Anschließend werden drei Modelle zu Vorhersage von Interaktionszeiten vorgestellt. Dabei gehen wir besonders auf das Konzept des Keystroke-Level Modell und Erweiterungen ein und stellen die verwandten Arbeite in diesem Bereich vor.

Zuletzt befassen wir uns mit multimodalen Interaktionen, sowie mit multimodaler Interaktion im automobilen Kontext.

2.1 Interaktionsraum beim Fahren

In heutigen Fahrzeugen wird dem Fahrer eine Vielzahl von Informationen und Entertainment Funktionen angeboten. Diese stellen einen multifunktionalen Raum für den Fahrer dar ([Kern and Schmidt, 2009](#)). Ausgestattet mit Medienfunktionen, Navigationssystemen, die mittels GPS ortszogene Funktionen bieten (zum Beispiel „finde die nächste Raststätte“), Kommunikationssystemen und dem angeschlossenen Smartphone, bieten Fahrzeuge auch einen vernetzten Interaktionsbereich.

Beim Fahren werden drei verschiedene Aufgabenbereiche unterschieden ([Geiser, 1985](#)). Die wichtigste Aufgabe des Fahrers ist das Fahren an sich, also die sichere Fortbewegung zum Ziel. Diese Aufgabe wird als primäre Aufgabe (Primary Task) kategorisiert und beinhaltet das Fahren, Bremsen, Spur halten und Lenken. Sekundäre Aufgaben (Secondary Tasks) bilden eine weitere Kategorie. Sie unterstützen mit Aktionen und Reaktionen die primäre Aufgabe und die Sicherheit des Fahrens. Hierzu gehören zum Beispiel das Blinken oder das Einschalten des Lichts bei Dunkelheit. Die letzte Kategorie sind die tertiären Aufgaben (Tertiary Tasks). Sie beinhalten alle Aufgaben, die unabhängig vom Fahren gemacht werden können, wie zum Beispiel die Bedienung von Komfortfunktionen, Kommunikationsfunktionen oder Entertainmentfunktionen, aber auch Interaktionen mit dem Beifahrer ([Geiser, 1985](#)), ([Kern and Schmidt, 2009](#)). Diese dritte Kategorie ist ein wesentlicher Bestandteil der Ablenkungen des Fahrers und führt oft zu Fahrfehlern oder



Abbildung 2.1.1: Aufteilung der Bereiche von primären, sekundären und tertiären Aufgaben. Bild entnommen aus Tonnis et al. (2006)

sogar Unfällen (Neale et al., 2005), (Rumar, 1988).

2.1.1 Design von Informationssystemen im Auto

Beim Design von Informationssystemen im Auto (IVIS) ist es wichtig im Kopf zu behalten, wer unsere Nutzer sind. Das Alter von Autofahrern kann sich von 16 bis über 90 Jahren erstreckt. In der Gruppe der 21-75 jährigen haben über 80% einen Führerschein (Green, 2003). Unsere Nutzergruppe spiegelt also einer sehr große Altersspanne wieder. Es ist zu erwarten, dass die verschiedenen Altersklassen mit unterschiedlicher Performanz Aufgaben verrichten (Green, 2003). Um Informationssysteme möglichst für jeden leicht verständlich zu gestalten, gibt es bereits viele Richtlinien, Prinzipien und Standards. Beim Design von Informationssystemen im Auto beachtet sollten diese beachtet werden (siehe unter anderem Alliance of Automobile Manufacturers (AAM) (Group et al., 2006), National Highway Traffic Safety Administration (NHTSA) (Administration et al., 2012), European Statement of Principles (ESoP) (of the European Communities, 2008), ISO, Society of Automotive Engineers (SAE) oder DIN). Green (2012) hat dazu auch einige der Standards verglichen und zusammengefasst.

2.1.2 Evaluation von Informationssystemen im Auto

Informationssysteme im Auto können das Fahrverhalten negativ durch Ablenkung beeinflussen. Deshalb ist es wichtig solche Systeme zu evaluieren, um den Effekt der Ablenkung, die mentalen Belastung und die Interaktionsdauer zu messen. Es gibt verschiedene Methoden um ein neues System oder einen neuen Prototypen zu testen. Burnett (2008) zeigt in seiner Abbildung 2.1.2 die verschiedenen Evaluationstypen von einer realen Feldstudie auf der Straße bis hin zu Laborstudien.

In realen Feldstudien zur Untersuchung des Fahrverhaltens werden die Fahrzeuge zur Erhebung gewünschter Werte instrumentiert und über einen Zeitraum beobachtet. Mit dieser Methode können unter realistischen Bedingungen Daten erhoben werden. Allerdings können Variablen wie das Wetter oder die Verkehrslage nicht beeinflusst werden. In einer High Fidelity Fahrsimulation können äußere Einflüsse wie die Verkehrslage kontrolliert werden, jedoch nimmt hier bereits das natürliche Fahrverhalten ab, da sich der Fahrer offensichtlich in einer Studie befindet und sich

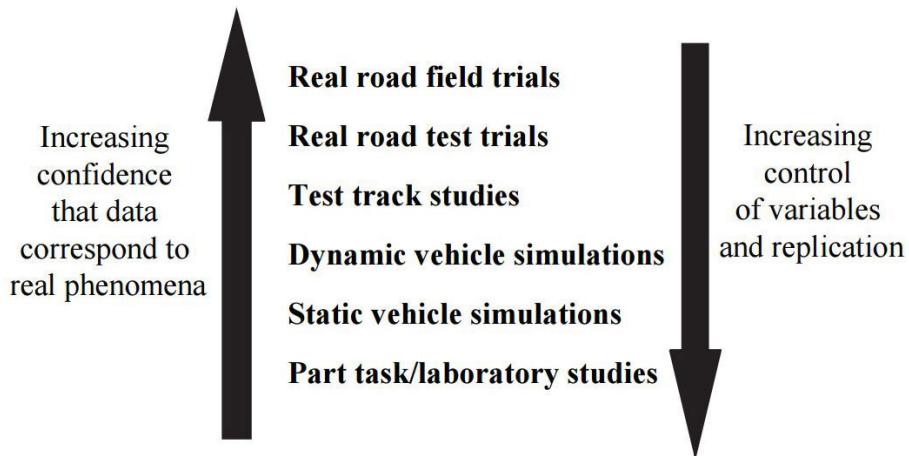


Abbildung 2.1.2: Evaluierungstypen und ihre Abhängigkeit zur Validität und deren Kontrollierbarkeit. Bei realen Feldstudien verhalten sich Fahrer sehr natürlich, jedoch ist es schwer die Beeinflussung von Variablen zu kontrollieren. Im Gegensatz dazu können in Laborstudien die Variablen genau kontrolliert werden, jedoch ist das Fahrverhalten konstruiert und entspricht nicht mehr dem natürlichen Verhalten. Idee und Bild aus [Burnett \(2008\)](#)

möglicherweise anders verhält. Außerdem sind die Kosten und der Aufwand einer solchen Studie sehr hoch. Eine Low Fidelity Fahrsimulation ist einfacher umzusetzen, jedoch nimmt das realistische Fahrverhalten weiter ab. Die Validität der Daten ist bei realen Feldstudien am besten und nimmt bis zu den Laborstudien immer mehr ab. Umgekehrt ist die Kontrolle von Variablen und deren Wiederholbarkeit bei Laborstudien am besten und bei realen Feldstudien am unkontrollierbarsten.

Bei der Wahl einer Evaluierungsmethode sollte die Abwägung zwischen Validität der Daten und die Kontrollierbarkeit der Variablen bewusst getroffen werden.

Wir stellen im Folgenden fünf Varianten vor, mit denen IVIS bereits evaluiert wurden. Die 15-Sekunden Regel, der NASA Task Load Index (NASA TLX), der Driving Activity Load Index (DALI), die Okklusions Methode und der Lane-Change Test. Diese Varianten können in einem stehenden Fahrzeug getestet werden.

15-Sekunden Regel

Die Länge einer sekundären oder tertiären Aufgabe steht in Korrelation mit dem Unfallrisiko. Da es einfacher ist die Gesamtdauer einer Aufgabe (Total Task Time) zu bestimmen, ist die 15 Sekundenregel eine vereinfachte Annahme, dass eine Aufgabe in einem stehenden Auto nicht länger als 15 Sekunden dauern soll ([Green, 1999b](#)). Diese Regel ist die Grundlage eines vom Society of Automotive Engineers (SAE) vorgeschlagenen Standards ([Green, 1999a](#)). [Green \(1999a\)](#) definiert die Dauer in einem stehenden Auto oder einer Attrappe gemessen, in dem der Proband nur die gewünschte Aufgabe ausführt. TODO(Satz kapier ich nich) Das entspricht der Total Task Time, die wir in unserer Studie ebenfalls verwenden werden.

NASA TLX

Hart and Staveland (1988) entwickelte den NASA Task Load Index, dessen Verfahren die mentale Belastung in 6 Dimensionen misst. Dafür verwenden wir die deutsche Übersetzung des NASA TLX. Die Beanspruchungen werden unterschieden in:

1. **Geistige Anforderungen:** Wie viel geistige Anstrengung war bei der Informationsaufnahme und -verarbeitung erforderlich (zum Beispiel Denken, Entscheiden, Rechnen, Erinnern, Hinsehen, Suchen)? War die Aufgabe leicht oder anspruchsvoll, einfach oder komplex, erforderte sie hohe Genauigkeit oder war sie fehlertolerant?
2. **Körperliche Anforderungen:** Wie viel körperliche Aktivität war erforderlich (zum Beispiel Ziehen, Drücken, Drehen, Steuern, Aktivieren)? War die Aufgabe leicht oder schwer, einfach oder anstrengend, erholsam oder mühselig?
3. **Zeitliche Anforderungen:** Wie viel Zeitdruck wurde hinsichtlich der Häufigkeit oder dem Takt, mit dem Aufgaben oder Aufgabenelemente auftraten, empfunden? War die Abfolge langsam und geruhsam oder schnell und hektisch?
4. **Leistung:** Wie erfolgreich wurde die vom Versuchsleiter (oder dem Probanden selbst) gesetzten Ziele erreicht? Wie zufrieden war der Proband mit seiner Leistung bei der Verfolgung dieser Ziele?
5. **Anstrengung:** Wie hart musste Der Proband arbeiten, um den Grad an Aufgabenerfüllung zu erreichen?
6. **Frustration:** Wie unsicher, entmutigt, irritiert, gestresst und verärgert (im Gegensatz zu sicher, bestätigt, entspannt und zufrieden mit sich selbst) fühlte sich der Proband während der Aufgabe?

Für jede dieser Beanspruchungen wird ein Wert zwischen 1 (gering) und 20 (hoch) abgefragt. Anschließend werden im zweiten Teil noch die 6 Beanspruchungen verglichen. Es werden alle Beanspruchungen gegenüber gestellt und der Proband muss sich immer für die Beanspruchung entscheiden, die aus seiner Sicht wichtiger ist. Somit wird eine Gewichtung ermittelt, um den Grad der Belastung genauer bestimmen zu können. Diese Bestimmung der Beanspruchung mit dem NASA TLX ist weitverbreitet und wird regelmäßig in der Forschung verwendet (Hart, 2006). Teilweise wird die kurzen Version (ohne die Gewichtung durch den zweiten Teil) verwendet.

DALI

Sehr ähnlich zu dem NASA TLX ist der DALI, der sich jedoch auf den automobilen Kontext bezieht und den Beanspruchungswert mit angepasster Gewichtung berechnet. Der DALI evaluiert die subjektive mentale Belastung eines Fahrers während der Fahrt (Pauzié, 2008), mit oder ohne die Unterstützung eines Informationssystems. Eine weitere Technik zur subjektiven Einschätzung der Belastung ist die „Subjective Workload Assessment Technique“ (SWAT), siehe (Reid et al., 1982).

Okklusions Methode

Hierbei wird die visuelle Belastung von IVIS für die sekundären Aufgaben gemessen. Diese Verschluss Methode soll die Abwendung des Blickes simulieren, die zwischen der Hauptaufgabe dem Fahren und dem Blick zum IVIS passiert. Die Hauptidee besteht darin die Sicht, durch Abdeckung, abwechselnd für 1,5 Sekunden zu verhindern und anschließend für 1,5 Sekunden zu ermöglichen. Heutzutage gibt es spezielle Brillen mit denen diese Methode durchgeführt werden kann (Pettitt et al., 2006).

Lane Change Test (LCT)

Dieser Test wurde in Kooperation von DaimlerChrysler und BMW von Mattes (2003) entwickelt. Er misst die Performanz von Doppelaufgaben. Der Proband muss eine simulierte Fahraufgabe lösen. Diese ist in abstrakter Weise dargestellt und es muss auf einer dreispurigen Autobahn zwischen den Spuren gewechselt werden. Wann gewechselt werden muss wird mit Schildern angezeigt und es werden hieraus verschiedenen Zeiten gemessen (wie lange dauert ein Wechsel, wurde jedes mal richtig gewechselt). Diese Fahraufgabe löst der Proband zuerst ohne sekundäre Aufgabe. Dies wird als Referenzwert verwendet. Anschließend soll der Proband sowohl die Fahraufgabe bewerkstelligen, als auch das zu Testende IVIS. Natürlich hat dabei die Fahraufgabe die höhere Priorität. Am Ende können die Zeiten von reiner Fahraufgabe und Fahraufgabe mit IVIS verglichen werden. Diese Methode kommt dem natürlichen Fahrverhalten deutlich näher, als der 15-Sekunden Regel und der Okklusion Methode.

2.2 Modelle zur Vorhersage von Interaktionszeiten

Zur Vorhersage von Bedien- oder Interaktionszeiten gibt es verschiedene Ansätze und Modelle. In den nächsten Unterkapiteln stellen wir drei Varianten vor. Dabei gehen wir vor allem auf das Keystroke-Level Modell und dessen Erweiterungen ein.

2.2.1 Fitts' Law

Der Psychologe Paul Fitts entwickelte 1954 ein Gesetz, dass die Zeit berechnet, die eine eindimensionale Bewegung einer Strecke zu einem Ziel einer bestimmten Größe benötigt (Fitts, 1954).

Dieses Gesetz bekam im Laufe der Zeit den Namen Fitts' Law und wurde in der Mensch-Maschine Interaktion oft verwendet. Die Formel hat ihren Ursprung bei MacKenzie (MacKenzie, 1992) und berechnet sich wie folgt:

$$MT = a + b * \log_2\left(\frac{D}{W} + 1\right)$$

Wobei MT der Bewegungszeit (movement time) entspricht, a und b Konstanten sind, D die Distanz des Startpunkts bis zur Mitte des Zielobjekts ist und W die Breite des Objekts darstellt, die entlang der Bewegungsrichtung gemessen wird. Aus Fitts' Law können wir entnehmen, dass Distanz und Bewegungszeit positiv korreliert sind, das heißt mit größer werdender Distanz sinkt die

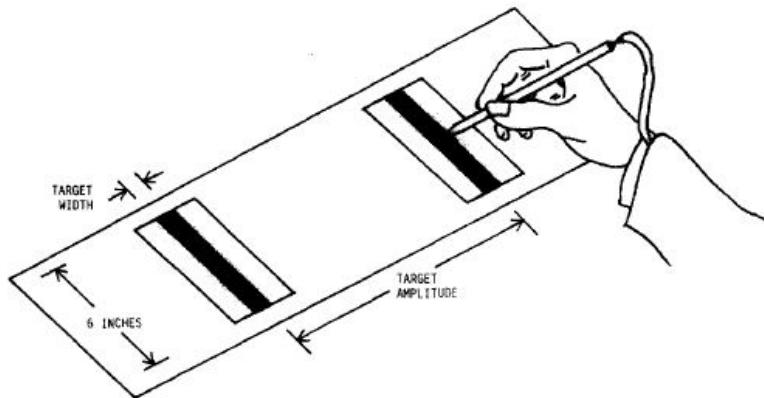


Abbildung 2.2.1: Orginaler Aufbau des Versuchs zur Erstellung von Fitts' Law. Die schraffierten Bereiche sollten abwechselnd mit dem Stift getroffen werden, ohne den Bereich zu überschreiten (Fitts, 1954).

Zeit verlängert. Zeit und Größe des Zielobjekt sind dagegen negativ korreliert, das heißt mit größer werdendem Zielobject verkürzt sich die Zeit. Bei Designentscheidungen kann mit Fitts' Law die schnellere Variante berechnet werden. Zum Beispiel bei zwei Konzepten für eine Interface mit unterschiedlich großen Buttons, die unterschiedlich positioniert sind.

2.2.2 GOMS

Ein weiteres Modell zur Vorhersage von Interaktionszeiten auf einem Bildschirm mit Eingabe durch Tastatur und Maus ist GOMS (Goal, Operator, Method und Selection Rules). Bei GOMS wird das Ziel durch ein Set von Methoden und dessen Operatoren erreicht und durch die Selektionsregeln werden die entsprechenden Methoden gewählt (Card et al., 1983). Interaktionsschritte können somit von oben nach unten (top-down) strukturiert und dargestellt werden (Butz and Krüger, 2014).

2.2.3 Keystroke-Level Modell (KLM)

Das Keystroke-Level Modell ist eine vereinfachte Variante von GOMS und baut sich umgekehrt von unten nach oben (bottom-up) auf. Das ursprüngliche Keystroke-Level Modell (KLM) wurde von Card et al. (1980) für Desktopanwendungen mit Tastatur entwickelt. Mit dem Keystroke-Level Modell kann die Zeit vorhergesagt werden, die ein Experte benötigt, um eine bestimmte Aufgabe zu lösen. Um diese Zeit vorhersagen zu können wird die Aufgabe in ihre Einzelschritte den Operatoren unterteilt und deren Zeiten aufsummiert. Diese resultierende Zeit entspricht der „Total Task Time“, also die Zeit, die benötigt wird bis ein Nutzer die gesamte Aufgabe vollendet hat. Das Set von Operatoren besteht aus einem mentalen Operator **M** und vier physischen Operatoren. Der mentale Operator entspricht einer Vorbereitungszeit, die ein Nutzer vor oder zwischen Operatoren benötigt. Um das Keystroke-Level-Model richtig anzuwenden werden noch Regeln vorgegeben, um den mentalen Operator **M** richtig zu platzieren. Es gibt 5 verschiedene Regeln mit deren Hilfe die Berechnung einer Aufgabe alle benötigten physischen und mentalen Operatoren vorgibt. Die 4 physischen Operatoren sind:

- **K (Keystroke):** Tastenklick

Begin with a method encoding that includes all physical operations and response operations. Use Rule 0 to place candidate Ms, and then cycle through Rules 1 to 4 for each M to see whether it should be deleted.

- Rule 0.** Insert Ms in front of all Ks that are not part of argument strings proper (e.g., text strings or numbers). Place Ms in front of all Ps that select commands (not arguments).
- Rule 1.** If an operator following an M is *fully anticipated* in the operator just previous to M, then delete the M (e.g., PMK → PK).
- Rule 2.** If a string of MKs *belong to a cognitive unit* (e.g., the name of a command), then delete all Ms but the first.
- Rule 3.** If a K is a *redundant terminator* (e.g., the terminator of a command immediately following the terminator of its argument), then delete the M in front of the K.
- Rule 4.** If a K *terminates a constant string* (e.g., a command name), then delete the M in front of the K; but if the K terminates a variable string (e.g., an argument string), then keep the M.

Abbildung 2.2.2: Plazierungsregeln des mentalen Operators

- **P (Pointing):** Maus zu einem Zielort verschieben
- **D (Drawing):** ein Set von geraden Linien mit der Maus malen
- **H (Homing):** der Wechsel zwischen Maus und Tastatur

Außerdem gibt es noch den Operator R(t), der die Antwortdauer des Systems darstellt.

Es wird grundsätzlich bei den Zeiten von Experten ausgegangen, die keine Fehler machen, jedoch werden beim Tippen von Texten (Operator K) 7 Kategorien unterschieden. Diese Zeiten haben eine Spanne vom besten Tipper bis zum schlechtesten Tipper, siehe Abbildung 2.2.3 auf der nächsten Seite.

Dieses Modell ermöglicht in einem sehr frühen Stadium der Entwicklungsphase, bereits vor der Implementation, vorherzusagen wie lange Experten für bestimmte Aufgaben benötigen. Dieses Modell wurde bereits in vielen Studien angewendet und immer wieder auf Weiterentwicklungen angepasst. Um bei beliebigen Interfaces schnell und ohne Fehler das KLM zu erzeugen entwickelten John et al. (2004) ein Tool (später auch CogTool genannt), das von neuen Interfaces das KLM automatisch generiert und somit das Testen von neuen Interfaces noch einfacher macht.

Mit dem Fortschritt der Technik kommen ständig neue Anwendungen mit neuen Modalitäten auf den Markt, die implementiert und getestet werden müssen. Deshalb wurde im Laufe der Zeit das KLM immer weiter an aktuelle Geräte und Interfacekonzepte angepasst. In den folgenden Abschnitten geben wir eine Zusammenfassung der Erweiterungen des Keystroke-Level Modell.

2.2.4 Erweiterung des KLM für tragbare Geräte

Luo and John (2005) haben das Modell für tragbare Geräte mit Stiftnutzung statt der Tastatur erweitert beziehungsweise auf diesen Kontext angepasst. Die Modelle wurden mit Hilfe der für

Operator	Description and Remarks	Time (sec)
K	Keystroke or button press. Pressing the SHIFT or CONTROL key counts as a separate K operation. Time varies with the typing skill of the user; the following shows the range of typical values:	
	Best typist (135 wpm) .08 ^a	
	Good typist (90 wpm) .12 ^a	
	Average skilled typist (55 wpm) .20 ^a	
	Average non-secretary typist (40 wpm) .28 ^b	
	Typing random letters .50 ^a	
	Typing complex codes .75 ^a	
	Worst typist (unfamiliar with keyboard) 1.20 ^a	
P	Pointing to a target on a display with a mouse. The time to point varies with distance and target size according to Pitts's Law. The time ranges from .8 to 1.5 sec., with 1.1 being an average time. This operator does not include the button press that often follows (.2 sec).	1.10 ^c
H	Homing the hand(s) on the keyboard or other device.	.40 ^d
D(n_D, l_D)	Drawing (manually) n_D straight-line segments having a total length of l_D cm. This is a very restricted operator; it assumes that drawing is done with the mouse on a system that constrains all lines to fall on a square .56 cm grid. Users vary in their drawing skill; the time given is an average value.	.9 $n_D + .18l_D$ ^e
M	Mentally preparing for executing physical actions.	1.35 ^f
R(t)	Response of t sec by the system. This takes different times for different commands in the system. These times must be input to the model. The response time counts only if it causes the user to wait.	t

Abbildung 2.2.3: Orginale Tabelle der Operatoren von Card et al. (1980)

KLM entwickelten Software CogTool generiert (John et al., 2004). Die vorhergesagten und die gemessenen Zeiten betragen nur einen Vorheragefehler von weniger als 8%. (Holleis et al., 2007) und (Li et al., 2010) ergänzen das KLM für die Handynutzung. Li et al. geht dabei mehr auf stiftbasierte Operatoren ein und Holleis et al. (2007) betrachten mehr die haptischen Tasten auf Handys und führt allgemeinere Operatoren hinzu. Holleis et al. (2007) erweitert das Keystroke-Level Modell unter anderem mit 7 neuen Operatoren, um die Dauer von fortgeschrittenen Interaktionen auf Handys vorhersagen zu können. Diese neuen Operatoren sind:

- **Macro Attencion Shift (S_{Macro}):** Wechsel zwischen dem Handy und der realen Welt.
- **Micro Attention Shift (S_{Micro}):** Wechsel zwischen Display, Tastatur und dem Hotkeybereich.
- **Distraction (X):** Ein Ablenkungsfaktor wird mit den Operationen multipliziert.
- **Action (A(t)): Aktionen, die nicht in Operatoren unterteilt werden können.**
- **Gesture (G):** Gestenerkennung wie zum Beispiel das Handy drehen oder Zahlen in die Luft malen.
- **Fingermovement (F):** Fingerbewegung von einer Position zu einer anderen.
- **Initial Act (I):** Zeit um zum Beispiel das Handy aus der Tasche zu holen.

Die Operatoren K, P und H des orginalen KLM werden etwas angepasst, M bleibt gleich und D wird nicht benötigt.

[Li et al. \(2010\)](#) kombinieren das KLM mit einem neuen Konzept von Operator Blöcken. Ein Operator Block enthält eine Sequenz von Operatoren, die häufig als Einheit in Aktionen vorkommen. Dem originalen KLM werden neun physische Operatoren hinzugefügt und fünf mentale Operatoren, wovon vier für die Stiftnutzung angelegt sind. Teils entsprechen die physischen Operatoren denen von [Holleis et al. \(2007\)](#) und teils werden deren Operatoren wieder verworfen.

2.2.5 Erweiterung des KLM für Toucheingaben

Nachdem bei mobilen Geräten wie Smartphones oder Tablets mehr und mehr die Stiftnutzung verschwindet und eingebaute Touchscreens verwendet werden wurde auch in diesem Bereich das KLM erweitert. [Abdulin \(2011\)](#) untersuchte das KLM für direkten Touch auf 7-17 Zoll großen Tablets. Die KLMs von drei verschiedenen Interfaces wurden mit dem CogTool generiert und mit Studienzeiten verglichen.

Ebenfalls auf Touchscreens fokussieren sich [El Batran and Dunlop \(2014\)](#) und [Rice and Lartigue \(2014\)](#). [El Batran and Dunlop \(2014\)](#) gewinnt in drei Studien Zeiten für drei Touch-Operatoren: Swipe, Zoom und Tap. [Rice and Lartigue \(2014\)](#) erstellen ein neues Set von Operatoren, dem Touch-Level Modell, zusammen. Dafür lassen sie die ursprünglichen Operatoren K, H, M und R(t) unverändert. Die Operatoren X und I werden von [Holleis et al. \(2007\)](#) übernommen und acht weitere Operatoren hinzugefügt:

- **Gesture (G):** Geste mit 1, 2 oder mehreren Fingern
- **Pinch (P):** 2-Fingergeste, wobei die Finger am Ende geschlossen sind
- **Zoom (Z):** 2-Fingergeste, wobei die Finger am Anfang geschlossen sind
- **Tap (T):** Finger-Tap
- **Swipe (S):** Finger auf dem Touchscreen bewegt sich in horizontaler oder vertikaler Richtung
- **Tilt (L(d)):** Neigung oder Rotation des ganzen Geräts um d Grad
- **Rotate (O(d)):** 2-Fingergeste mit der auf dem Screen etwas um d Grad rotiert wird
- **Drag (D):** 1-Fingergeste mit der meist ein Objekt von einer Position zu einer anderen gebracht wird

Das Touch-Level Modell soll in zukünftigen Studien noch verifiziert und evaluiert werden.

2.2.6 Erweiterung des KLM im automobilen Kontext

Die Infotainmentsysteme in Autos werden immer komplexer und auch hier ist es wichtig für Designer in einem frühen Stadium Interfaces testen zu können. Es gibt im Auto nicht nur ein Gerät das betrachtet werden kann, sondern mehrere Bereiche mit verschiedensten Bedienelementen. Diese Bedienelemente können ursprüngliche haptische Knöpfe und Schalter sein, aber auch Touchdisplays und mögliche Einrichtungen für Sprach- oder Gestensteuerungen.

Pettitt et al. (2007) befassen sich 2007 mit diesem Thema und erweitern das KLM für Informationssysteme im Auto, mit Hinzunahme der Occlusion Technik. Die Zeiten dieses KLM werden dann mit den Zeiten von einer Occlusion Studie verglichen. Sie verwenden die Operatoren von Card et al. (1980) und der Operator R_f (Reach far) wurde von Green (2003) adaptiert, genau wie der Operator (Reach near), dessen Zeit für die Bewegung innerhalb eines Systems mit Fitt's Law berechnet wird.

Ebenfalls im automobilen Kontext erweitert Schneegaß (2009) (siehe auch (Schneegaß et al., 2011)) das KLM und bezieht sich dabei im speziellen auf die Interaktionen von Autoradios. In dieser Studie werden vier Aufgaben in diesem Kontext mit insgesamt 78 Operatoren aufgezeichnet und deren Zeiten berechnet. Der interessanteste und für den automobilen Kontext neu hinzugefügte Operator ist der Turn (T) Operator. Dieser Operator wird für alle haptisch drehbaren Knöpfe verwendet und es wurden jeweils drei verschiedene Drehwinkel für Linksdrehungen (negativ) und Rechtsdrehungen (positiv) untersucht ($-180^\circ, -90^\circ, -45^\circ, 45^\circ, 90^\circ, 180^\circ$).

Operator	Description	Time (sec)
H	Homing Wheel – System	0.89
	Homing System – Wheel	0.81
K	K (button pressed once)	0.54
	K (button pressed twice)	1.76
	K (button pressed X times)	$2.12 + 0.22 \cdot (X-3)$
T	Turn 45° clockwise	1.10
	Turn 90° clockwise	1.16
	Turn 180° clockwise	1.74
	Turn 45° counter-clockwise	0.80
	Turn 90° counter-clockwise	1.14
	Turn 180° counter-clockwise	1.40
F	Move finger between controls	1.14
R	Response time depending on system	t
AS	Predictable list	0.30
	Unpredictable list	1.12
M	After R operator	1.35
	After T operator	1.18

Abbildung 2.2.4: Tabelle der Operatoren, deren Beschreibung und den dazugehörigen Zeiten.

Schneegaß verweist darauf, dass die Berechnung der Zeiten für den Turn(T) Operator in Zukunft noch genauer untersucht werden sollte, indem zusätzlich untersucht wird, ob in einen bestimmten Bereich gedreht oder ein bestimmter Winkel der Drehung verlangt wird. Die genauen Zeiten der einzelnen Operatoren können aus Abbildung 2.2.4 von Schneegaß et al. (2011, Seite 75) entnommen werden. Ein weiterer wichtiger Aspekt aus seinen Ergebnissen ist das Feedback der Teilnehmer dieser Studie. Es wurde angemerkt, dass die Aufgaben, mit bis zu sechs Unteraufgaben, zu lang waren. Somit war die kognitive Auslastung, sich diese Aufgaben zu merken, zu hoch.

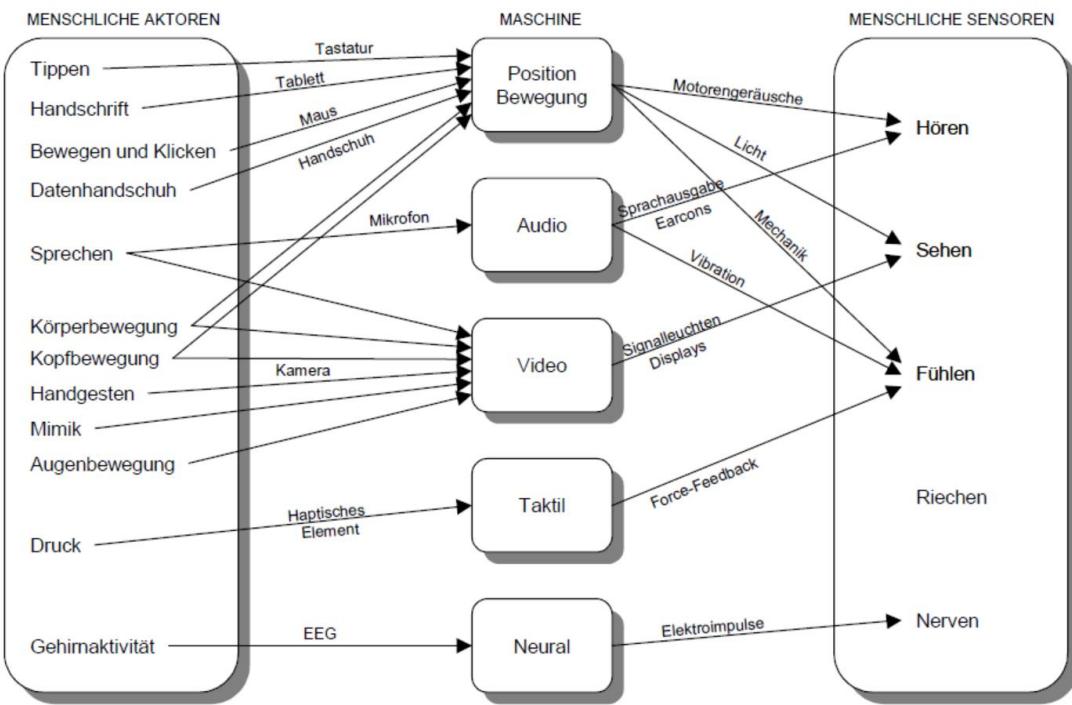


Abbildung 2.3.1: Die Abbildung von (Neuss, 2001) beschreibt das Zusammenspiel von menschlichen Aktoren und Sensoren in Interaktion mit der Maschine. Das Modell ist abgeleitet von (Sharma et al., 1998)

Außerdem reagierte das verwendetet System langsamer, als die Teilnehmer es von eigenen Autos gewohnt waren. Dieses Feedback versuchen wir in unserer Studie aufzugreifen, indem wir die Aufgaben möglichst kurz halten.

2.3 Multimodale Interaktion

Multimodale Interaktion ist für Menschen etwas ganz natürliches, das unbewusst bei den meisten Kommunikationen zwischen Menschen gemacht wird. Zwei Leute unterhalten sich zum Beispiel und nebenbei zeigt einer zusätzlich auf einen Gegenstand, schauen in eine Richtung und kombinieren all diese zusätzlichen Informationen, die das gesamte Gespräch bereichern und erleichtern. Um unsere Außenwelt besser verstehen zu können schauen, hören, tasten und riechen wir zugleich (Sharma et al., 1998). Um auch Interaktionen zwischen Mensch und Maschine intuitiver und natürlicher für zu gestalten könnte das Hinzufügen zusätzlicher Modalitäten eine große Bereicherung sein.

Multimodale Systeme können durch mehrere Kommunikationskanäle bedient werden. Die Mensch-Maschine-Interaktion geschieht zwischen den Akteuren und Sensoren des Menschen, siehe Grafik 2.3.1 von (Neuss, 2001). Doch wann kann ein Nutzer eine bestimmte Modalität verwenden oder zu einer anderen wechseln? Neuss (2001) differenziert in seiner Dissertation vier verschiedenen Ausprägungen der Multimodalität hinsichtlich der Eingabe.

- **seriell-redundant:** Modalität kann beliebig gewechselt werden

- **seriell-exklusiv:** Modalität kann nur nach bestimmten Schritten gewechselt werden
- **parallel-ergänzend:** „Simultan getätigte Eingaben mit verschiedenen Modalitäten ergänzen sich.“ ([Neuss, 2001](#))
- **parallel-verifizierend:** „Simultan getätigte Eingaben mit verschiedenen Modalitäten bestätigen sich gegenseitig“ ([Neuss, 2001](#)). Ein Beispiel wäre die Kombination aus Sprache und Lippenlesen.

Je nach Kontext wird die beste Variante gewählt. Alle Varianten, ob alleine oder in Kombination geben dem Nutzer mehr Freiheit das Ziel zu erreichen und erhöhen zudem die Effizienz ([Neuss, 2001](#)).

Eine Herausforderung ist ein gutes Design von multimodalen Interfaces, sodass die Vorteile der Modalitäten genutzt werden können, aber auch die Bedienung den Nutzer unterstützt und nicht verwirrt. Dazu haben [Reeves et al. \(2004\)](#) ein Regelwerk für das Design von multimodalen Interfaces erstellt. Deren Ziel ist es, dass in Zukunft Interaktionen natürlicher und intuitiver werden und die Robustheit durch redundante Informationen und Interaktionsmodalitäten gesteigert wird. Einige wichtige Aspekte sind hierbei das konsistente Design von Input und Output. Es soll zum Beispiel vermieden werden, dass redundante Informationen in verschiedenen Modalitäten präsentiert werden, wenn der Nutzer sich dabei auf zwei verschiedene Quellen konzentrieren muss. Der Output soll im gleichen Stil des Inputs gestalten werden. Außerdem ist es wichtig, dass ein multimodales Interface Design adaptiv und konsistent ist. Es soll ausreichend Feedback geben und eine gute Fehlerbehandlung lässt den Nutzer immer wissen, welche Fehler ihn zu einem Status gebracht haben.

2.4 Multimodale Interaktion im automobilen Kontext

Multimodale Interaktion gibt es bereits in aktuellen Autos und dieser Bereich umfasst ein großes Feld an Möglichkeiten. Eine der bekanntesten Varianten multimodaler Interaktion im Auto, ist die Sprachbedienung mit der zum Beispiel Anrufe getätigt werden. Auch Touchdisplays von externen Navigationssystemen oder Informationssystemen werden immer häufiger verwendet. Einige interessante Forschungsansätze von multimodaler Interaktion im Auto stellen wir Ihnen vor.

[Pieraccini et al. \(2004\)](#) entwickelte einen multimodalen Prototypen in einen Ford Model U. Dieser basiert auf ein sprach basiertes System, dass durch einen visuellen und haptischen Touch-Screen unterstützt wird. Diese Kombination soll Anfängern das Erlernen von Sprachbefehlen erleichtern, indem visuell Hilfen der Sprachbefehle auf dem Touch-Screen in jeder Ebene des Interfaces zu sehen sind. Für Experten ist die Nutzung von ganzen Sätzen möglich, ohne für jede Ebene ein eigenes Kommando verwenden zu müssen. Nach jedem Sprachbefehl ist es möglich durch eine Touch-Geste den Modus zu Wechseln.

Redundante Modalitäten haben viel Potenzial für das Design von modernen Interfaces, wie auch [Muller and Weinberg \(2011\)](#) in ihrem Artikel feststellten haben sie zwei signifikante Vorteile im Gebrauch von Autos: Es ist zu einem dem Fahrer möglich, die für ihn passenste Modalität je nach Situation zu wählen. Zum anderen können lange Interaktionen, von einer Modalität zur anderen, ohne Probleme gewechselt werden.

[Bertoldi and Filgueiras \(2010\)](#) haben die Multimodalität in Fahrer Assistenz Systemen untersucht,

mit dem Ergebnis, dass es in diesem Feld noch viel zu Forschen und zu verbessern gibt. Aber es hat das Potenzial die Sicherheit zu erhöhen, die Reaktionszeiten und die Eyes-off-the-road Zeit zu verkürzen und das Bewusstsein der Fahrersituation zu erhöhen.

Döring et al. (2011) erweiterten ein Lenkrad in der Mitte mit einer Multi-Touch Oberfläche und untersuchten deren Möglichkeiten für Touch-Gestensets. Die visuelle Ablenkung konnte durch Touchgesten reduziert werden.

Eine ähnliche Idee hatten Pfleging et al. (2012). Sie entwickelten, unter Rücksichtnahme von guter Lernbarkeit, Sichtbarkeit, Granularität und der Möglichkeit Aktionen rückgängig zu machen, ein im Lenkrad integriertes multimodales System, das Sprache und Touchgesten kombiniert. Es stellte sich heraus, dass Sprache und Gesten langsamer sind als der Gebrauch von haptischen Buttons, jedoch bei ähnlicher Fahrperformanz die visuelle Anstrengung geringer ist. Ein Nachteil ist das nötige Erlernen und Erinnern von Sprachbefehlen und Gesten. In deren System wurde zuerst die Sprache verwendet um Objekte oder Funktionen direkt ohne hierarchische Struktur zu benennen und anschließend konnte mit einer Touchgeste deren Parameter justiert werden, siehe auch (Pfleging et al., 2011). Bei den Sprachbefehlen wurden von 82,1% die Benennung von Objekten bevorzugt. Bei den Touchgesten benutzten 78,1% der Teilnehmer 1-Finger Touchgesten.

Stracke (2014) untersuchte in seiner Bachelorarbeit Touch Interactionen auf der Mittelkonsole. In dieser Arbeit wurde Direct-Touch-Steuerung mit „drei Formen positionsunabhängigen Wischgesetzen“ (Stracke, 2014, Seite 57) dem Serial Swipe, dem Relative Swipe und einem Relative Swipe mit Multi-Touch verglichen. Mit der Direct-Touch-Variante konnte eine Verkürzung der Interaktionszeit festgestellt werden. Jedoch schnitt diese Variante bei der Fehlerrate schlechter als die anderen Varianten ab, allerdings konnte keine Signifikanz bewiesen werden. Sehr beliebt war die Direct-Touch-Steuerung zur Einstellung der Luftverteilung und für Ein-/Ausschaltoptionen.

Im nächsten Schritt verschaffen wir uns in einem Workshop einen Überblick über multimodale Interaktionen im Auto.

Kapitel 3

Workshop zur Bestimmung eines Aktionssets

Um ein Konzept für unser Modell abzuleiten, interessiert uns welche Arten von Interaktionen, in einem Auto stattfinden und aus welchen Einzelschritten diese bestehen. Darüber hinaus wollen wir aktuelle und mögliche Umsetzungen der Modalitäten Haptik, Touch, Sprache und Gestik zu den Einzelschritten sammeln. Die Interaktionen sollen in die Kategorien Kommunikation, Navigation, Medien, Komfortfunktionen und Einstellungen unterteilt werden.

Zu diesem Zweck wollen wir im Rahmen eines Workshops unsere Ziele erarbeiten. Dafür ist eine Fokusgruppe geeignet, die sich im automobilen Bereich befindet. Bei BMW gibt es wöchentlich die Möglichkeit eine Brainstorming Runde (Connected Minds) mit freiwilligen internen Mitarbeitern anzumelden. Diese Möglichkeit nutzen wir, um unseren Workshop durchzuführen. Am 24.08.2016 führten wir mit insgesamt neun Teilnehmern unseren Workshop durch. Wir bereiteten unsere Ziele vor, die es zu erreichen galt und wie folgt lauten:

1. **Ziel A:** Anwendungsbeispiele von Interaktionen im Auto sammeln und diese einer der Kategorien zuordnen (Kommunikation, Navigation, Medien, Komfortfunktionen und Einstellungen). Anschließend diese Anwendungsbeispiele in einzelnen Aktionen zerlegen.
2. **Ziel B:** Die Aktionen aus den Anwendungsbeispielen gruppieren und Umsetzungen für die verschiedenen Modalitäten (Haptik, Sprache, Touch und Geste) finden beziehungsweise sammeln.

3.1 Organisation und Ablauf des Workshops

Nach einer kurzen Vorstellrunde wurden die Teilnehmer in das Thema eingewiesen und die zwei zu erreichenden Ziele vorgestellt. Für das Ziel A sammelten wir im ersten Schritt Anwendungsbeispiele und ordneten sie den Themengebieten Kommunikation, Navigation, Medien, Komfortfunktionen oder Einstellungen zu (siehe Abbildung 3.1.2 auf Seite 31 linkes Bild). Dazu wurden die Anwendungsbeispiele auf Karten geschrieben und auf einem White-Board in einer Spalte zur passenden Kategorie gepinnt.

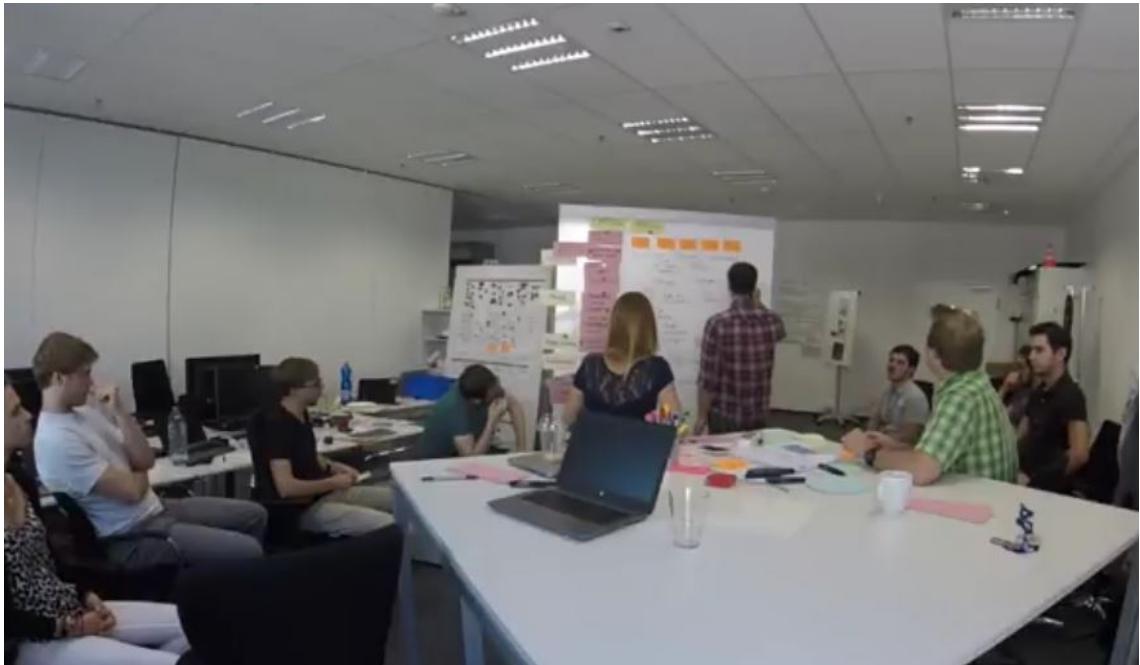


Abbildung 3.1.1: Connected Minds bei BMW

Als nächsten Schritt gliederten wir jedes Anwendungsbeispiel in ihre Einzelschritte. Diese Einzelschritte sollen Aktionen darstellen, die eine Ebene über den Operatoren eines Keystroke-Level Modells stehen. Zum Beispiel würde die Aktion einen Button auszuwählen oder zu drücken, im Keystroke-Level Modell aus einem mentalen Operator, einem Homing Operator und einem Keystroke Operator bestehen. Wir betrachten das Auswählen eines Buttons als eine komplette Aktion, die diese Teilschritte enthält. Bei der Einteilung der Anwendungsbeispiele in ihre Aktionen achteten wir darauf, dass sie möglichst abstrakt und Interface-unabhängig formuliert werden. Ein Beispiel hierfür ist die Abfolge Auswählen, Inkrementieren und Bestätigen. Alle Aktionen eines Anwendungsbeispiels wurden nacheinander neben das Anwendungsbeispiel geschrieben (siehe Abbildung 3.1.2 auf der nächsten Seite linkes Bild).

Die gesammelten Aktionen wurden für Ziel B gruppiert und auf Karten geschrieben. Diese Aktionen wurden anschließend auf einem anderen White-Board untereinander gepinnt. Für jede Aktion wurden sowohl aktuell bestehende als auch potentiell mögliche Umsetzungen für die Modalitäten Haptik, Touch, Geste und Sprache gesammelt (siehe Abbildung 3.1.2 auf der nächsten Seite rechtes Bild und Tabelle 3.2.1 auf Seite 32).

Der Workshop dauerte insgesamt 1,5 Stunden, davon 30 Minuten für das Erreichen des Ziels A und eine Stunde für das Erreichen von Ziel B. Der Workshop wurde auf Video aufgezeichnet.

3.2 Ergebnisse des Workshops Connected Minds

Im Workshop wurden insgesamt sechs verschiedene Aktionen identifiziert, aus denen alle Bedienbeispiele zusammengesetzt werden können.

- **Bestätigung (B):** Eingabe oder Auswahl bestätigen.



Abbildung 3.1.2: Connected Minds Ergebnisse - Ziel 1 (links), Ziel 2 (rechts)

- **Direktauswahl (DA):** Auswahl aus sichtbaren Elementen.
- **Ein-/Ausschalten (E/A):** Ein- und Ausschalten oder Aktivieren und Deaktivieren von Funktionen.
- **Listen-Navigation (L):** Seitenweise inkrementieren und anschließende Direktauswahl (DA).
- **De-/Inkrementieren (Inkr):** Scrollen, Erhöhen/Verringern von Werten.
- **Texteingabe (T):** Wiederholte Eingabe eines Buchstabens und anschließende Bestätigung (B).

Für die sechs Aktionen wurden für die Modalitäten Haptik, Touch, Geste und Sprache vorhandene oder mögliche Umsetzungen gesammelt. Tabelle 3.2.1 auf der nächsten Seite listet alle Vorschläge der verschiedenen Modalitäten auf. Wir entschieden uns bei dem Touchmodus immer für den direkten Touch auf einem Display. TODO(den satz laenger und umformulieren) In der schon erwähnten Bachelorarbeit von [Stracke \(2014\)](#) erwiesen sich diese Varianten bei Touch für eine gute Umsetzung mit schnellen Interaktionszeiten. Auch [Rümelin and Butz \(2013\)](#) bekamen beim direkten Touch die besten Zeiten für das Vollenden einer Aufgabe.

Um den Umfang der geplanten Studie im realistischen Bereich zu lassen, haben wir uns entschieden die Haptik als Modalität nicht miteinzubeziehen. Unser Fokus soll bei den Modalitäten Touch, Geste und Sprache liegen, auch weil Haptik im automobilen Kontext bereits in Studien umfassend untersucht wurde ([Pettitt et al., 2007](#); [Schneegäß, 2009](#); [Schneegäß et al., 2011](#)). Die haptische Bedienung stellt natürlich weiterhin einen wichtigen Bestandteil der Interaktion im Auto dar. Da dieser Bereich für uns wegfällt benötigen wir somit die Aktion Ein- und Ausschalten (E/A) nicht. Im nicht haptischen Kontext unterscheidet sich diese Aktion nicht von einer gewöhnlichen Aktivierung eines Buttons.

Aktion	Haptik	Touch	Gesten	Sprache
B	Button drücken Dreh-Drücker	Tap	Daumen hoch, Tauch OK, bestimmte Geste halten	Ja, OK, passt
DA	Button drücken Dreh-Drücker	Swipe Tap	3D-Pointing	Objekt/Funktion benennen
E/A	Button (on/off) Kippschalter	Tap, Swipe	von offener zu geschlossener Hand	Funktionsname „ein“/„aus“
L	Button drücken Dreh-Drücker	Swipe Tap Scroll	relative Handbewegung (Wischgeste, Swipe), Drehbewegung des Fingers	hoch/runter mit Start/Stop, weiter/zurück, Tonhöhe variieren
Inkr.	Button drücken Dreh-Drücker Hebel drücken	Swipe Tap Scroll	relative Handbewegung (Wischgeste, Swipe), Drehbewegung des Fingers	hoch/runter mit Start/Stop, weiter/zurück, Tonhöhe variieren
T	Tastatur Dreh-Drücker Morsecode	X mal Tap Schreiben	In der Luft schreiben, 3D-Pointing auf Tastatur, L auf Tastatur	diktieren, sprechen

Tabelle 3.2.1: Ergebnisse der Umsetzungsmöglichkeiten der Modalitäten für die verschiedenen Aktionen

Geeignete Anwendungsbeispiele: Aus den im Workshop genannten Beispielen wurden fünf stellvertretende Beispiele identifiziert, die sowohl alle Aktionen abdecken, als auch zu einem der fünf Themengebiete Navigation, Medien, Komfortfunktionen, Einstellungen und Kommunikation zuzuordnen sind.

- Navigieren zu „Rom“, „Dorfweg“ und „Kirchengasse“ (Navigation): DA (Navigation) + DA (Zieleingabe) + T (Ziel) + B (Bestätigen)
- Song „Happy“ aus beliebten Songs wählen (Medien): DA (Medien) + DA (beliebte Songs) + L („Happy“ suchen) + DA („Happy“ auswählen)
- Temperatur um 3 Grad erhöhen (Komfortfunktionen): DA (Temperatur) + Inkr. (+ 3°)
- Lautstärke erhöhen (Einstellungen): DA (Einstellungen) + Inkr. (von 50% auf ca. 80%)
- „Maria Müller“ aus Kontakten anrufen (Kommunikation): DA (Telefon) + DA (Kontakte) + L („Maria Müller“ suchen) + DA („Maria Müller“ auswählen)

Bei der Einstellung der Lautstärke haben wir uns bewusst für einen groben Zielwert entschieden. Die Lautstärke soll in ein Intervall zwischen 75% und 85% gesetzt werden, da ein genauer Wert (zum Beispiel auf exact 80% erhöhen) eine unnatürliche Aufgabenstellung wäre (Stracke, 2014). Unsere Anwendungsbeispiele stellen übliche Anwendungen dar und sind teils in vereinfachter Weise dargestellt, um eine zu hohe kognitive Belastung zu vermeiden. Da die Zeiten von Experten gemessen werden sollen, ist es wichtig Fehler zu eliminieren und Bedenkzeiten zu minimieren.

Die Inkrementation eines Wertes stellen wir in unseren Beispielen in zwei verschiedenen Varianten dar. Zum einen soll die Temperatur um 3 Grad erhöht werden, indem durch dreimaliges inkrementieren schrittweise der Wert verändert wird. Im anderen Beispiel, die Lautstärke von 50% auf ca.

80% zu inkrementieren, wählen wir einen Slider als Darstellung. Hier muss nicht eine Aktion 30 mal angewendet werden, um von 50 auf 80 zu inkrementieren, sondern eine direkte Inkrementation (für die Modalitäten Touch und Geste) soll den Wert verändern. Wir unterscheiden also zwischen der schrittweisen Inkrementation (Inkr. (s)), die für kleine Wertunterschiede geeignet ist, und der direkten Inkrementation (Inkr. (d)), die bei größeren und größeren Werteinstellungen vorteilhaft ist.

Bei der Texteingabe eines Ziels haben wir zur Unterscheidung von kurzen und langen Wörtern drei verschiedene Ziele mit unterschiedlicher Länge gewählt („Rom“, „Dorfweg“ und „Kirchengasse“). Außerdem wurde darauf geachtet möglichst einfach zu schreibende Wörter zu verwenden, um negative Einflüsse in Bezug auf Rechtschreibkenntnissen zu vermeiden. Es soll in erster Linie die Zeit gemessen werden, die benötigt wird einen Button bestimmter Größe zu treffen. Die mentale Zeit soll sich lediglich auf die Interaktion und das Interface beziehen, nicht jedoch auf die Rechtschreibung.

3.3 Abgeleitete Idee für ein multimodales Modell

Mit unserem Modell sollen sich multimodale Interaktionszeiten vorhersagen lassen, indem sie sich aus den Aktionen Direktauswahl aus sichtbaren Elementen (DA), Bestätigung (B), Listen-Navigation (L), schrittweiser Inkrementation (Inkr.(s)), direkter Inkrementation (Inkr.(d)), und der Texteingabe (T) zusammensetzen.

Bei den Keystroke-Level Modellen auf Operatorebene gibt es eine „richtige“ Lösung mit einer bestimmten Anzahl an Operatoren. Das ist der schnellste Weg, den Experten verwenden, um eine bestimmte Aufgabe zu lösen. Bei Interaktionen im Auto ist es deutlich schwieriger eine Abfolge von Operatoren zu bilden. Je nach Nutzer und Situation im Auto werden zum Beispiel mehrere Bewegungen vom Lenkrad hin zum Interaktionsbereich gemacht als nötig. Deshalb erscheint es uns eine gute Annäherung die Gesamtdauer in unserem Modell nicht durch Operatoren, sondern durch Aktionen zusammenzusetzen. Je nach Modalität unterscheiden sich die Zeiten einer Aktion.

Hinzu kommen Wechselkosten, die auftreten wenn ein Wechsel von einer Modalität zu einer anderen entstehen. Auch diese Zeiten gilt es zu berechnen und zu berücksichtigen. Die Wechselkosten, enthalten unter anderem einen mentalen Operator und einen Homing Operator (Card et al., 1980) bzw. einen Reach Far Operator (Green, 2003).

Eine Aktion soll in unserem Fall eine durchschnittliche Dauer darstellen, die ein geübter Nutzer benötigt in einem bestimmten Modus die Aktion fehlerfrei auszuführen. Wie bei dem KLM soll auch bei unserem Modell von Experten ausgegangen werden. Es sollen nur Durchgänge gewertet werden, die keine Fehler vom System oder dem Nutzer mit dem System aufweisen. Ob ein Nutzer sich dabei Zeit lässt oder einmal mehr oder weniger die Hand vom Interaktionsbereich zurück zum Lenkrad nimmt stellt dabei keinen Fehler dar. Dies soll in den durchschnittlichen Zeiten berücksichtigt werden.

Kapitel 4

Umsetzung des multimodalen Prototyps

Im nächsten Schritt sollen die Anwendungsbeispiele in geeigneter Weise umgesetzt werden, um in einer Studie Interaktionszeiten erheben zu können.

Unser erster Ansatz war das Interface als Klickdummy zu erstellen und mit der „Wizard of Oz“ Methode umzusetzen ([Salber and Coutaz, 1993](#)). In der „Wizard of Oz“ Methode simuliert der sogenannte „Wizard“ die Funktionalität eines Systems, indem er Events durch seine Beobachtungen manuell auslöst. Das heißt die Funktionalität, bestimmte Gesten oder Sprachbefehle zu erkennen, muss in diesem Fall nicht implementiert werden und ist somit auch nicht anfällig für Fehler. Dies ist ein großer Vorteil für ein Modell, dass von Experten ausgeht, die mit dem System vertraut sind und keine Fehler machen. Auch Fehler des Systems können mit dieser Methode ausgeschlossen werden.

Da in der „Wizard of Oz“ Methode, Events nur durch den „Wizard“ ausgelöst werden können, ist es schwieriger diese Events zu protokollieren. Studien dieser Art werden meist per Video aufgezeichnet und anschließend mit einer Videoanalyse ausgewertet. Natürlich gibt es einige Hilfen, um eine Videoanalysen zu optimieren, allerdings bleibt der Zeitaufwand hoch, weshalb wir uns dafür entschieden haben diese Methode nicht zu verwenden. Mit einem vollständig funktionsfähigem Prototypen können Events mit Zeitstempeln protokolliert werden, um später die Zeiten für unsere Aktionen und deren Wechselkosten berechnen zu können.

Im folgenden werden die Anwendungsbeispiele und deren Implementation erläutert.

4.1 Beschreibung und Umsetzung der Anwendungsbeispiele

Die im Workshop erarbeiteten Anwendungsbeispiele werden in vereinfachter und abstrakter Darstellung im Prototypen umgesetzt. Damit wollen wir vermeiden, dass ein auffälliges Design den Nutzer ablenkt. Es sollen lediglich die Interaktionszeiten untersucht werden und nicht eine bestimmte Umsetzung eines Interfacedesigns. In Abbildung 4.1.1 auf der nächsten Seite werden die Anwendungsbeispiele gezeigt welche im Folgenden beschrieben werden, ohne auf die Verwendung der Modalitäten einzugehen. Auf die Modalitäten gehen wir im Abschnitt Implementation unter der jeweiligen Modalität ein.

Diese vier Anwendungsbeispiele sollen in der Studie sowohl unimodal, als auch mit allen multimodalen Kombinationen getestet werden. Jeder Proband kann sich in einem Probendurchgang mit der Aufgabe vertraut machen. Anschließend folgen zwei Messdurchgänge, in denen die Interaktionszeiten erhoben werden.

In Abbildung 4.1.1 werden die vier umgesetzten Anwendungsbeispiele in einer Übersicht horizontal dargestellt. Hier wird auch deutlich, ab welchen Screens ein Moduswechsel stattfinden soll. Der Moduswechsel wurde festgelegt und ist immer mit einem Screenwechsel verbunden. Durch diese Vorgabe ist unser Prototyp exklusiv redundant, da nur innerhalb eines Screens die Modalität gewechselt werden soll.

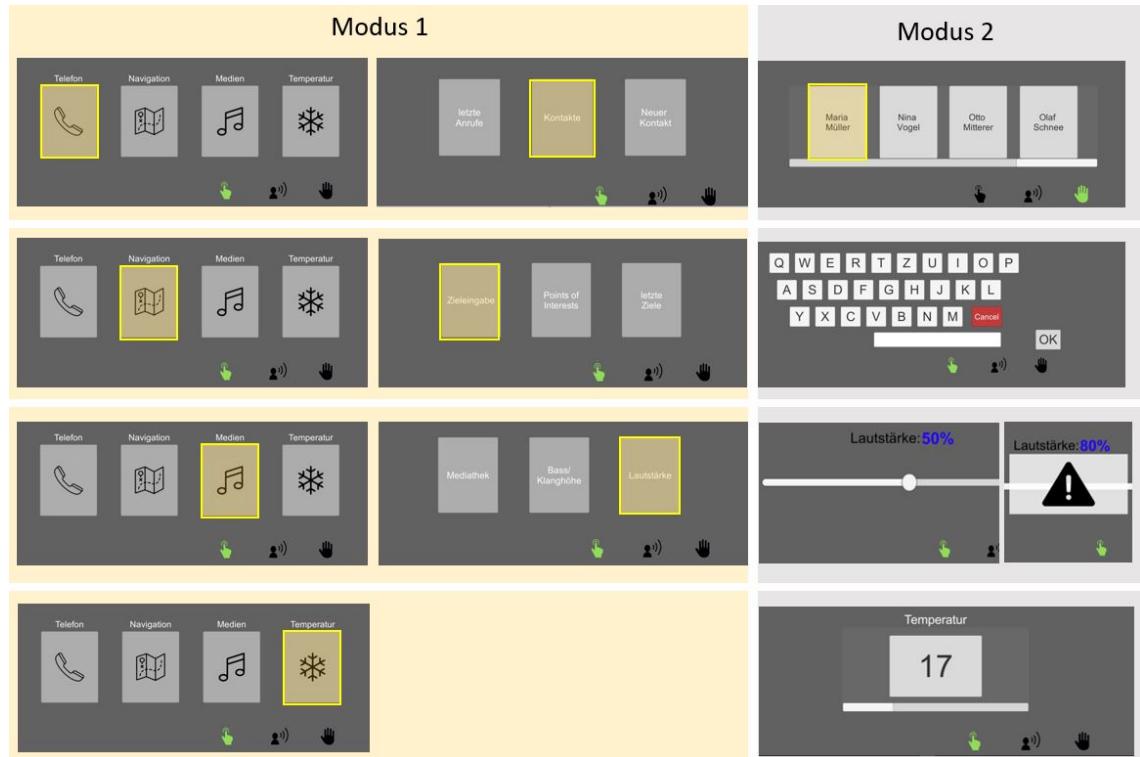


Abbildung 4.1.1: Übersicht der Screenabfolge, der 4 Anwendungsbeispiele. Ein Anwendungsbeispiel besteht aus den horizontal abgebildeten Screens. Vertikal werden die Anwendungsbeispiele in ersten Modus und zweiten Modus geteilt. Dazwischen findet der Moduswechsel statt. Referenz zu den Icons siehe Kapitel 8

Im ersten Anwendungsbeispiel soll „Maria Müller“ angerufen werden. Dafür muss im Hauptmenü die Kachel „Telefon“ auf der linken Seite ausgewählt werden. Sobald diese Kachel mit einem Modus aktiviert wurde, wechselt der Screen in ein Untermenü mit drei Optionen, von denen die Kachel Kontakte auszuwählen ist. Jetzt muss durch eine horizontale Liste von Kontakten navigiert werden. Die Liste lässt sich seitenweise scrollen. Auf der dritten Seite befindet sich auf der linken Kachel der gewünschte Kontakt „Maria Müller“, der auszuwählen ist. Dieses Anwendungsbeispiel besteht aus einer zweifachen Direktauswahl aus sichtbaren Elementen, einer Listennavigation bestehend aus drei Swipes gefolgt von einer Direktauswahl aus sichtbaren Elementen. Unsere Aktionen sind:

$$2 * \text{DA} + 3 * \text{L} + \text{DA}$$

In der zweiten Anwendung soll zu einem von drei verschiedenen Zielen navigiert werden. Dazu

wird im Hauptmenü die Navigationskachel ausgewählt und anschließend im nächsten Untermenü die Kachel Zieleingabe. Um das Ziel einzugeben ist eine vereinfachte Tastatur abgebildet, die im Touchmodus verwendet werden kann. Mit dem „OK“ Button soll die Zieleingabe bestätigt werden. Hier bestehen die Aktionen aus zwei Direktauswahlen aus sichtbaren Elementen, einer Texteingabe und der Bestätigung. Unsere Aktionen sind:

2 * DA + Xmal Buchst. + B

In der dritten Anwendung soll die Lautstärke von 50% auf 80% erhöht werden. Dazu wird im Hauptmenü die Kachel „Medien“ selektiert. Im darauf folgenden Untermenü ist die Kachel „Lautstärke“ auszuwählen. Die Lautstärke muss mit einem horizontalen Slider auf einen Wert zwischen 75% und 85% gestellt werden. Dann erscheint ein Warnung als Popup, dass bestätigt werden muss. Auch hier setzt sich die Interaktion aus einer zweimaligen Direktauswahl aus sichtbaren Elementen, einer direkten Inkrementation des Sliders und einer Bestätigung zusammen. Unsere Aktionen sind:

2 * DA + Inkr. (d) + B

Das letzte Anwendungsbeispiel hat zum Ziel die Temperatur von 17 auf 20 Grad zu erhöhen. Dazu muss im Hauptmenü die Kachel „Temperatur“ selektiert werden. Die aktuelle Temperatur wird angezeigt und kann durch schrittweise Inkrementation einer horizontalen Liste erhöht werden. Es ist jeweils nur ein Element der Liste, das heißt eine Temperatur, sichtbar. Dieses Anwendungsbeispiel ergibt sich aus einer Direktauswahl aus sichtbaren Elementen und einer dreifachen Inkrementation des Wertes. Unsere Aktionen sind:

DA + 3 * Inkr. (s)

Das Anwendungsbeispiel „Einen Song aus einer Liste auszuwählen“ wurde nicht umgesetzt, da ein weiteres Anwendungsbeispiele die Studiendauer deutlich erhöht hätte. Außerdem ähnelt diese Interaktion sehr dem Anwendungsbeispiel einen Kontakt aus einem Liste zu wählen und hätte somit keinen wesentlichen Mehrwert. Wir werden dieses Anwendungsbeispiel für die Evaluation verwenden.

Die Direktauswahl aus sichtbaren Elementen wurde immer im ersten Modus ausgeführt. Anschließend folgt ein Moduswechsel. Die drei möglichen Modalitäten wurden mit kleinen schwarzen Symbolen für Touch, Sprache und Geste in jedem Screen angezeigt. Der Modus, den der Proband verwenden soll wurde grün hervorgehoben. In Abbildung 4.1.2 soll beispielsweise mit dem Modus Touch interagiert werden.



Abbildung 4.1.2: Symbolische Anzeige des aktiven Modus, der von den Probanden ausgeführt werden soll. Referenz zu den Icons siehe Kapitel 8

Ist die Modalität Geste aktiv, haben wir eine zusätzliche Information in die Gestensymbolik eingebaut. In der aktiven grünen Hand ist ein Punkt zu sehen, der wie bei einer Ampel entweder rot, gelb oder grün ist. Dieser Punkt dient dazu besser erkennen zu können, ob und wo die rechte Hand

vom Leap Motion Controller erkannt wird. Wie in Abbildung 4.1.3 auf der nächsten Seite zu sehen ist, ist der Punkt rot, wenn die Hand von dem Controller nicht erkannt wird. Wird die Hand vom Controller erkannt, befindet sich aber nicht im Interaktionsbereich, indem Gesten ausgeführt werden können, ist der Punkt gelb. Grün wird der Punkt, sobald die Hand erkannt wurde und sich im Interaktionsbereich befindet. Mit dieser Visualisierung soll für den Beobachter veranschaulicht



Abbildung 4.1.3: Ampeldarstellung zur Gestenerkennung. Der rote Punkt bedeutet, dass die rechte Hand vom Leap Motion Controller nicht erkannt wurde. Gelb wird der Punkt, wenn die Hand erkannt wurde, sich aber nicht im Interaktionsbereich befindet. Grün ist der Punkt wenn sich die Hand im Interaktionsbereich befindet. Referenz zu den Icons siehe Kapitel 8

werden, falls Gesten nicht erkannt wurden.

Wird ein Button angewählt bekommt dieser eine gelbe Markierungsfarbe. Bei einer Selektion bekommt der Nutzer ein zusätzliches Audio Feedback eines Klickgeräusches.

4.2 Implementation des multimodalen Prototypen

Der multimodale Prototyp wurde in Unity (Version 5.4 ([Technologies, b](#))) auf einem Surface umgesetzt, was die Touch Interaktion ermöglicht. Für die Gestenerkennung wird ein Leap Motion Controller ([Motion](#)) verwendet. In Unity wurde ein Software Developer Kit namens Orion ([Technologies, a](#)) verwendet, welches in die Unity Szene eingebunden wird. Um die Spracherkennung zu gewährleisten wurde in Unity ein Skript eingebunden, dass auf die Windows integrierte Spracherkennung zugreift. Die Skripte wurden mit Visual Studio in C# geschrieben.

Jeder Screen wurde in Unity in eine eigene Szene eingebaut. Die Selektion eines Buttons durch ein Touch-Event, eine Selektionsgeste oder durch einen Sprachbefehl, lädt die nächste Szene. Im ersten Screen wird vom Versuchsleiter die Proband ID eingetragen und die passende Permutation geladen (mehr zur Permutation siehe Kapitel 5). In der Permutation steckt das aktuelle Anwendungsbeispiel, sowie die zu verwendenden Modalitäten. Die permutierte Reihenfolge der Kombinationen verhindert das eventuell entstandene Lerneffekte unsere Messungen beeinflussen.

Der Versuchsleiter kann mit den Pfeiltasten einer externen Tastatur das nächste Anwendungsbeispiel laden, einen Messdurchgang aktiviert und den Durchgang gestartet. Zur Vorbereitung wird als erstes ein schwarzer Screen für drei Sekunden gezeigt. Anschließend wird der Hauptscreen geladen, mit dem die Interaktion beginnt.

Um Fehler zu reduzieren ist es mit den Modalitäten Geste und Sprache nicht möglich eine falsche Auswahl zu tätigen. Wir haben uns entschieden für die Sprache keinen Push To Talk Button zu

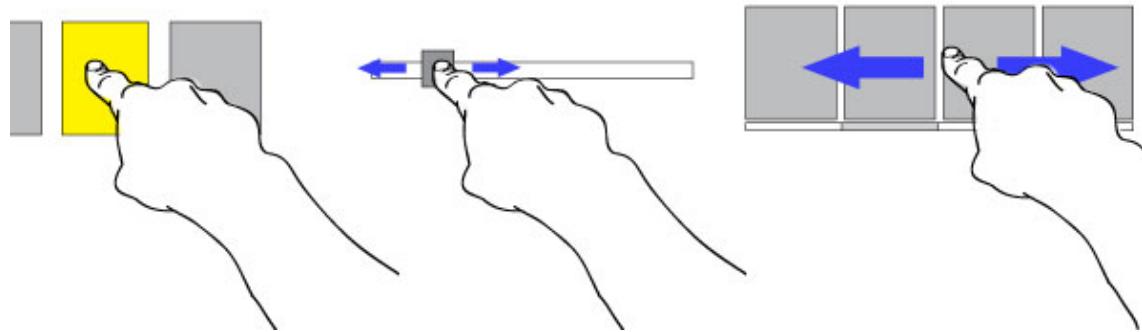


Abbildung 4.2.1: Touchgesten: Tap, Slide und Swipe

verwenden, da wir Aktionszeiten messen wollen und wir uns auf die Modalitäten Touch, Geste und Sprache fokussieren. Sollte dieser haptische Knopfdruck verwendet werden, müsste die Zeit für diese Aktion hinzugefügt werden.

4.2.1 Realisierung der Toucheingabe

[Wobbrock et al. \(2009\)](#) untersucht von Nutzern definierte Touchgesten auf einem Surface. Sie fanden heraus, dass die Anzahl der Finger für eine Touchgeste für die meisten Nutzer irrelevant ist und das einhändige Touchgesten bevorzugt werden. Bei unseren Touchgesten entschieden wir uns für drei verschiedene Touchgesten:

- Einen direkten Touch (Tap), um einen Button auszuwählen. Die Selektion geschieht erst wenn der Finger den Touchscreen verlässt.
- Eine Swipegeste, um eine Liste seitenweise zu scrollen oder einen Wert zu inkrementieren.
- Eine kontrollierte Slidegeste, um einen Slider zu verstellen. Hierbei wird der Regler des Sliders ähnlich wie bei Drag and Drop direkt verschoben, allerdings nur entlang des Achse des Sliders. Der Finger bleibt während der Verschiebung auf der Touchfläche.

In der [4.2.1](#) sind die 3 Varianten veranschaulicht.

4.2.2 Realisierung der Gesteneingabe

Auch bei der Erkennung der Gesten wurden drei verschiedene Arten implementiert. Um einen Button zu selektieren wird zuerst geprüft, ob sich eine rechte Hand im festgelegten Interaktionsbereich befindet. Dieser Interaktionsbereich ist je nach Anzahl der Buttons in drei oder vier Bereiche entlang der x-Achse unterteilt. Je nachdem in welchem Bereich sich die Hand befindet ändert der passende Button dieses Bereichs die Farbe zu gelb. Dies entspricht der Markierungsfarbe. Somit erkennt der Nutzer welcher Button gerade angewählt ist. Um einen Button nun zu selektieren, muss der Zeigefinger schnell nach unten bewegt werden, ohne dabei die Hand zu bewegen. Nur wenn ein entsprechender Button gelb markiert ist und die Selektionsgeste mit dem Zeigefinger ausgelöst wurde, wird dieser Button selektiert und somit die nächste Szene geladen.

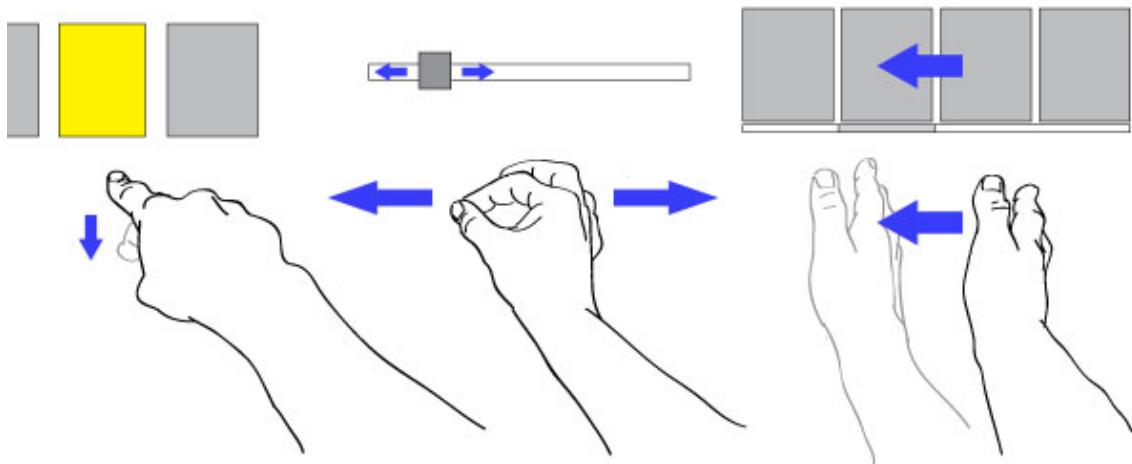


Abbildung 4.2.2: Selektionsgeste, Slidegeste und Swipegeste

Die nächste Geste ist eine Wischgeste, um in einer Liste seitenweise zu scrollen oder einen Wert zu inkrementieren. Das entspricht unseren Aktionen (L) und (Inkr. (s)). Diese Geste wurde in den Anwendungsbeispielen verwendet, um die Temperatur zu verändern und um die Liste der Kontakte zu durchsuchen. Wie zuvor muss die rechte Hand erkannt werden und sich im Interaktionsbereich befinden. Wenn sich die Hand entlang der x-Achse mit mindestens 80 Millimeter pro Sekunde von rechts nach links bewegen, löst dies eine Animation aus, die zur nächsten Seite navigiert. Damit die Hand nicht unbeabsichtigt zwei Seiten auf einmal oder direkt hintereinander scrollen lässt, wird die Erkennung der Wischgeste nach jeder Erkennung für eine Sekunde deaktiviert.

Die dritte und letzte Geste ist eine Geste, um einen Slider zu verstellt. Hierfür müssen sich Daumen und Zeigefinger berühren und somit einen geschlossenen Kreis über dem Leap Motion Controller bilden. Damit dieser Kreis von der Leap erkannt wird, ist es wichtig die anderen Finger abzuspreizen, siehe 4.2.2. Ist dieser Kreis geschlossen wird der Slider aufgenommen und kann verschoben werden. Um den Slider nach rechts oder links zu verstellen muss die Geste entlang der x-Achse durch Bewegung der Hand in eben diese Richtung verändert werden. Ist der gewünschte Wert erreicht, muss der Kreis geöffnet werden, indem sich Zeigefinger und Daumen wieder lösen.

4.2.3 Realisierung der Spracheingabe

Um die Spracherkennung zu ermöglichen wurden in Unity in jeder Szene die dementsprechenden Sprachbefehle als String definiert. Im Beispiel des Hauptscreens waren dies die Sprachbefehle: „Telefon“, „Navigation“, „Medien“ und „Temperatur“. Die in Windows integrierte Spracherkennung verarbeitet gesprochene Befehle zu Strings. Stimmt ein String mit einem der vordefinierten String überein, wird der passende Button selektiert.

Damit der Nutzer bei der Direktauswahl aus sichtbaren Elementen zusätzlich zum Soundfeedback Feedback bekommt, wird der selektierte Button vor dem tatsächlichen Klick farblich gelb hervorgehoben. Um diesen Effekt erzielen zu können musste eine Verzögerung von einer halben Sekunden eingebaut werden.

Im Beispiel der Listen wird zur Veranschaulichung erst zum entsprechenden Wert geswiped, das heißt eine Animation ist sichtbar und anschließend wird die Kachel selektiert. Im Beispiel der

ms Gesamt	Events	Szenen	Probe	Messdurchgang	Anw.	Mo	Modus	Fr
59886607	SCENE	MainUI2	2	TRUE	Sound	0	0	0
59889376	Touch in Main erkannt		2	TRUE	Sound	0	0	0
59889461	Button geklickt: Media		2	TRUE	Sound	0	0	0
59889477	SCENE	MainUI_Medien_Sound	2	TRUE	Sound	0	0	0
59889494	Touch in Main endet.		2	TRUE	Sound	0	0	0
59889515	Touch in Main erkannt		2	TRUE	Sound	0	0	0
59889531	Touch in Main endet.		2	TRUE	Sound	0	0	0
59891424	Touch in Options erkannt		2	TRUE	Sound	0	0	0
59891642	Button geklickt: Lautstärke		2	TRUE	Sound	0	0	0
59891660	SCENE	MainUI_Sound	2	TRUE	Sound	0	0	0
59891679	Touch in Options endet.		2	TRUE	Sound	0	0	0
59891694	Touch in Options erkannt		2	TRUE	Sound	0	0	0
59891709	Touch in Options endet.		2	TRUE	Sound	0	0	0
59893159	Touch in Sound: Slide startet bei 0.5		2	TRUE	Sound	0	0	0
59894977	Touch in Sound: Slide endet bei 0.8080809. Anzeige ist: 80%		2	TRUE	Sound	0	0	0
59896041	Touch in Sound: Slide startet bei 0.8080809		2	TRUE	Sound	0	0	0
59896110	Button geklickt: Popup		2	TRUE	Sound	0	0	0
59896127	SCENE	LastScene_Smileys	2	TRUE	Sound	0	0	0

Abbildung 4.2.3: Ausschnitt aus der Logdatei vom Anwendungsbeispiel Lautstärke, das mit den Modalitäten Touch und Touch, also unimodal ausgeführt wurde

Texteingabe wird das Inputfeld mit dem gesprochenen Ziel aktualisiert.

4.2.4 Protokollierung der relevanten Events

Um die Interaktionszeiten messen zu können werden die relevanten Aktionen mit Zeitstempeln in einer Textdatei protokolliert (Logdatei). Diese Logdatei ist so strukturiert, dass Werte durch Tabs separiert sind, um später in Excel leichter bearbeitet werden zu können. An relevanten Stellen im Code wird die Logdatei geöffnet und der Zeitstempel zusammen mit dem Event in die Datei geschrieben und wieder geschlossen. Zur Veranschaulichung ist in Abbildung 4.2.3 ein Auszug des Protokolls zusehen. Die Logdatei wurde hier bereits angepasst und in Excel als Tabelle formatiert (mehr dazu in Kapitel 5).

Sobald ein Anwendungsbeispiel startet werden die Einstellungsinformationen siehe Abbildung 4.2.4 auf der nächsten Seite protokolliert. Diese enthalten die ID des Probanden, das aktuelle Anwendungsbeispiel, die zu verwendenden Modalitäten und ob es sich um einen Probe- oder um einen Messdurchgang handelt. Nach jedem Messdurchgang werden die Antworten protokolliert, die die Probanden über Eignung und Gefallen der gerade ausgeführten Moduskombination eingaben. TODO(kamen die smileys schon? wenn nich kannst du daruber noch schreiben) Nach jedem Durchgang erscheint wieder der Einstellungsscreen, indem der Versuchsleiter das nächste Anwendungsbeispiel laden oder den Messdurchgang aktiviert kann. Für jede Runde werden die entsprechenden Infos vom Setting protokolliert.

Während des Anwendungsbeispiels wird der aktuelle Screen protokolliert, sobald er sichtbar geladen wurde. Jeder Touch, jede erkannte Geste und jeder erkannte Sprachbefehl wird ebenfalls protokolliert. Außerdem wird bei Listen die aktuelle Seite und beim Slider die Start- und Endwerte während einer Bewegung durch Touch oder Geste protokolliert. Bei jedem Button wird der Name des Buttons protokolliert, sobald der Button ausgelöst wurde. Mit diesen Logzeiten sollen nach der Studie die Zeiten der Operatoren berechnet werden. Dazu mehr in Kapitel 5.



Abbildung 4.2.4: Einstellungen vor jedem Durchlauf

Kapitel 5

Studie zur Erhebung multimodaler Interaktionszeiten und Auswertung

In diesem Kapitel wird das Studiendesign und die Durchführung der Studie beschrieben, sowie auf die anschließende Auswertung eingegangen.

5.1 Studiendesign zur Erhebung multimodaler Interaktionszeiten

Für die Studie verwenden wir ein within-subject Design, bei dem jeder Proband alle Kombinationen der vier Anwendungsbeispiele im stehenden Auto durchführen soll.

5.1.1 Permutation der Anwendungsbeispiele

Jedes Anwendungsbeispiel verfügt über eine Moduswechsel. Durch die drei verschiedenen Modalitäten (Touch, Gestik und Sprache) kommen wir somit auf neun Kombinationen pro Anwendungsbeispiel. Eine Ausnahme ist das Navigationsanwendungsbeispiel, bei dem wir uns entschieden haben die Zieleingabe nicht für die Gestik umzusetzen. Somit fallen die Varianten Geste & Geste, Sprache & Geste und Touch & Geste weg. Damit kommen wir bei diesem Anwendungsbeispiel lediglich auf sechs multimodale Kombinationen. Zu den multimodalen Varianten testen wir auch die drei unimodalen Varianten der vier Anwendungsbeispiele und kommen insgesamt auf $(9 * 4 - 3 =) 33$ Varianten für alle vier Anwendungsbeispiele, die jeder Proband durchzuführen hat.

Um Lerneffekte zu vermeiden wurden die vier Anwendungsbeispiele mittels Balanced Latin Square permutiert (vier Permutationen). Die drei unimodalen Kombinationen wurden mittels Latin Square permutiert (drei Permutationen) und die sechs multimodalen Kombinationen erneut mittels Balanced Latin Square (sechs Permutationen). Die Permutation der Modalitäten wurde für jeden Proband fixiert, das heißt für jedes Anwendungsbeispiel gleich angewendet.

Bei zwölf Probanden ist es möglich die vier verschiedenen Permutationen der Anwendungsbeispiele drei mal zu wiederholen. Die drei unimodalen Kombinationen werden vier mal wiederholt

Anwendungsbeispiele					Unimodale Varianten				Multimodale Varianten				Unimodale Varianten			
S: Sound	N: Navi	T: Touch (00)	G: Geste (11)	S: Sprache (22)	0: Touch	1: Geste	2: Sprache	T: Touch (00)	G: Geste (11)	S: Sprache (22)						
P: Phone	K: Klima															
Proband 1	A	S	P	K	N	A	T	G	S	A	02	01	21	12	20	10
Proband 2	B	P	N	S	K	B	G	S	T	B	01	12	02	10	21	20
Proband 3	C	N	K	P	S	C	S	T	G	C	12	10	01	20	02	21
Proband 4	D	K	S	N	P					D	10	20	12	21	01	02
Proband 5	A	S	P	K	N					E	20	21	10	02	12	01
Proband 6	B	P	N	S	K					F	21	02	20	01	10	12
Proband 7	C	N	K	P	S					A	02	01	21	12	20	10
Proband 8	D	K	S	N	P					B	01	12	02	10	21	20
Proband 9	A	S	P	K	N					C	12	10	01	20	02	21
Proband 10	B	P	N	S	K					D	10	20	12	21	01	02
Proband 11	C	N	K	P	S					E	20	21	10	02	12	01
Proband 12	D	K	S	N	P					F	21	02	20	01	10	12

Abbildung 5.1.1: Permutation der Anwendungsbeispiel, der unimodalen Varianten und der multimodalen Varianten pro zwölf Probanden

und die sechs Permutationen der multimodalen Kombinationen werden zwei mal innerhalb von zwölf Probanden wiederholt. Zudem wird die Reihenfolge von unimodalen und multimodalen alterniert (siehe Abbildung 5.1.1). Dieses Vorgehen wird für die nächsten zwölf Probanden wiederholt. Die drei verschiedenen Zieleingaben werden ebenfalls mit dem Latin Square permutiert und auf die permutierte Moduskombination verteilt. Mit diesen Vorkehrungen können wir davon ausgehen, dass mögliche Lerneffekte gleich verteilt sind.

5.1.2 Fragebögen der Studie

Es wurden zwei verschiedene Fragebögen erstellt (siehe Kapitel 10). Im ersten Fragebogen werden demografischen Angaben wie Name, Alter, Händigkeit und die Muttersprache abgefragt, sowie die bereits gesammelten Erfahrungen im Umgang mit Interaktionen durch Touch, Geste und Sprache. Auf die Frage der Nutzungshäufigkeit wurden vier Antwortmöglichkeiten unterschieden:

- ja, benutze ich regelmäßig
- ja, benutze ich gelegentlich
- ja, aber nur sehr wenig
- nein

Nach jedem durchgeführten Messdurchgang eines Anwendungsbeispiels wurde innerhalb der Anwendung eine Bewertung abgefragt, wie geeignet die Probanden die eben ausgeführte Aufgabe fanden und wie sehr ihnen die Aufgabe gefallen hat. Mit Touch als Eingabemodalität soll eines der fünf Symbole gewählt werden (siehe Abbildung 5.1.2 auf der nächsten Seite).

Der zweite Fragebogen wird nach der Durchführung der Studie ausgefüllt. Darin wird abgefragt, wie geeignet die Modalitäten für die verschiedenen Screentypen sind. Dies wird mit einer Likerts-

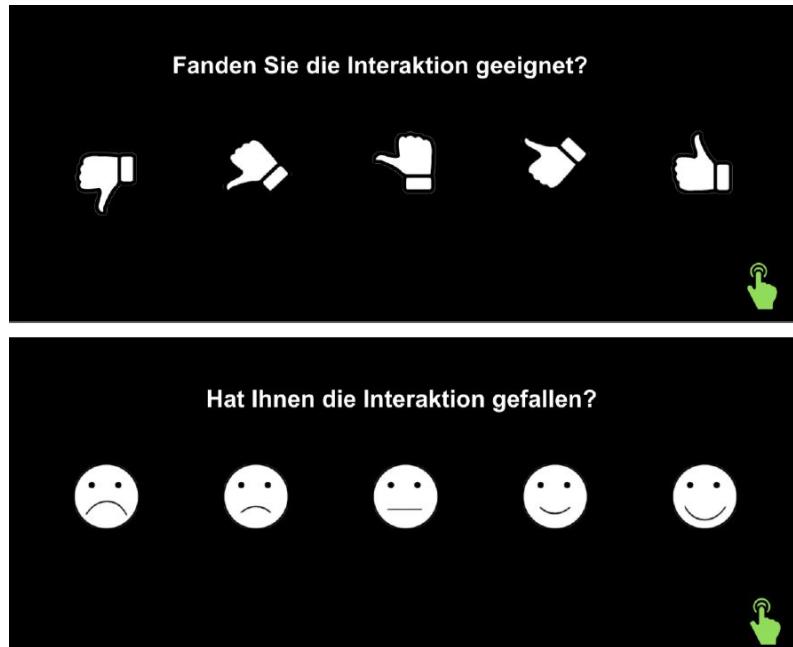


Abbildung 5.1.2: Eignung und Gefallen eines Anwendungsbeispiels. Verwendung der Icons siehe Kapitel 8.

kala mit fünf Auswahlmöglichkeiten für jede Modalität abgefragt. Die Auswahlmöglichkeiten gehen von nicht geeignet bis geeignet. Außerdem wird ein Ranking für die verschiedenen Screentypen abgefragt, indem sich der Proband entscheiden muss welcher Modus für die unterschiedlichen Screentypen am besten, zweitbesten und am schlechtesten geeignet ist.

Die kompletten Fragebögen, Einverständniserklärung, sowie der Studienleitfaden sind am Ende dieser Arbeit angehängt (siehe Kapitel 10).

5.2 Durchführung der Studie

Die Studie fand in Garching-Hochbrück in der Parkgarage von BMW in einem 6er Gran Coupé statt (siehe Abbildung 5.2.1 auf der nächsten Seite). Die Studiendauer betrug circa 1,5 Stunden. Das Auto wurde ans Stromnetz angeschlossen, um das Surface zu laden und eine Dauerhafte Stromversorgung für Licht sowie Sitzheizung zu ermöglichen. Das Surface wurde vor dem ursprünglichen Display am Dashboard angebracht. Der Interaktionsbereich befand sich somit auf gleicher Höhe (siehe Abbildung 5.2.2 auf der nächsten Seite), jedoch etwas näher am Fahrer. Dies ist für die Touchbedienung geeigneter. Die Auflösung des Surface musste auf 1240x800 Pixel reduziert werden, damit die Darstellung groß genug war. Die Kacheln auf dem ersten und zweiten Screen hatten somit eine Größe von 3,5 mal 4,5 cm.

Der Leap Motion Controller, zur Erkennung der Gesten, wurde zwischen Gangschaltung und Dashboard angebracht. Somit konnte die Hand bereits am Lenkrad erkannt werden, befand sich allerdings noch nicht im Interaktionsbereich, um Gesten auszuführen. Pro Tag konnten bis zu vier Probanden an der Studie teilnehmen. Bevor der erste Durchgang begann wurde der Akku der GoPro Kamera ausgetauscht und eine leere Speicherkarte eingelegt. Im Kofferraum konnten Ersatzbatterien für die Kamera geladen werden. Nach jedem Probanden wurde der Akku gewechselt

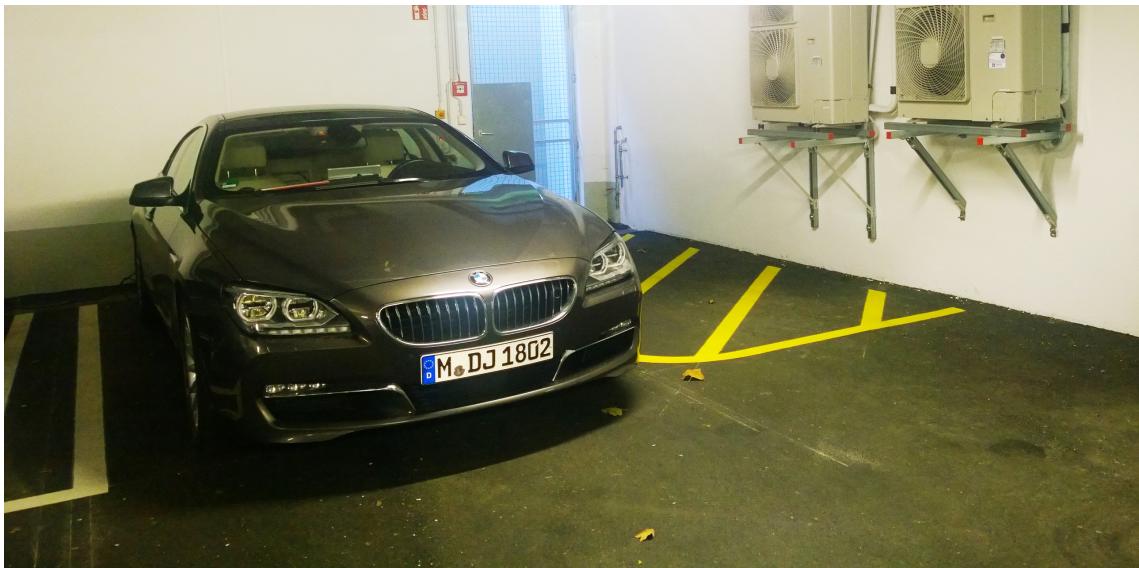


Abbildung 5.2.1: BMW 6er Gran Coupé in der Parkgarage

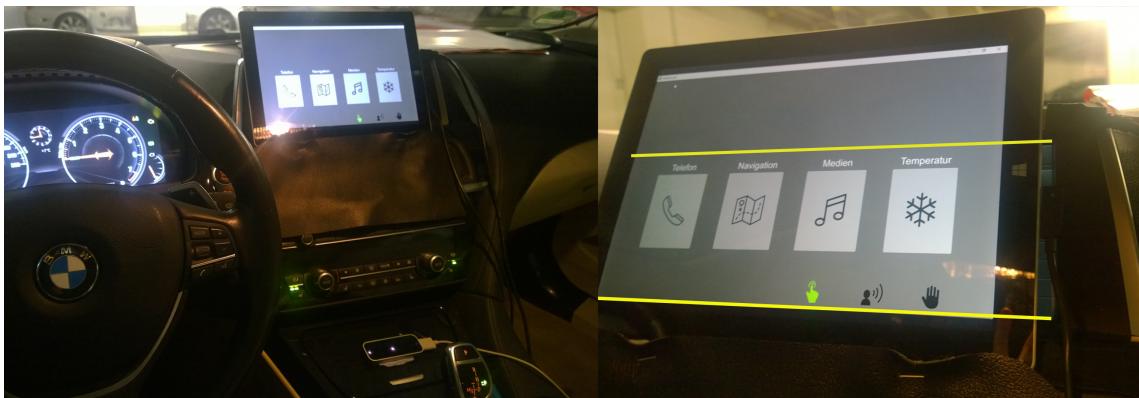


Abbildung 5.2.2: Anbringung des Surface im Auto

und der gebrauchte erneut geladen. Die Speicherkarten von 32GB wurden nach jedem zweiten Probanden gewechselt und anschließend auf eine Festplatte kopiert. Die Kamera wurde mittig, zwischen Fahrer und Beifahrer an der Decke befestigt und Richtung Surface ausgerichtet (siehe Abbildung 5.2.3 auf der nächsten Seite).

Das Surface wurde wie oben beschrieben angebracht und der Leap Motion Controller, sowie die externe Tastatur über einen Hub angeschlossen. Der Verteiler war ebenfalls am Stromnetz angeschlossen und wurde auf der Beifahrerseite zwischen Gurt und Mittelkonsole eingeklemmt siehe Abbildung 5.2.4 auf der nächsten Seite.

Jeder Proband wurde am Empfang am Parkings 19 abgeholt und begrüßt. Es wurde darauf hingewiesen, dass die Studie bis zu 1,5 Stunden dauern wird und das jetzt die beste Gelegenheit wäre, um wenn nötig noch auf die Toilette zu gehen. Anschließend begaben wir uns in die Parkgarage zum Testfahrzeug, indem der Proband auf dem Fahrersitz Platz nehmen sollte, während der Studienleiter am Beifahrersitz Platz nahm. Der Proband wurde darauf hingewiesen seinen Sitz einzustellen, als würde er oder sie das Auto fahren.

Nachdem sichergestellt wurde, dass alle Smartphones auf lautlos gestellt sind, wurde das Thema

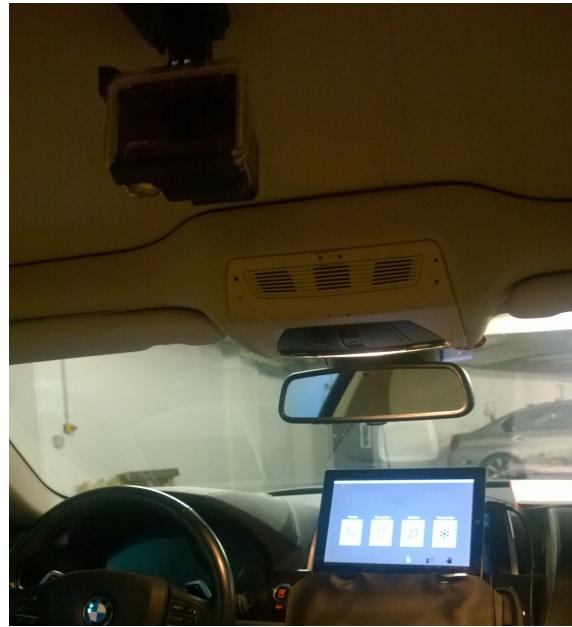


Abbildung 5.2.3: Anbringung der Kamera im Studienauto



Abbildung 5.2.4: Anschluss für die Leap Motion und der externen Tastatur

kurz erläutert und darauf aufmerksam gemacht, dass zur Erhebung der Interaktionszeiten verschiedene Daten protokolliert werden und zusätzlich die Studie mit einer GoPro aufgezeichnet wird. Darüber aufgeklärt musste jeder Proband eine Einverständniserklärung unterschreiben (siehe Kapitel 10). Zusätzlich sollte ein kurzer Fragebogen zu demografischen Daten und den Vorerfahrungen zum Umgang mit den Interaktionen Touch, Geste und Sprache ausgefüllt werden. In der Zwischenzeit startete der Studienleiter das Programm, stellte die richtige ID ein und überprüfte die neu geladene Permutation.

Jetzt wurde das Vorgehen der Studie anhand des ersten Anwendungsbeispiel mit Hilfe der ausgedruckten Übersicht der Anwendungsbeispiele 4.1.1 erläutert. Sobald das erste Anwendungsbeispiel und die Vorgehensweise klar war, wurde die GoPro gestartet und der Proband konnte mit dem ersten Probendurchlauf beginnen. Dazu hatte der Proband seine Hände am Lenkrad und der Studienleiter startete die Anwendung mit einer externen Tastatur. Der Screen wird nach dem Start zuerst für drei Sekunden schwarz bis dann der Hauptscreen zu sehen ist und die Interaktion beginnen kann. Insgesamt wurden 33 verschiedene Durchläufe von jedem Proband getestet. Zu jedem dieser Varianten, bestehend aus den vier Anwendungsbeispielen und deren Moduskombinationen, sollte mindestens ein Probendurchlauf gemacht werden.

Der Probedurchlauf diente dazu den Probanden mit der Aufgabe vertraut zu machen und sich an die möglicherweise neue Interaktionsmethode zu gewöhnen. Die Probedurchläufe wurden so oft wiederholt, bis die Aufgabe klar verstanden und fehlerfrei ausgeführt wurde.

Nach dem Probedurchlauf wurden zwei Messdurchgänge durchgeführt, die für die Erhebung der Interaktionszeiten verwendet werden. Da wir nur fehlerfreie Messdurchgänge benötigen, wurde im Falle eines Fehlers in einem Messdurchgang, dieser wiederholt. Passierte dies kennzeichnete der Studienleiter dies in seinen Notizen zu dieser Variante, um später nur die gültigen Messdurchgänge zu verwenden. Auch interessante Anmerkungen oder Auffälligkeiten wurden während der Studie notiert.

Nach den mindestens 99 Durchgängen ($((1 * \text{Probedurchgang} + 2 * \text{Messdurchgänge}) * 33\text{Varianten})$) konnte die Kamera gestoppt werden und die Probanden sollten den anschließenden Fragebogen ausfüllen. Währenddessen konnten Anmerkungen vom Versuchsleiter notiert werden. Als Dankeschön bekamen die Probanden nach der Studie Schokolade.

Nach jedem Studiendurchgang wurde der Akku der Kamera gewechselt. Die Speicherkarte wurde nach jedem zweiten Probanden gewechselt.

5.3 Quantitative Auswertung der Studienergebnisse

Im Folgenden werden die quantitativen Ergebnisse der Studie präsentiert. In einem Zeitraum von zwei Wochen nahmen insgesamt 22 Probanden an der Studie zur Erhebung der Interaktionszeiten teil. hierbei wurden über 22 Stunden Videomaterial aufgezeichnet.

5.3.1 Studienteilnehmer

Das Alter der 22 Studienteilnehmer (14 männlich und 8 weiblich) beträgt im Durchschnitt 30,55 Jahre. Der Altersbereich erstreckt sich von 22 bis 58 Jahren. Mode und Median ergaben je 25 Jahre. Neun der 22 Probanden sind rechtshändig, zwei linkshändig und ein Proband gab an beidhändig zu sein. Alle Teilnehmer sprechen deutsch als Muttersprache.

Die Vorerfahrungen mit der Bedienung von Touch, Gesten und Sprache der Teilnehmer wurde von den Probanden als Selbsteinschätzung abgefragt. Auf die Frage, ob sie Erfahrung mit der Bedienung von Sprache, Touch und Geste haben, konnte aus vier Optionen gewählt werden (1: nein, 2: ja, aber nur sehr wenig, 3: ja, benutze ich gelegentlich und 4: ja, benutze ich regelmäßig).

Die Vorerfahrung von Sprache ergab im Durchschnitt 2,32, bei Touch 3,95 und bei Geste 1,86. Die genauen Angaben der Probanden können aus Abbildung 5.3.1 auf der nächsten Seite entnommen werden.

5.3.2 User Experience

Nach jedem Messdurchgang wurden die Nutzer gefragt wie geeignet sie diese Interaktion fanden und ob sie Ihnen gefallen hat. Wie bei einer Likert-Skala werden 5 Antwortoptionen unterschieden.

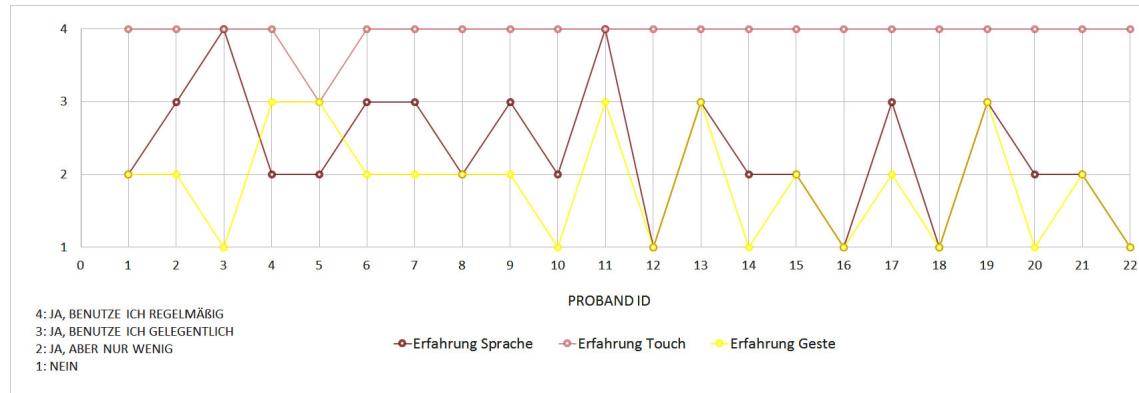


Abbildung 5.3.1: Vorerfahrung der Probanden mit der Bedienung von Touch, Sprache und Geste. Einschätzung der Probanden auf die Frage: Haben Sie Erfahrung bei der Bedienung von Sprach-, Touch- und Gestensteuerung.

Zwei negative, eine neutrale und zwei positive (siehe Abbildung 5.3.2). Aus den Antworten ist deutlich zu sehen, dass die unimodale Variante mit Sprache in beiden Bereichen stets am besten abschnitt. Die Unterschiede zwischen Eignung und Gefallen sind nicht sehr groß, allerdings fällt auf, dass vor allem die Interaktionen mit Geste den Nutzern besser gefällt als sie deren Eignung einschätzen. Am zweit beliebtesten in beiden Kategorien war die Kombination von Touch und Sprache. Bei dem Anwendungsbeispiel Lautstärke war die Kombination Touch und Sprache sogar in Eignung und Gefallen noch etwas besser als die unimodale Sprachvariante. Die dritt beliebteste Variante im Anwendungsbeispiel Lautstärke war die Kombination Sprache und Touch. Bei allen anderen Anwendungsbeispielen war die dritt beliebteste Kombination Geste und Sprache.

Grundsätzlich lässt sich sagen, dass die Mehrheit der Probanden die Sprache am geeignetsten empfanden. Nur für die Direktauswahl aus sichtbaren Elementen schnitt Touch etwas besser ab als Sprache. Die Geste hat bei der Direktauswahl die besten Ergebnisse verglichen zu den anderen Operatoren. Ein sehr eindeutiges Ergebnis aus den Fragebögen bekommen wir bei der Texteingabe, für die Sprache eindeutig besser geeignet ist, als per Touch das gewünschte Ziel einzutippen siehe Abbildung 5.3.2.

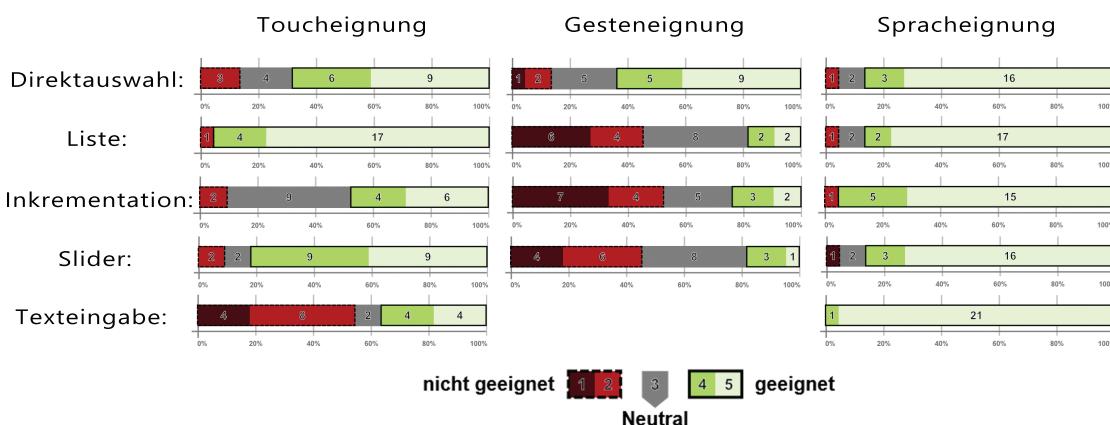


Abbildung 5.3.2: Einschätzung der Probanden, wie geeignet Sie die verschiedenen Screentypen für die jeweilige Modalität halten. Die Balkendiagramme zur Darstellung von Likert-Skalen wurde mit <http://likertplot.com/> generiert.

Im zweiten Teil des Fragebogen sollten die Screenarten noch mit einem Ranking versehen werden. Hierbei mussten sich die Probanden entscheiden welche Modalität am besten, zweitbesten und am schlechtesten geeignet ist. Bei der Direktauswahl aus sichtbaren Elementen halten 10 von 22 die Toucheingabe, 8 die Sprachbedienung und 4 die Gestensteuerung am geeignetensten. Bei den restlichen Screenarten wird die Sprachbedienung am häufigsten für am geeigneten gehalten. Sehr eindeutig sind die Ergebnisse der Texteingabe. 21 von 22 Probanden finden die Spracheingabe für Texteingaben geeigneter, als das Ziel Buchstabe für Buchstabe einzutippen.

5.3.3 Ermittlung aller Zeiten der Aktionen

Um die Zeiten der Operatoren zu ermitteln werden die Differenzen zwischen den gewünschten protokollierten Events berechnet. Dafür musste zuerst die Protokollierung aufbereitet werden. Als erstes wurde die Logdatei in eine Excelltabelle in Tabellenformat geladen. Somit können für die Auswertung nach beliebigen Kriterien die Spalten gefiltert und sortiert werden.

Fehlerhafte Messdurchgänge wurden, anhand der Notizen während der Studie, als fehlerhaft markiert. In einigen wenigen Fällen gab es bei manchen Kombinationen drei statt zwei Messdurchgänge. Hier wurde der dritte Durchgang immer behalten und von den ersten beiden der schlechteste verworfen. Bei Unklarheiten konnte das Videomaterial zur Überprüfung hinzugezogen werden, bis es zu jedem Probanden zwei fehlerfreie Messdurchgänge in jeder Kombination gab.

Für die Berechnungen der verschiedenen Zeiten wurden die protokollierten Events verwendet. Ein Beispiel dafür ist die Zeit, die ein Nutzer benötigt, bis der erste Button gedrückt wird (DA des ersten Screens). Als Startzeit wurde die protokolierte Zeit vom Event „MainUI“ benutzt. Der Screen des Hauptmenüs heißt MainUI. Da unsere Interaktion begann, sobald der Hauptscreen zu sehen war, entspricht das Laden dieser Szene unserem Startpunkt. Dafür wurde eine neue Spalte der Excelltabelle angelegt und immer wenn das protokolierte Event „MainUI“ in der Spalte „Szenen“ gefunden wird, wird die protokolierte Zeit aus der Spalte „ms Gesamt“ in die neue Spalte geschrieben. Wenn die Bedingung nicht erfüllt ist, wird der Wert der Zelle darüber in die Zelle geschrieben. Gibt man folgende Formel in die Spalte ein wird sie automatisch in der kompletten Spalte berechnet.

=IF ([@Szenen] = "MainUI";[@[ms Gesamt]]; AJ59)

Das Ende der Aktion ist das Event „Button geklickt: (Name der 4 Buttons)“. Dafür wurden alle vier Buttons (Telefon, Navigation, Medien und Temperatur) des Hauptscreens verwendet. Um daraus die Dauer einer Aktion zu berechnen wurde die Differenz der Zeitpunkte berechnet und durch 1000 geteilt um die Zeit in Sekunden zu bekommen.

=IF ([@ Endzeit Spalte] - [@ Startzeit Spalte] > 0;
([@ Endzeit Spalte] - [@ Startzeit Spalte]) / 1000; 0)

Mit Berechnungen dieser Art wurden alle Zeiten ermittelt. Die Tabelle kann nach Anwendungsbeispiel, Moduskombination und Event sortiert werden, um die gewünschten Zeiten für die Auswertung zu verwenden. Anzumerken ist, dass diese Art der Berechnung auch Antwortzeiten des Systems beinhaltet. Wir werden uns relevante Größen separat anschauen, um diese später als konstanten Wert im Modell zu berücksichtigen. Zum Beispiel haben wir eine Animation einer scrollenden Liste eingebaut, um für den Nutzer ein besseres Feedback zu gewährleisten. Diese Animationszeit wird separat als Antwortzeit unseres Prototypen betrachtet.

Im folgenden schauen wir uns die Ergebnisse der Studie an. Wir gehen in diesem Abschnitt noch nicht auf signifikante Unterschiede ein, sondern schauen uns zunächst die Ergebnisse und deren Auffälligkeiten näher an.

Die ersten beiden Screens für die Anwendungsbeispiele Telefon, Navigation und Medien und der erste Screen für das Anwendungsbeispiel Temperatur, bestehen je aus einer Direktauswahl aus sichtbaren Elementen (DA). Der Startpunkt der Berechnung war das Laden des Hauptmenüs. Die Nutzer wurden angewiesen mit der Interaktion zu beginnen, sobald sie den Hauptscreen sehen. Das Ende der Direktauswahl aus sichtbaren Elementen war das protokolierte Event das der Button geklickt wurde. Je nach Anwendungsbeispiel (Telefon, Navigation, Medien und Temperatur) musste ein anderer Button selektiert werden. Es wurden alle 4 verschiedene Button einmal verwendet, siehe Abbildung 4.1.1 auf Seite 36.

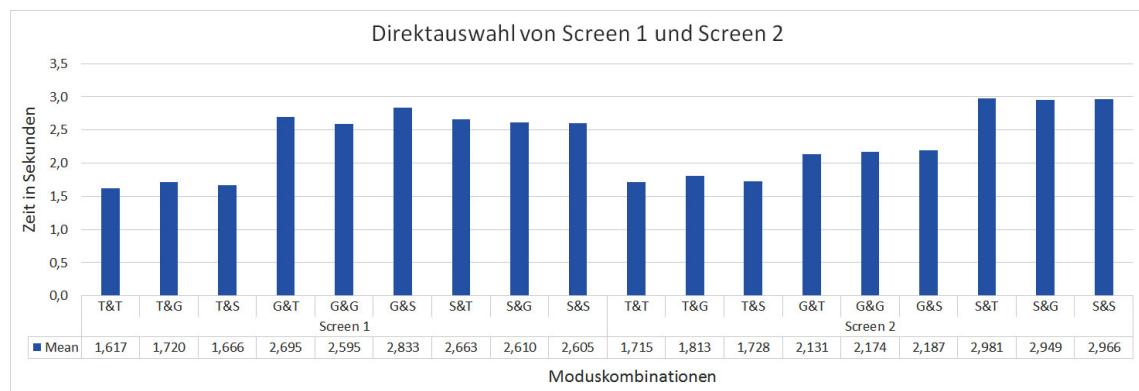


Abbildung 5.3.3: Durchschnittszeiten und Medianwerten in Sekunden der Direktauswahl beider Screens für alle Moduskombinationen. T steht für Touch, S für Sprache und G für Geste.

In Abbildung 5.3.3 werden die Durchschnittszeiten von allen Moduskombinationen der beiden Screens abgebildet. Für die erste Aktion spielt auch nur der erste Modus eine Rolle, da der Moduswechsel erst anschließend statt findet. Die Auswahl der Kachel per Touch ist auf beiden Screens am schnellsten. Im ersten Screen liegen die Interaktionszeiten für Geste und Sprache sehr nah zusammen. Im zweiten Screen dauert die Spracheingabe deutlich länger als per Gestensteuerung. Wie zu erwarten scheint der nachfolgende Modus keinen großen Einfluss auf die Interaktionszeit zu nehmen. Uns interessiert im speziellen der Einfluss nach einem Wechsel.

Als nächstes stellen wir Aktionen vor, die nach einem potentiellen Moduswechsel statt finden. Die Modalität, die in dieser Aktion ausgeführt wird ist somit die jeweils zweite Modalität. Im Falle der unimodalen Variante findet kein Moduswechsel statt. Findet ein Moduswechsel statt, bezeichnen wir im folgenden die zusätzliche Zeit als Wechselkosten. Sie beinhalten einen mentalen Operator und den Homing Operator, da bei einem Moduswechsel zu Geste oder Touch die Position der Hand geändert wird.

Die erste Aktion, die wir uns nach einem Moduswechsel ansehen, ist die Aktion Listen-Navigation (L). Die Probanden mussten im Anwendungsbeispiel Telefon die Liste um drei Seiten inkrementieren und schließlich den Kontakt „Maria Müller“ durch eine Direktauswahl anrufen. Der erste Swipe beinhaltet die Wechselkosten, weswegen dieser auch deutlich länger dauert, als die anderen beiden Swipes siehe Abbildung 5.3.4 auf der nächsten Seite. Der Startpunkt des ersten Swipes ist der Zeitpunkt, sobald der Screen mit der Liste geladen wurde. Wurde die Liste um eine Seite gescrollt, ist dies die Entzeit des ersten Swipes und gleichzeitig die Startzeit des zweiten Swipes. Das selbe gilt für den dritten Swipe.

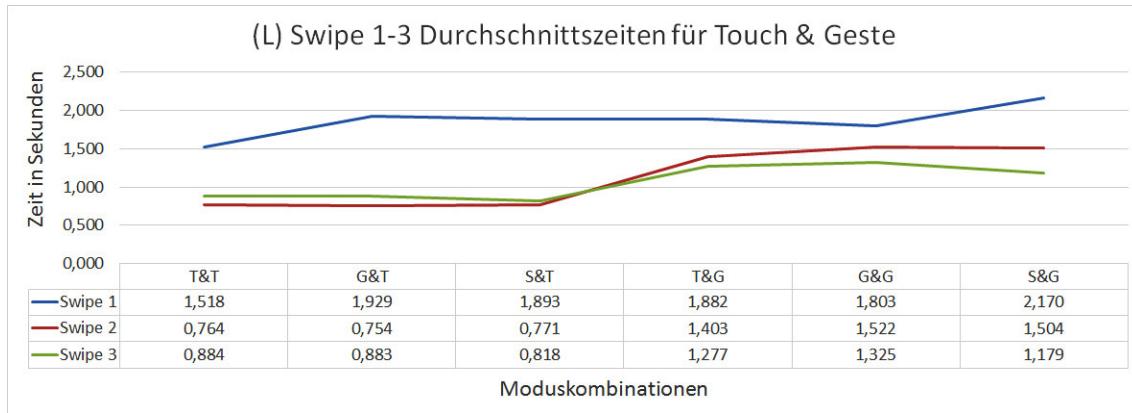


Abbildung 5.3.4: Durchschnittszeiten für Swipe 1, 2 und 3 in Sekunden der Liste für die Moduskombinationen von Touch und Geste

Es fällt auf, dass beim ersten Swipe die unimodalen Varianten schneller sind als die multimodalen Varianten. Der Wechsel einer Modalität benötigt etwas mehr Zeit, da der mentale Operator wahrscheinlich etwas höher ist. Des Weiteren fällt die Umpositionierung der Hand weg.

Anschließend folgt die Aktion Direktauswahl innerhalb der Liste von „Maria Müller“, siehe Abbildung 5.3.5. Hier ist eine deutlich kürzere Zeit zu beobachten als bei Direktauswahl des ersten Screens, da sich die Hand des Nutzer bereits im richtigen Interaktionsbereich befindet. Daher sollten wir diese Aktionen zusätzlich unterscheiden. Die Direktauswahl aus sichtbaren Elementen innerhalb der Liste startet unmittelbar nach dem dritten Swipe und endet mit der Selektion der Kachel „Maria Müller“.

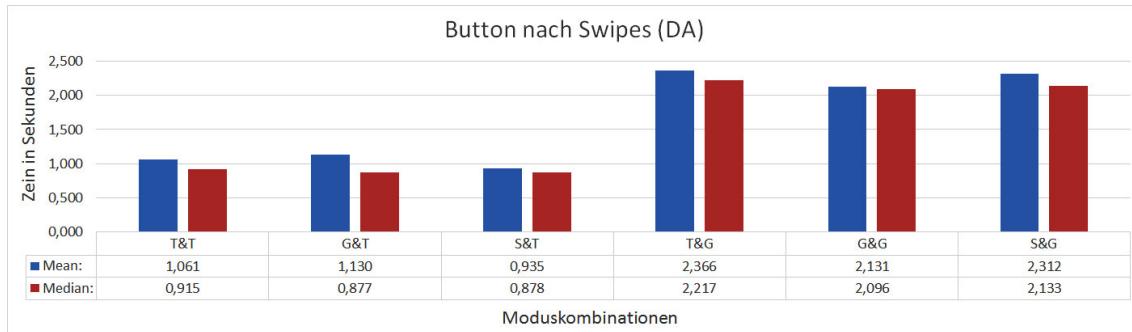


Abbildung 5.3.5: Durchschnittszeiten in Sekunden für die Direktauswahl innerhalb der Listenaktion

Die Touchzeiten der Texteingabe durch Buchstaben, der drei verschiedenen Ziele können aus Abbildung 5.3.6 auf der nächsten Seite entnommen werden. Der Start der Texteingabe beginnt wieder mit dem Laden der Szene des Texteingabescreens. Jeder Buttonklick eines Buchstabens ist die Endzeit des vorherigen Buchstabens und die Startzeit des nächsten Buchstabens. Auch hier ist ein deutliches Muster zu erkennen. Der erste Buchstabe dauert am längsten, da hier erneut die Wechselkosten hinzukommen. Auch bei dieser Aktion ist die unimodale Variante mit 1,765 Sekunden (Touch-Touch) im Durchschnitt schneller als die Multimodalen Varianten mit 1,990 Sekunden (Geste-Touch) und 1,983 Sekunden (Sprache-Touch). Der zweite und dritte Buchstabe ist schon deutlich schneller und alle weiteren Buchstaben nähern sich dem Wert einer halben Sekunde an.

In 5.3.6 lässt sich deutlich erkennen, dass sich der zweite und dritte Buchstaben nicht deutlich un-

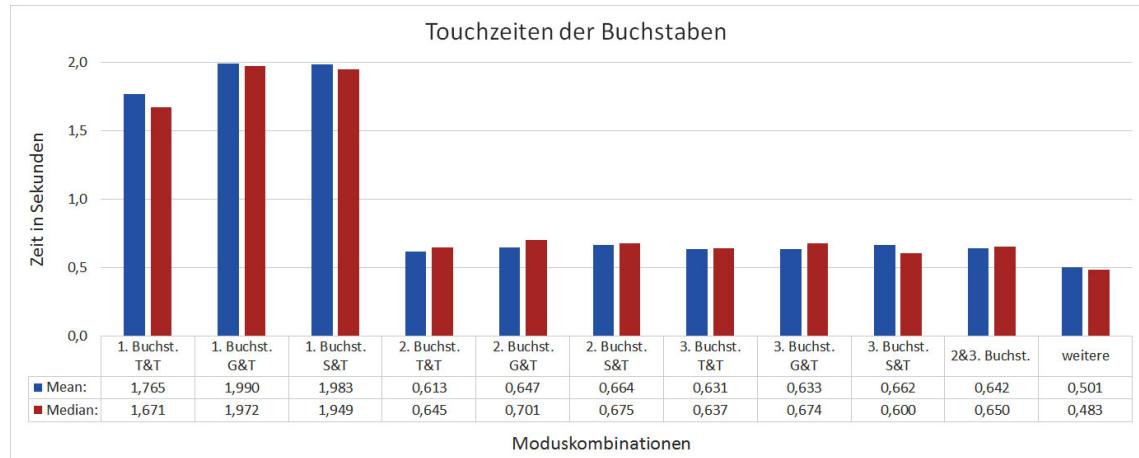


Abbildung 5.3.6: Durchschnittszeiten für Touch in Sekunden der einzelnen Buchstaben

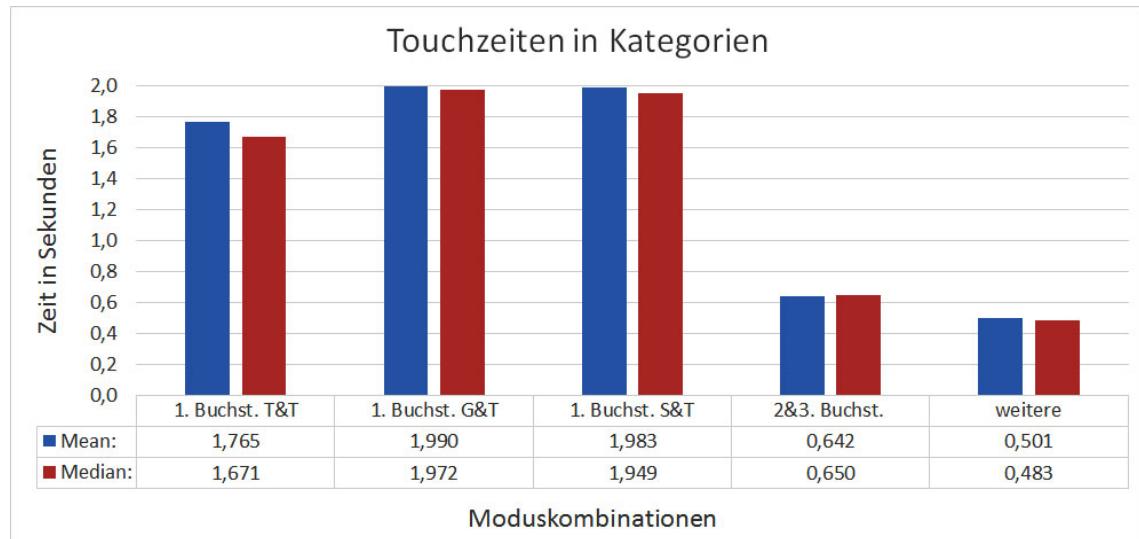


Abbildung 5.3.7: Kategorien der Buchstaben für Wörter. Hier wurden von den drei verschiedenen Zielen („Rom“, „Dorfweg“ und „Kirchengasse“) jeweils die ersten, zweiten und dritten und alle weiteren Buchstaben zusammengefasst.

terscheiden und die restlichen Buchstaben nochmals schneller eingegeben wurden. Wir haben uns deshalb entschlossen beim ersten Buchstaben den vorherigen Modus zu unterscheiden. Den zweiten und dritten Buchstaben zusammenzufassen und für alle weiteren Buchstaben den Durchschnitt dieser zu verwenden. Wir haben somit drei Kategorien. Der erste Buchstabe mit Unterscheidung der vorherigen Modalität, eine gemeinsame Zeit für den zweiten und dritten Buchstaben und als letztes, eine Zeit für alle weiteren Buchstaben [5.3.7](#).

Bei der Sprache ist das Ergebnis weniger deutlich (siehe [5.3.8](#)). Allerdings ist klar zu erkennen, dass „Rom“ am schnellsten eingegeben werden konnte. Zwischen „Dorfweg“ und „Kirchengasse“ ist keine eindeutiger Unterschied zu erkennen. Auch hier ist das Laden der Szene die Startzeit und die Endzeit ist die Zeit, zu der das Inputfeld mit dem Ziel gefüllt wird.

Nach dem Touch und den Spracheingaben der Ziele wird noch die Bestätigung des „Ok“-Buttons gemessen. Die Aktion startet mit dem letzten Buchstaben der Toucheingabe oder dem Aktualisieren des Inputfeld durch die Spracheingabe. Der Moduswechsel fand bereits statt und wie zu erwar-

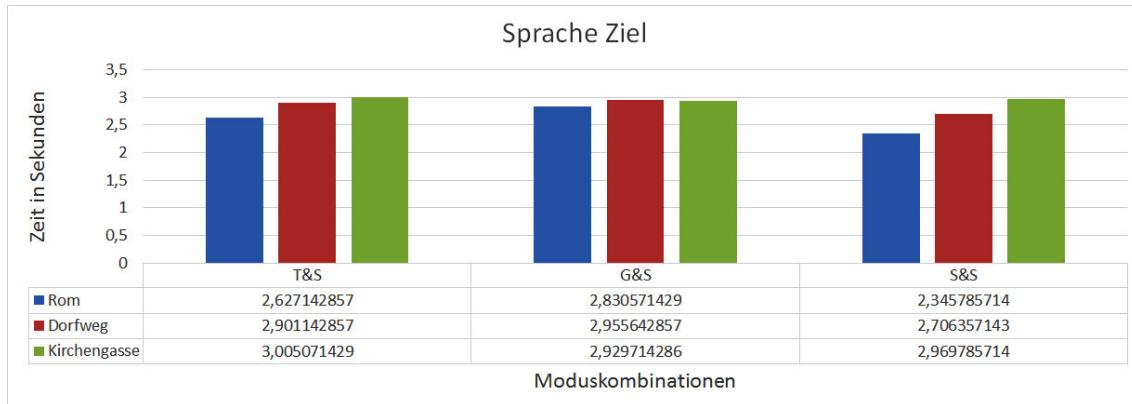


Abbildung 5.3.8: Durchschnittszeiten in Sekunden der Ziele für Sprache

ten sind die Abweichungen zum vorherigen Modus sowohl bei Touch, als auch bei der Spracheingabe sehr gering (siehe 5.3.9).

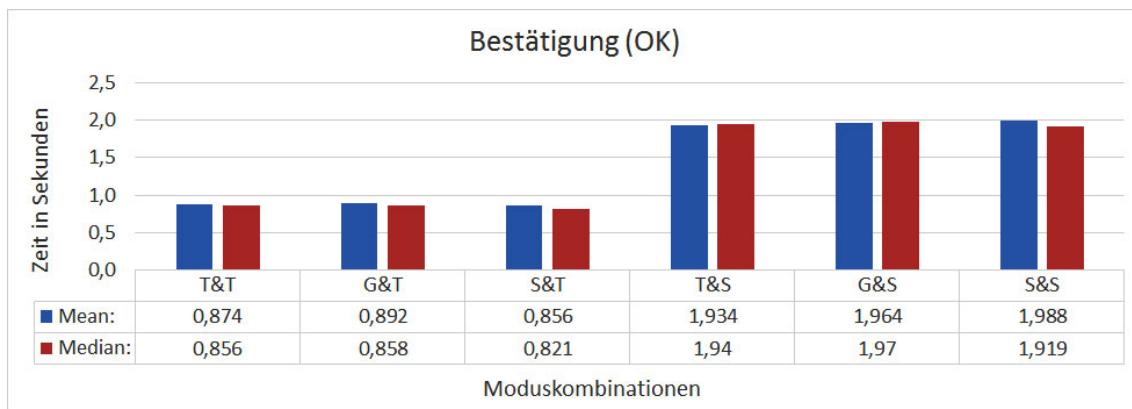


Abbildung 5.3.9: Durchschnittszeiten und Medianwerte in Sekunden für die Bestätigung der Texteingabe.

Im Anwendungsbeispiel Medien folgt im dritten Screen die Aktion direkte Inkrementation (Inkr. (d)) durch den Slider (siehe Abbildung 5.3.10 auf der nächsten Seite) mit einem anschließendem Popup (siehe Abbildung 5.3.11 auf Seite 56). Auch hier ist das Laden der Szene die Startzeit der Aktion. Sobald der Slider im gewünschten Bereich von 75-85% losgelassen wird endet die Aktion und das Popup wird sichtbar. Das Ende der Aktion Inkr. (d) ist der Start der Bestätigungsaktion B. Diese wiederum endet mit der Selektion des Popups. Auch bei den Zeiten für (Inkr. (d)) bestätigt sich erneut, dass die unimodalen Varianten schneller sind als die multimodalen Varianten. Bei dem anschließenden Popup hat der Moduswechsel bereits stattgefunden und es ist zu erkennen, dass die Zeiten hier vom vorherigen Modus wieder unabhängig sind.

Als letzte Aktion bleibt noch die stufenweise Inkrementation Inkr. (s) im Anwendungsbeispiel Temperatur (siehe Abbildung 5.3.12 auf Seite 56). Die Aktion beginnt mit dem Laden der Szene für die Einstellung der Temperatur. Der erste Swipe endet genau wie bei der Listen-Navigation (L) mit dem Erreichen der nächsten Seite. Hier startet gleichzeitig der nächste Swipe, bis wir nach dem dritten Swipe bei der gewünschten Seite angekommen sind. Es folgt eine Verzögerung bis der Wert eingeloggt wird. Diese wird in die Gesamtzeit nicht miteinbezogen, da das Ziel bereits erreicht wurde.

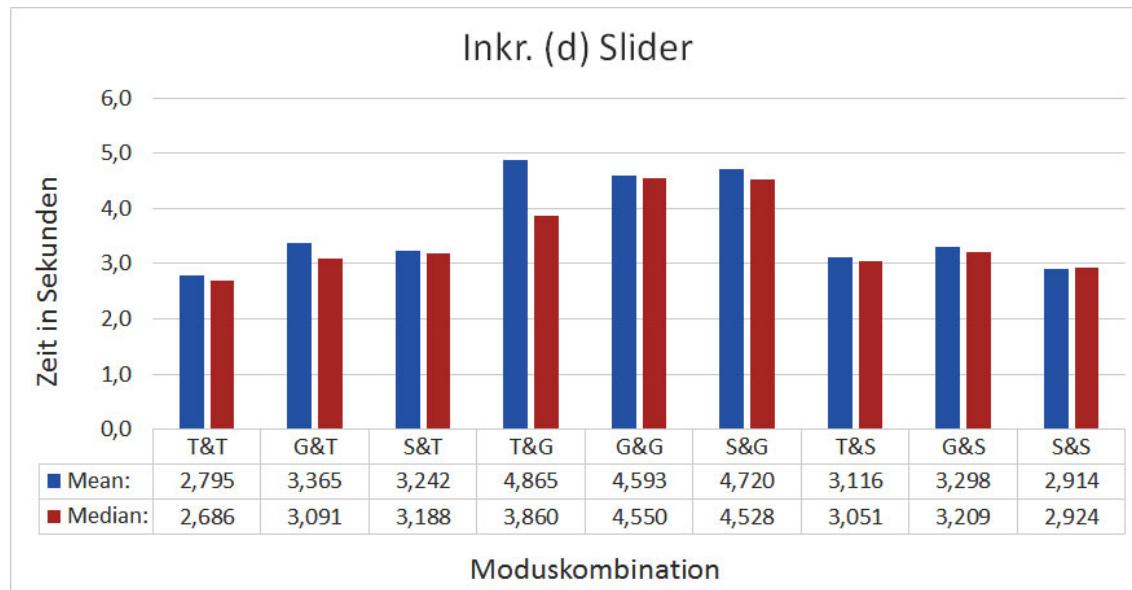


Abbildung 5.3.10: Durchschnittszeiten in Sekunden für die Aktion Slider von 50 auf 75 bis 85 Prozent für alle Moduskombinationen

Auch hier ist deutlich zu sehen, dass der erste Swipe am längsten dauert, Swipe 2 und 3 hingegen sehr ähnlich sind. Beim ersten Swipe sind die unimodalen Varianten ebenfalls schneller als die multimodalen Varianten.

Wir haben jetzt unsere Aktionen ausgewertet und einen Blick auf die verschiedenen Zeiten geworfen. Im Folgenden werden die Zeiten auf statistisch signifikante Unterschiede untersucht.

5.3.4 Statistische Tests der Aktionszeiten

Mit einer One-Way Repeated Measures ANOVA wurde mit Hilfe von SPSS für jede Aktion geprüft, ob der zweite Modus einen signifikanten Einfluss auf den ersten Modus hat und vor allem, ob der erste Modus einen signifikanten Einfluss auf den zweiten Modus hat. Die anhängige Variable war dabei der Modus (Touch, Geste und Sprache) und es wurden die wiederholten Messungen der Zeiten verglichen.

Bei diesem Test muss beachtet werden, dass die Sphärität nicht verletzt wird. [Field and Hole \(2002\)](#) erklärt in seinem Buch „How to Design and Report Experiments“ wie dies in SPSS berücksichtigt wird. Ist bei Mauchly's Test der Sphärität das Ergebnis signifikant wird die Sphärität verletzt und es muss eine Korrektur vorgenommen werden. Wir verwenden bei unseren Ergebnissen für eine Korrektur, die Greenhouse-Geisser Korrektur. Ist bei Mauchly's Test der Sphärität das Ergebnis nicht signifikant, ist die Sphärität auch nicht verletzt und es kann beim Test der Inner-Subjekt Effekten der Wert aus der Zeile entnommen werden, indem die Sphärität angenommen wird ([Field and Hole, 2002](#)).

Diese Auswertung zur Überprüfung von signifikanten Unterschieden wurde für jede Aktion innerhalb ihrer Modalität überprüft. Die Ergebnisse werden im Abschnitt [5.4](#) präsentiert.

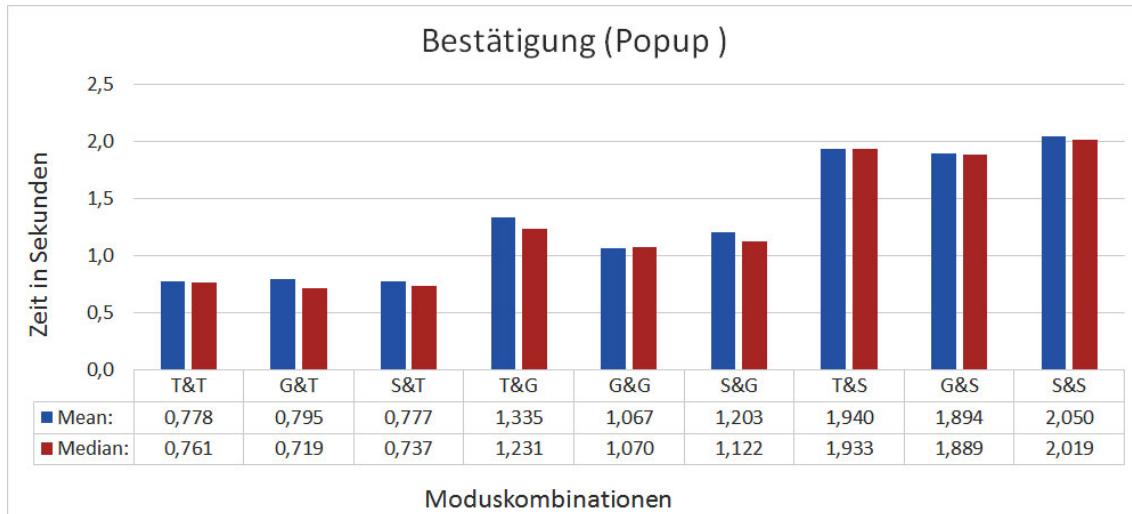


Abbildung 5.3.11: Durchschnittszeiten in Sekunden für die Bestätigung des Popups für alle Moduskombinationen

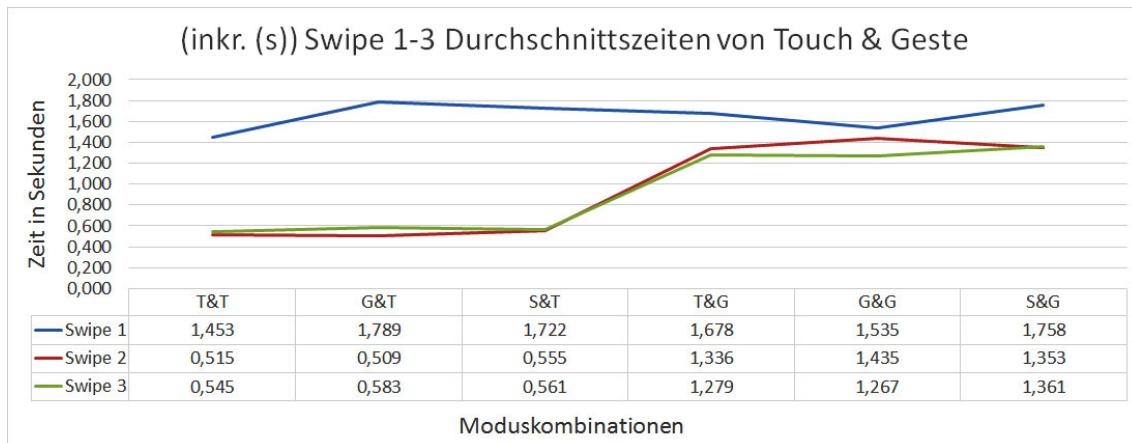


Abbildung 5.3.12: Durchschnittszeiten für Swipe 1, 2 und 3 in Sekunden der Wertinkrementation für die Moduskombinationen von Touch und Geste

5.4 Ableitung des multimodalen Modells aus den gewonnenen Interaktionszeiten

Aus unserer Studie haben wir Zeiten für verschiedene Aktionen in den jeweiligen Moduskombinationen erhoben. Die Zeiten in den folgenden Tabellen sind die Durchschnittszeiten für eine Aktion, der in einer bestimmten Moduskombination ausgeführt wurde. Wird eine Aktion in der Modalität Touch (T) ausgeführt gibt es 3 mögliche Modalitäten, die nach einem Wechsel kommen können. Entweder ebenfalls Touch und somit Touch&Touch (TT) oder es folgt Geste und ergibt Touch&Geste (TG) oder Touch&Sprache (TS), wenn Sprache nach dem Touch folgt. Die gleichen Kombinationen gibt es auch wenn Geste oder Sprache der erste Modus ist.

Aus unseren Anwendungsbeispielen haben sich verschiedene Aktionen ergeben, die wir jetzt analysieren wollen, um daraus unser Modell abzuleiten. In der folgenden Liste werden die einzelnen Aktionen der Anwendungsbeispiele erläutert (siehe auch Anwendungsbeispiele in Abbildung 4.1.1 auf Seite 36):

- **DA₁, DA₂:** Direktauswahl aus sichtbaren Elementen des ersten beziehungsweise zweiten Screens. Dieser beinhaltet beim ersten Screen 4 Buttons, die durch Symbole und Titel gekennzeichnet sind und beim zweiten 3 Buttons, die durch Text auf den Buttons gekennzeichnet sind.
- **L₁, L₂ und L₃:** Der erste (zweite oder dritte) Swipe der Liste aus dem Anwendungsbeispiel Telefon.
- **DA (L):** Direktauswahl aus sichtbaren Elementen, nachdem im Anwendungsbeispiel Telefon bereits drei mal geswiped wurde und nur noch Maria Müller aus vier Elementen ausgewählt werden muss.
- **Inkr. (s)₁, Inkr. (s)₂ und Inkr. (s)₃:** Der erste (zweite oder dritte) Swipe der Wertinkrementation aus dem Anwendungsbeispiel Temperatur.
- **Inkr. (d)** Verschiebung des Sliders im Anwendungsbeispiel Medien. Beim Modus Sprache entspricht das dem Sprachbefehl „achtzig Prozent“.
- **Popup:** Bestätigung des Popup, das nach dem Slider angezeigt wird.
- **1. Bu., 2. Bu., 3. Bu.:** Der Touch des ersten (zweiten oder dritten) Buchstabens im Anwendungsbeispiel Navigation.
- **Bu. >3:** Alle weiteren Buchstaben im Anwendungsbeispiel Navigation.
- **Maria M.:** Der Sprachbefehl „Maria Müller“, der den Kontakt aus einer Liste direkt auswählt. Die Animationszeit, der scrollenden Liste wird nicht berücksichtigt (siehe R(Swipe)).
- **80 %:** Der Sprachbefehl „80%“, um den Lautstärkelider zu verstellen.
- **Wort S, M, L:** Kurzes Wort (Rom), mittellanges Wort (Dorfweg) und langes Wort (Kirchengasse).
- **OK:** Bestätigung der Zieleingabe im Anwendungsbeispiel Navigation.
- **R(Screen):** Konstante Zeit, die vom System benötigt wird von einem Screen auf den nächsten zu wechseln. Ein Wechsel dauert 0,016 Sekunden pro Screenwechsel.
- **R(Swipe):** Beim Modus Sprache wird nach dem Sprachbefehl: „Maria Müller“ die Liste mit einer Animation zur richtigen Seite geswiped bevor der gewünschte Kontakt ausgewählt wird. Die Zeit dieser Animation nennen wir R(Swipe) und dauert mit unserem Prototypen 1,5 Sekunden.
- **R(DA):** Beim Modus Sprache wurde bei der DA nach jedem Sprachbefehl 0,5 Sekunden gewartet um die gelbe Markierung als Feedback zu sehen. Diese 0,5 Sekunden müssen also bei DA₁ und DA₂ für die Gesamtzeit hinzugefügt werden.

Diese Einzelaktionen wollen wir in diesem Kapitel im Hinblick der Modalitätsunterschiede untersuchen und diese so weit es sinnvoll ist aggregieren, um ein übertragbares Modell zu bekommen.

Für die Sprachbefehle modellieren wir die Zeiten von verschiedenen Wortlängen.

Für uns am interessantesten in diesem Modell ist, ob ein vorheriger Modus Einfluss auf den nachfolgenden Modus hat und ob und wie sich die Interaktionszeiten dadurch unterscheiden. Die Zeiten

	Touch (T)			Geste (G)			Sprache (S)		
Aktion	TT	TG	TS	GT	GG	GS	ST	SG	SS
DA ₁	1,617	1,720	1,666	2,695	2,595 (GS)*	2,833 (GG)*	2,663	2,610	2,605
DA ₂	1,715	1,813 (TS)**	1,728 (TG)**	2,131	2,174	2,187	2,981	2,949	2,966

Tabelle 5.4.1: Durchschnittszeiten in Sekunden der Aktionen vor dem Moduswechsel. Signifikanz wurde gemessen innerhalb einer Modalität vor dem Wechsel (zum Beispiel TT, TG und TS). Fettgedruckte Zeiten sind signifikant. Unter der Zahl steht zu welcher Kombination diese Zeit signifikant ist. Dabei gilt zusätzlich: * Signifikanz $\leq 0,05$, ** Signifikanz $\leq 0,01$ und *** Signifikanz $\leq 0,001$

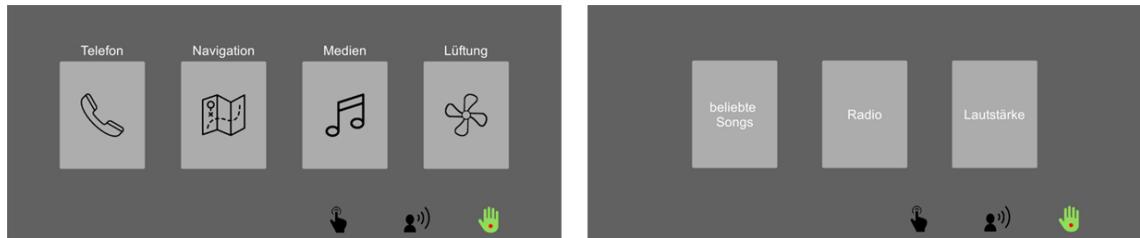


Abbildung 5.4.1: Unterschiedliche Varianten einer Direktauswahl

wurden als Durchschnitt der Zeiten der 22 Probanden ermittelt. Die Durchschnittszeiten aller Aktionen sind aus der Tabelle 5.4.1, Tabelle 5.4.2 auf der nächsten Seite und Tabelle 5.4.3 auf Seite 60 zu entnehmen.

Um das Modell zu verallgemeinern, fassen wir einige der Zeiten zusammen. Alle Aktionen, die nicht unmittelbar nach einem Moduswechsel stattfinden sollen zusammengefasst werden. Diese Aktionen unmittelbar nach einem Moduswechsel sind L₁ (der erste Swipe aus den Kontakten), Inkr. (s)₁ (der erste Swipe, um die Temperatur zu inkrementieren), Inkr. (d) (Einstellung des Sliders) und 1. Bu. (der erste Touch des ersten Buchstabens). All diese Aktionen weisen auch signifikante Unterschiede untereinander auf (siehe Tabelle 5.4.2 auf der nächsten Seite, Tabelle 5.4.3 auf Seite 60). Die Signifikanz wurde gemessen innerhalb einer Modalität nach einem potentiellen Wechsel. Das heißt, die Interaktionszeiten der Modalität Sprache werden auf signifikante Unterschiede zum vorherigen Modus untersucht (TS, GS und SS). Fettgedruckte Zeiten sind signifikant. Unter der Interaktionszeit steht, zu welcher Kombination diese Zeit signifikant ist. Dabei gilt zusätzlich: * Signifikanz $\leq 0,05$, ** Signifikanz $\leq 0,01$ und *** Signifikanz $\leq 0,001$.

Als ersten fassen wir die Zeiten der direkten Auswahl (DA) zusammen, die vor einem Moduswechsel stattfinden und somit keinen Einfluss eines Wechsel haben können. Die Aktion DA₁ weist signifikante Unterschiede der unimodalen Variante Geste (GG) zur multimodalen Variante Geste und Sprache (GS) auf. Im zweiten Screen gibt es ebenfalls einen signifikanten Unterschied in der Aktion DA₂ für Touch zwischen TS und TG (siehe Tabelle 5.4.1). Wir werden die Interaktionszeiten trotzdem für jede Modalität zusammenfassen, da wir den möglichen Einfluss eines bevorstehenden Moduswechsels nicht berücksichtigen wollen. Jetzt haben wir noch sechs verschiedene Durchschnittszeiten (je Touch, Geste und Sprache für DA₁ und DA₂).

Die Zeiten von DA₁ und DA₂ unterscheiden sich, jedoch ist es schwer zu begründen an welchen Variablen dieser Unterschied liegt. Die Größe der Buttons ist zwar gleich, jedoch ist die Anzahl der Buttons unterschiedlich. Es gibt auf dem zweiten Screen nur 3 Buttons, statt der 4 Buttons im

Aktion	Touch (T)			Geste (G)		
	TT	GT	ST	TG	GG	SG
L ₁	1,518 (GT, ST)***	1,929 (TT)***	1,893 (TT)***	1,902	1,797 (SG)**	2,190 (GG)**
L ₂	0,764	0,758	0,771	1,403	1,529	1,504
L ₃	0,884	0,883	0,818	1,271	1,336	1,191
DA (L)	1,056	1,141	0,945	2,366	2,118	2,291
Inkr. (s) ₁	1,453 (GT, ST)***	1,789 (TT)***	1,726 (TT)***	1,672	1,536 (SG)***	1,745 (GG)***
Inkr. (s) ₂	0,523	0,509	0,555	1,336	1,435	1,350
Inkr. (s) ₃	0,552	0,583	0,561	1,279	1,266	1,361
Inkr. (d)	2,795 (GT, ST)**	3,365 (TT)**	3,242 (TT)**	4,865	4,593	4,720
Popup	0,778	0,795	0,777	1,335 (GG)*	1,067 (TG)*	
1. Bu.	1,765 (GT)*, (ST)**	1,990 (TT)*	1,983 (TT)**			
2. Bu.	0,613	0,647	0,664			
3. Bu.	0,631	0,633	0,662			
Bu. >3		0,501				
OK	0,838	0,892	0,856			

Tabelle 5.4.2: Durchschnittszeiten von Touch und Geste in Sekunden der Aktionen nach dem Moduswechsel. Signifikanz wurde gemessen innerhalb einer Modalität nach dem Wechsel (zum Beispiel TT, GT und ST). Fettgedruckte Zeiten sind signifikant. Unter der Zahl steht zu welcher Kombination diese Zeit signifikant ist. Dabei gilt zusätzlich: * Signifikanz $\leq 0,05$, ** Signifikanz $\leq 0,01$ und *** Signifikanz $\leq 0,001$

ersten Screen. Außerdem sind im Hauptscreen Icons und Texte dargestellt und auf dem zweiten Screen lediglich Text abgebildet (siehe Abbildung 5.4.1 auf der vorherigen Seite). Wir wollen eine einheitliche Durchschnittszeit der drei Modalitäten für diese Aktion Direktauswahl aus sichtbaren Elementen (DA). Also die durchschnittliche Zeit, die für eine Auswahl von 3-4 Buttons der Größe 3,5cm x 4,5cm mit Text oder Icons benötigt wird. Deshalb haben wir uns entschlossen die Zeiten zusammenzufassen (siehe Tabelle 5.4.4 auf Seite 61).

Als nächstes fassen wir Aktionen zusammen, bei denen der Moduswechsel bereits stattgefunden hat. Der erste Swipe beispielsweise ist unmittelbar nach einem potenziellen Moduswechsel. Ein Moduswechsel beinhaltet meist einen Homing Operator wie aus dem ursprünglichen KLM. Wechselt die Modalität zum Beispiel von Sprache zu Touch, muss die Hand vom Lenkrad genommen werden und sich zum Touchbereich bewegen. Nach Green (2003) würde das dem Operator R_f (Reach Far) Operator entsprechen. Eine Aktion unmittelbar nach einem Moduswechsel dauert also länger, was sich aus unseren ermittelten Zeiten auch bestätigen lässt.

Deshalb fassen wir für den zweiten und dritten Swipe der Liste (L₂), (L₃) die Modalität vor dem Moduswechsel zusammen. Subtrahieren wir die Durchschnittszeit, die für einen zweiten oder dritten Swipe benötigt wird, von (L₁) bekommen wir sozusagen die eine Vorbereitungszeit. Diese Vorbereitungszeit enthält zum Beispiel eine Umpositionierung der Hand und eine mentale Vorbereitung der Aufgabe. Es entsteht also die einheitliche Aktion (L) für alle seitenweise Swipes und zusätzlich eine Vorbereitungszeit, die immer zum ersten Swipe addiert werden muss. Die Vorbereitungszeit unterscheiden wir für jede Modalität und somit entsteht die Aktion V(L), siehe

Aktion	Sprache (S)		
	TS	GS	SS
Maria M.	2,961 (GS)*, (SS)**	3,112 (TS)*, (SS)***	2,802 (TS)**, (GS)***
80 %	3,126 (SS)**	3,298 (SS)***	2,914 (TS)**, (GS)***
Popup	1,941	1,906 (SS)**	2,100 (GS)**
Wort S	2,627 (SS)*	2,785 (SS)***	2,346 (TS)*, (GS)***
Wort M	2,829	2,956 (SS)*	2,706 (GS)*
Wort L	3,055	2,930	2,970
OK	1,933	1,964	1,988

Tabelle 5.4.3: Durchschnittszeiten von Sprachbefehlen in Sekunden der Aktionen nach dem Moduswechsel. Signifikanz wurde gemessen innerhalb einer Modalität nach dem Wechsel (zum Beispiel TT, GT und ST). Fettgedruckte Zeiten sind signifikant. Unter der Zahl steht zu welcher Kombination diese Zeit signifikant ist. Dabei gilt zusätzlich: * Signifikanz $\leq 0,05$, ** Signifikanz $\leq 0,01$ und *** Signifikanz $\leq 0,001$

Tabelle [5.4.5](#) auf der nächsten Seite.

Das gleiche Vorgehen wenden wir für die Aktion Inkr. (s) an. Auch hier ist der erste Swipe im Schnitt ca. eine Sekunde langsamer als die folgenden. Wir wollen eine einheitliche Aktion Inkr. (s) und zusätzlich eine Vorbereitungszeit V (Inkr. (s)), die für die erste Ausführung benötigt wird.

Bei der Eingabe von Buchstaben per Touch unterscheiden wir die Zeiten zusätzlich in 2 Gruppen. Wie wir bereits aus der Abbildung [5.3.7](#) abgeleitet haben gibt es einerseits auch hier eine deutlich längere Zeit für den ersten Buchstaben. Zusätzlich nimmt die Zeit der Toucheingabe ab dem vierten Buchstaben nochmal deutlich ab (0,501 statt 0,642 Sekunden). Wir modellieren zum einen eine Zeit für das Tippen der ersten 3 Buchstaben und addieren für die Vorbereitung des ersten Buchstabens eine feste Zeit V(1. Bu) hinzu, wie schon bei den Aktionen L und Inkr. (s). Für alle weiteren Buchstaben gibt es eine neue geringere Durchschnittszeit, die verwendet werden soll, siehe Tabelle [5.4.5](#) auf der nächsten Seite.

Im nächsten Schritt werden wir unsere Aktionen für die Modalität Sprache noch vereinfachen, um unterschiedlich lange Wörter unterscheiden zu können.

Bei der Aktion Swipe Liste muss für die Modalität Sprache lediglich „Maria Müller“ gesagt werden, um in der Liste zur richtigen Position zu scrollen. Bei der Aktion Inkr. (d) muss „80 Prozent“ gesagt werden, um den Slider auf 80 Prozent zu verschieben. Da beides die Dauer darstellt, die für 2 Wörter benötigt wird, wollen wir diese Zeiten trotz signifikanter Unterschiede zusammenfassen (siehe Tabelle [5.4.6](#) auf Seite [62](#)).

Ähnlich dazu, wollen wir die Zeiten für Touch und Sprache des Bestätigungsdialogs und des Ok Buttons verbinden und diese als einheitliche Aktion Bestätigung (B) zusammen fassen. Sowohl bei Touch, als auch bei Sprache gibt es keine signifikanten Unterschiede zwischen dem Popup und dem Ok Button.

Als letztes fassen wir die Wortlängen M (Dorfweg) und L (Kirchengasse) zusammen. Diese sind beide signifikant unterschiedlich zur Wortlänge S (Rom), aber gegenseitig weisen sie keine signifikanten Unterschiede auf. Die Unterscheidung der Modalitäten wird allerdings erhalten, da unmittelbar davor der Moduswechsel statt gefunden hat. Bei diesen Aktionen wird jetzt lediglich unterschieden, ob sie einsilbig oder mehrsilbig sind. Diese Aktionen kürzen wir ab mit Wort(e) für ein einsilbiges Wort und Wort(m) für ein mehrsilbiges Wort (siehe Tabelle 5.4.6 auf der nächsten Seite).

Aktion	Touch			Geste			Sprache		
	TT	TG	TS	GT	GG	GS	ST	SG	SS
DA	1,710			2,436			2,795		

Tabelle 5.4.4: Zusammengefasste Zeiten der Operatoren vor dem Moduswechsel

Aktion	Touch (T)			Geste (G)		
	TT	GT	ST	TG	GG	SG
V(L)	0,705	1,116	1,080	0,530	0,425	0,818
L		0,813			1,372	
DA(L)		1,047			2,258	
V(Inkr. (s))	0,906 + 0,547	1,242 + 0,547	1,179 + 0,547	0,334 + 1,338	0,198 + 1,338	0,407 + 1,338
Inkr. (s)		0,547			1,338	
Inkr. (d)	2,795 (GT, ST)**	3,365 (TT)**	3,242 (TT)**	4,865	4,593	4,720
B		0,829			1,200	
V(1. Bu.)	1,123 + 0,642	1,348 + 0,642	1,341 + 0,642			
1-3. Bu.		0,642				
Bu. >3		0,501				

Tabelle 5.4.5: Durchschnittszeiten von Touch und Geste in Sekunden der Aktionen nach dem Moduswechsel. Signifikanz wurde gemessen innerhalb einer Modalität nach dem Wechsel (zum Beispiel TT, GT und ST). Fettgedruckte Zeiten sind signifikant. Unter der Zahl steht zu welcher Kombination diese Zeit signifikant ist. Dabei gilt zusätzlich: * Signifikanz $\leq 0,05$, ** Signifikanz $\leq 0,01$ und *** Signifikanz $\leq 0,001$

5.5 Multimodales Modell mit Wechselkosten

In den letzten Schritten haben wir die gesammelten Durchschnittszeiten gruppiert und zusammengefasst. Um die verschiedenen Aktionen mit ihren Interaktionszeiten übersichtlich zu einem Modell zusammenzufassen, betrachten wir ab jetzt die Aktionen vorerst unimodal (TT, GG und SS) und addieren je nach Moduswechsel die dementsprechenden Wechselkosten hinzu.

Die unimodalen Interaktionszeiten waren immer schneller als die multimodalen Interaktionszeiten. Wir verwenden diese als unsere Ausgangszeiten, indem wir sie von den Multimodalen Varianten subtrahieren. Dadurch entstehen dann die sogenannten zusätzlichen Wechselkosten. Zum Beispiel die V(L) dauert für unimodalen Touch (TT) 0,705 Sekunden und für die multimodale Variante Geste und anschließend Touch (GT) 1,116 Sekunden. Ziehen wir die unimodale Zeit von der multimodalen ab, erhalten wir die Wechselkosten von 0,411 Sekunden, die bei einem Wechsel von Geste zu Touch (W_{GT}) anfallen.

Aktion	Sprache (S)		
	TS	GS	SS
2 Wörter	3,043 (GS)*, (SS)**	3,205 (TS)*, (SS)***	2,858 (TS)**, (GS)***
B	1,972		
Wort (e)	2,627 (SS)*	2,785 (SS)***	2,346 (TS)*, (GS)***
Wort (m)	2,917	2,944 (SS)*	2,839 (GS)*

Tabelle 5.4.6: Durchschnittszeiten von Sprachbefehlen in Sekunden der Aktionen nach dem Moduswechsel. Signifikanz wurde gemessen innerhalb einer Modalität nach dem Wechsel (zum Beispiel TT, GT und ST). Fettgedruckte Zeiten sind signifikant. Unter der Zahl steht zu welcher Kombination diese Zeit signifikant ist. Dabei gilt zusätzlich: * Signifikanz $\leq 0,05$, ** Signifikanz $\leq 0,01$ und *** Signifikanz $\leq 0,001$

Wir haben die Aktion (DA) vor einem Moduswechsel für jede Modalität (ergibt 3 Aktionszeiten) und 13 Aktionen nach einem möglichen Moduswechsel für Touch und Geste (ergibt 17 Aktionszeiten) und 4 Aktionen nach einem möglichen Moduswechsel für Sprache identifiziert. Dies ergibt 24 Aktionen und 20 verschiedene Arten von Wechselkosten. Insgesamt kommen wir auf 44 verschiedene Interaktionszeiten für unser Modell.

In Tabelle 5.5.1 sind die zusammengefassten Zeiten der Aktion DA, die vor dem Moduswechsel stattfinden und in Tabelle 5.5.2 sind alle Aktionen mit ihren Interaktionszeiten aufgelistet. Wird eine dieser Aktionen mit einer anderen Modalität ausgeführt als die DA müssen die Wechselkosten addiert werden. Die dazugehörigen Wechselkosten für multimodale Interaktionen können aus Tabelle 5.5.3 auf der nächsten Seite entnommen werden.

Aktion	Touch			Geste			Sprache		
	TT	TG	TS	GT	GG	GS	ST	SG	SS
DA	1,710			2,436			2,795		

Tabelle 5.5.1: Zusammengefasste Zeiten der Operatoren vor dem Moduswechsel

Aus unserem Prototyp ergeben sich noch 3 Antwortzeiten. Einmal der Wechsel zwischen den Screens mit einer Zeit von 0,016 Sekunden. Um dem Nutzer beim Anwendungsbeispiel Telefon Feedback zu geben, wurde bei der Modalität Sprache nach dem Sprachbefehl „Maria Müller“ eine Animation eingebaut, die die Liste scrollen lässt, bevor der Kontakt markiert wird. Diese Animation dauert 1,5 Sekunden. Bei der Spracheingabe zur Direktauswahl aus sichtbaren Elementen wurde 0,5 Sekunden gewartet, um die Markierungsfarbe zu sehen. Solche konstanten Zeiten sind nicht übertragbar, sondern müssen für jedes neue Interface das getestet wird berechnet und in die Vorhersage miteinbezogen werden. Für die Evaluation benötigen wir diese Zeiten.

5.6 Qualitative Auswertung der Studie

Diese Art der Auswertung kann nicht in Zahlen ausgedrückt werden, sondern beinhaltet Kommentare der Probanden, sowie Beobachtungen des Studienleiters während der Studie.

Aktion	Unimodal		
	Touch	Geste	Sprache
V (L_1)	0,705	0,425	
je (L)	0,813	1,372	
DA(L)	1,047	2,258	
V(Inkr. (s))	0,906	0,198	
Inkr. (s)	0,547	1,338	
Inkr. (d)	2,795	4,593	
Bestätigung	0,829	1,200	1,972
V(1. Bu.)	1,123		
1-3. Bu.	0,642		
Bu. >3	0,501		
Wort(e)			2,346
Wort(m)			2,839
2 Wörter			2,858

Tabelle 5.5.2: unimodale Interaktionszeiten.

Aktion	Multimodalmodale Wechselkosten						
	Geste zu Touch	Sprache zu Touch	Touch zu Geste	Sprache zu Geste	Touch zu Sprache	Geste zu Sprache	
L_1	+0,411	+0,375	+0,105	+0,393			
Inkr. (s) ₁	+0,336	+0,273	+0,136	+0,209			
Inkr. (d)	+0,570	+0,447	+0,272	+0,127			
1. Buchstabe	+0,225	+0,218					
Wort(e)					+0,281	+0,439	
Wort(m)					+0,079	+0,105	
2 Wörter					+0,185	+0,347	

Tabelle 5.5.3: Zusatzkosten (Wechselkosten) in Sekunden bei einem Moduswechsel. Diese Wechselkosten betreffen nur Aktionen direkt nach dem Moduswechsel. Wechselkosten müssen bei einem Moduswechsel dementsprechend aufaddiert werden.

Die Studiendauer von 1,5 Stunden war sehr lange. Es konnte beobachtet werden, dass die Konzentration mit der Zeit oft nach ließ. Hinzu kam, dass es Anfang Dezember teils sehr kalt in der Parkgarage war. Je nach Vorlieben konnte die Sitzheizung eingeschaltet werden, jedoch wurden bei einigen mit der Zeit die Hände kalt.

Die Verwendung von Gesten als Eingabemodalität erwies sich als sehr unterschiedlich bei den Probanden. Obwohl einige Schwierigkeiten bei der Gestenerkennung auftauchten, machte es den meisten sehr viel Spaß damit zu interagieren. Es fiel auf, dass Probanden, bei dem die Gesteneraktion sehr gut funktionierte natürlich auch der Eindruck viel positiver war. Ein Grund weshalb die Gestenerkennung unterschiedlich gut funktioniert hat, war der festgelegt Interaktionsbereich zwischen Gangschaltung und Dashboard, siehe Abbildung 5.6.1 auf der nächsten Seite.

Je nach Einstellung des Sitzes und vor allem je nach Armlänge der Probanden befand sich die Hand mehr oder weniger bequem im Interaktionsbereich. Am besten erschien eine Variante, bei dem der Proband seinen Arm auf der Armlehne ablegen konnte. Ist der komplette Arm in der Luft, wird er schnell schwer und die Gesten sind schwieriger auszuführen. Befindet sich die Hand zu weit hinten kommen die Infrarotstrahlen der Leap nicht durch die Gangschaltung und die Gestenerkennung funktioniert nicht so gut oder gar nicht. Es ist wichtig, dass dieser Bereich individuell einstellbar

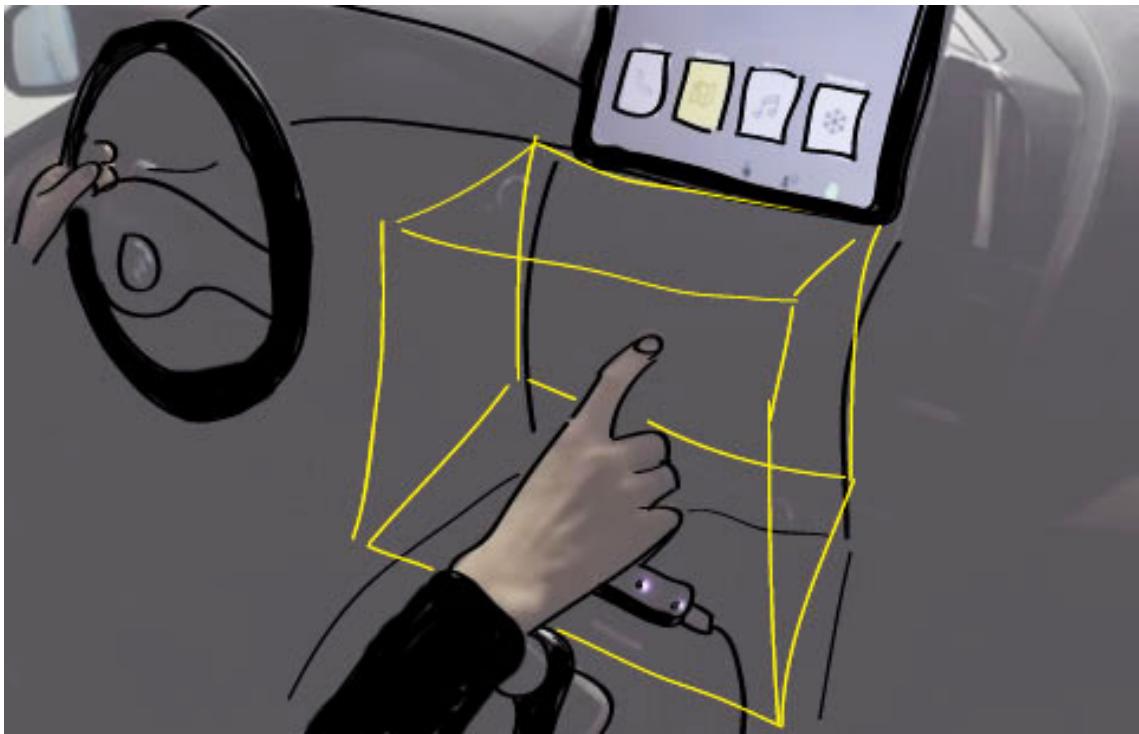


Abbildung 5.6.1: Skizze des Interaktionsbereichs zur Gestenerkennung

sein sollte, um die Gestenerkennung so bequem und robust wie möglich zu gestalten.

Bei der Toucheingabe ist aufgefallen, dass sie zwar keinerlei Schwierigkeiten bei der Interaktion machte, allerdings befand sich der Touchbereich etwas zu weit hinten. Einige der Probanden mussten sich etwas nach vorne lehnen, um ihn optimal zu erreichen. Das stellt für die Bedienung im stehenden Auto kein Problem dar, jedoch während Fahrt sollte das vermieden werden, um möglichst wenig abzulenken. Das Inputfeld im Screen der Texteingabe befand sich unter der Tastatur. Bei der Toucheingabe wurde es durch die Hand verdeckt. Es sollte sich besser über der Tastatur befinden.

Aus Kommentaren einiger Probanden kam heraus, dass sie einerseits die Sprachbedienung für während der Fahrt am geeignetsten halten. Andererseits wurde angemerkt, dass sie die Sprachbedienung eher allein im Auto nutzen würden als mit Mitfahrern. Das deutet wiederum auf die Vorteile von multimodale Interaktion hin, bei der eine beliebige Variante je nach Situation gewählt werden kann.

Die multimodale Ausführung von Interaktionen kam insgesamt sehr gut an. Immer wieder wurde jedoch angemerkt, dass haptische Bedienelemente für die Inkrementation besser geeignet wären.

5.7 Diskussion zur Studie und Erstellung des Modells

Bei der Entwicklung des Prototypen und der Erhebung der Interaktionszeiten aus der Studie gab es einige Einschränkungen und Entscheidungen, die im Zusammenhang der Studienergebnisse im folgenden interpretiert werden sollen.

Als erste Einschränkung haben wir die Interaktion mit klassischen haptischen Bedienelementen aus unserer Studie ausgeschlossen. Dies lag dem großen Mehraufwand zugrunde, der dabei entstanden wäre. Um haptische Bedienelemente befriedigend zu modellieren, müssten, um nur ein Beispiel zu nennen verschiedene Drehwinkel eines Drehrades gemessen werden (Schneegaß, 2009). Allein durch unsere drei Modalitäten mit Touch, Sprache und Geste entstanden mit unseren Aktionen ein Studienumfang von 1,5 Stunden, in der alle Kombinationen der drei Modalitäten getestet wurden. Wir haben uns entschieden uns in dieser Arbeit auf neuere Modalitäten (Touch, Sprache und Geste) zu fokussieren.

Die Bedienung im Auto durch haptische Knöpfe, Schalter und Regler ist und bleibt ein wichtiger Bestandteil, um Funktionen im Auto zu betätigen. Ein großer Vorteil ist, dass die Eingabe oft fast blind getätigert werden kann, ohne die Aufmerksamkeit zu viel von der Straße zu lenken. Jedoch wurde bereits in vorherigen Studien die Verwendung von haptischen Bedienelementen im Auto untersucht (Pettitt et al., 2007), (Schneegaß, 2009) und (Schneegaß et al., 2011).

In unserer Studie zur Erhebung von multimodalen Interaktionszeiten verwenden wir keine Fahrtaufgabe. Stattdessen messen wir die Total Task Time, die in einem stehenden Auto benötigt wird um eine Aufgabe zu lösen (Green, 1999a). Um den Grad der visuellen Ablenkung, solcher multimodalen Interaktionen messen zu können, sollte unser Modell zusätzlich noch dahingehend erweitert werden, die visuelle Ablenkung miteinzuberechnen (siehe (Pettitt et al., 2007)).

Unsere Aktionen des multimodalen Modells sind nicht so feingranular wie die Operatoren aus dem Keystroke-Level-Modell. Statt separate Zeiten für die mentale Vorbereitung, der Bewegung vom Lenkrad zum Interaktionsbereich und der Bewegung innerhalb eines Interaktionsbereichs zu modellieren, haben wir diese als durchschnittliche Vorbereitungszeit für 3 verschiedenen Aktionen je nach Modalität zusammengefasst. Bei der Modellierung des Keystroke-Level-Modell wird genau vorgegeben, wann der Experte welche Bewegung macht und wann er sich mental vorbereitet.

Diese Herangehensweise war uns für unser Modell zu genau und hätte per Videoanalyse ausgewertet werden müssen. Um die Akkudauer der Go Pro zu verlängern, haben wir die Auflösung verringert. Die Bildqualität reicht aus, um die Varianten nachzuvollziehen, für eine Videoanalyse zur Bestimmung von Operatoren ist sie eher ungeeignet. Es wurden über 22 Stunden Videomaterial aufgenommen.

Für unser Modell wollten wir Zeiten, die ein geübter Nutzer im Durchschnitt benötigt, um eine bestimmte Aktion in einem bestimmten Modus auszuführen. Das Modell soll eine grobe Abschätzung der Interaktionszeiten in Abhängigkeit der Modalität darstellen. Das ist auch der Grund warum wir unser Modell bis zum jetzigen Stand aggregiert haben. Doch selbst diese grobe Abschätzung sollte Vergleiche zwischen verschiedenen Moduskombination der Aktionen zulassen. Es ist jedoch zu erwarten, dass die Vorhersage der Interaktionszeiten eventuell nicht ganz so genau sind wie die eines Keystroke-Level Modells.

Um für alle Kombinationen ausreichend Zeiten erheben zu können musste jeder Nutzer im Within Subject Design alle 33 Kombinationen durchführen. Da wir in unserem Modell, wie bei dem Keystroke-Level-Modell (Card et al., 1980), von Experten ausgehen wollen, mussten die Probanden der Studie vorher die Interaktion üben.

Jude et al. (2014) zeigte in seinem Paper zu multimodaler Interaktion von Geste und Sprache, dass eine sehr schnelle Interaktionssteigerung mit relativ geringem Training zu beobachten ist. Außerdem entschlossen wir uns von jedem Probanden zwei Messdurchgänge zu erheben, um somit die Durchschnittszeit beider Messdurchgänge verwenden zu können. Diese Vorkehrungen bedeuteten

allerdings, dass die Studiendauer bis zu 1,5 Stunden dauerte (je nachdem wie viel geübt werden musste). Die Studiendauer von 1,5 Stunden ist sehr lange und es konnte beobachtet werden, dass die Konzentration bei einigen Probanden mit der Zeit abnahm. Durch unsere Permutation aller Modalitäten und Anwendungsbeispielen, sollten diese eventuellen negativen Effekte auf alle Varianten gleichermaßen verteilt sein.

Die Vorerfahrung von Sprache ergab im Durchschnitt 2,32, bei Touch 3,95 und bei Geste 1,86. Die große Erfahrung mit der Touchbedienung lässt sich auf die weitverbreitete Smartphonenuutzung zurückführen. Mit der Gestensteuerung haben die Probanden wie erwartet am wenigsten Erfahrung. Acht der 22 Probanden hatten keine Erfahrungen mit der Gestensteuerung und keiner davon nutzt sie regelmäßig. Diese Erfahrungswerte liegen der wenig verbreiteten Anwendungen mit Gesteninteraktionen zu Grunde.

Für die Sprachmodalität haben wir nicht alle Aktionen abgebildet. Statt zum Beispiel in einer Liste aus Kontakten durch einen Sprachbefehl „nächste Seite“ zu scrollen , erschien es uns geeigneter direkt den Kontakt „Maria Müller“ auszuwählen. Ebenso beim Einstellen der Temperatur oder der Lautstärke wurde der gewünschte Wert direkt, ohne Inkrementation, gesagt.

Bei unserem Modell konnten wir für die Modalität Sprache nur zwischen einsilbigen Wörtern und mehrsilbigen Wörtern, sowie einer Durchschnittszeit von zwei Wörtern unterscheiden. In den meisten gängigen IVIS mit Sprachbedienung können auch ganze Sätze gesprochen werden wie zum Beispiel „navigiere in die Kirchengasse Nummer 17“. Um die Dauer eines Satzes zu modellieren müsste unser Modell noch dahingehend erweitert werden.

Die Zeit der Toucheingabe ab dem vierten Buchstaben nochmal deutlich ab (0,501 statt 0,642 Sekunden). Warum? die Dauer darstellt, die für 2 Wörter benötigt wird, wollen wir diese Zeiten trotz signifikanter Unterschiede zusammenfassen. Warum? Weitere Messungen notwendig? Wir haben mehr Aktionen als erwartet. Warum und warum machen diese Sinn?

Die Aktion Direktauswahl aus sichtbaren Elementen sollte noch erweitert werden, dass sie auch nach einem Wechsel modelliert stattfinden kann. In unseren Anwendungsbeispielen kam diese Variante nicht vor und konnte somit nicht modelliert werden.

Der Gesteninteraktionsbereich war nicht für alle Probanden optimal, je nach Armlänge konnten die Gesten der Probanden besser und schlechter ausgeführt werden. Probanden, bei denen sich der Interaktionsbereich zu weit weg befand, mussten sich mehr darauf konzentrieren die Geste richtig auszuführen, was die Interaktion verlangsamt und die potenzielle Ablenkung vergrößert. In der Studie von [Riener et al. \(2013\)](#) wurde der Bereich für Gesten untersucht. Sie stellten fest, dass die häufigste Gesteninteraktion sich in dem Dreiecks-Bereich von Lenkrad, Rückspiegel und Gangschaltung befindet. Unsere Gestenerkennung befand sich somit innerhalb dieses Bereichs.

Auch der Touchdisplay war für einige Probanden zu weite vorne positioniert. Diese Aspekte sollten bei der Entwicklung von multimodalen Informationssystemen im Auto mehr berücksichtigt werden. Zur Optimierung der Interaktion mit großen Touchdisplays haben [Rümelin and Butz \(2013\)](#) Touchvarianten verglichen und zum Beispiel ein haptisches Element als Orientierungshilfe in den Screen eingebaut, mit dem die Interaktion eines Kuchenmenüs von diesem Punkt aus fast blind getätigkt werden kann. Um visuelle Überlagerungen von Karten zu verbessern haben [Lee et al. \(2013\)](#) die Suche von Elementen durch die präattentive Wahrnehmung durch das Hervorheben von Elementen verbessert und konnten somit Ablenkung reduzieren. Es gibt also nicht nur für die Kombination von Modalitäten, sondern auch für jede einzelne Modalität noch Verbesserungspotenzial.

Grundsätzlich lässt sich sagen, dass die Sprachbedienung am beliebtesten bei den Probanden war. Jedoch wurde oft angemerkt, dass Sie diese Modalität nicht mit Mitfahrern verwenden würden. Das ist ein gutes Beispiel dafür, warum es sinnvoll ist bei einem IVIS mehrere Modalitäten anzubieten. Nach der reinen Sprachbedienung war die zweitbeliebteste Kombination Touch und Sprache, dicht gefolgt von Geste und Sprache. Das zeigt, dass multimodale Interaktionen durchaus bei den Probanden gewünscht sind.

Kapitel 6

Validierung des multimodalen Modells

Wir haben jetzt aus unserer Studie ein Modell gebaut, dass die Zeiten von verschiedenen Aktionen in Abhängigkeit des vorherigen Modus darstellt. Damit sollen für zukünftige multimodale Interfaces die Dauer für eine bestimmte Moduskombination vorhergesagt werden können. Außerdem können verschiedene Varianten von Moduskombinationen auf ihre Gesamtdauer verglichen werden.

Um unsere vorhergesagten Zeiten aus denen der Studie zu vergleichen verwenden wir die gesamte Dauer der Aufgabe (Total Task Time) ([Teo and John, 2006](#)). [Teo and John \(2006\)](#) erklärt, dass es zwei Nachteile bei der Verwendung der gesamten Aufgabendauer gibt. Zum einen inkorrekte Messungen und falsch gesetzte Operatoren.

In unserem Modell haben wir verschiedene Aktionen je nach Modalität und deren Wechselkosten modelliert. Eine Aktion enthält allerdings, anders als bei den bisherigen Keystroke-Level Modellen auch mehrere Operatoren. Dies macht unser Modell etwas größer, allerdings ist die Platzierung der Aktionen einfacher. Nimmt ein Nutzer die Hand zum Beispiel einmal mehr zurück zum Lenkrad, ist dies kein Fehler, sondern hängt von der Situation ab. Diese Unterschiede in der Ausführung der Aktionen steckt in unseren Zeiten mit drinnen. Deshalb halten wir die gesamte Dauer der Interaktion für ein geeignetes Maß für den Vergleich.

6.1 Studiendesign zur Validierung

Genau wie in der ersten Studie zur Erhebung von Interaktionszeiten verwenden wir ein Within Subject Design. Zuerst wird mit Übungsbeispielen geübt und dann kommen unsere Anwendungsbeispiele.

Die fünf Anwendungsbeispiele haben wir in ihrer Moduskombination explizit so gewählt, dass sie in ihrer Eignung besonders gut waren siehe Abbildung [5.3.2](#) auf Seite [49](#) oder deren Kombination uns interessiert hat. Die Texteingabe zum Beispiel schnitt für die Sprache deutlich besser ab, als Touch, weshalb wir diese nur mit der Modalität Sprache verwendeteten. Außerdem testen wir auch Beispiele mit mehr als einem Moduswechsel, um die Übertragbarkeit unseres Modells zu testen.

Wir testen in der Evaluation das Modell also nur stichprobenartig (nicht alle Kombinationen), um

auch die Studie in kürzerem Rahmen zu halten. Die 5 Anwendungsbeispiele wurden mit dem Latin Square permutiert, um auch in dieser Studie mögliche Lerneffekte zu eliminieren. Die Übungsbeispiele am Anfang waren jedoch immer in gleicher Reihenfolge.

6.1.1 Übungsbeispiele der Validierungsstudie

Dafür bauen wir den Prototypen etwas um und testen in einer weiteren Studie mit 10 Teilnehmern. Bevor wir mit den Anwendungsbeispielen für die Evaluation beginnen, soll in der Evaluationsstudie erst geübt werden. Dafür verwenden wir die Variante des Prototypen aus der ersten Studie. Die Anwendungsbeispiele sind dort kürzer und so können sich die Probanden mit den Interaktionen vertraut machen. Wir haben folgende Übungsvarianten aus der ersten Studie gewählt (siehe Abbildung 4.1.1 auf Seite 36):

1. Navigation (Geste, Sprache)
2. Lautstärke (Geste, Touch)
3. Lautstärke (Sprache, Geste)
4. Temperatur (Sprache, Touch)
5. Telefon (Touch, Sprache)
6. Telefon (Geste, Touch)

Mit diesen Übungsbeispielen wurde die Direktauswahl aus sichtbaren Elementen (DA) mit Geste, Sprache und Touch geübt. Die Listennavigation (L) mit Sprache und Touch. Die direkte Inkrementation (Inkr. (d)) des Sliders mit Touch und Geste und zuletzt die schrittweise Inkrementation (Inkr.(s)) der Temperatureinstellung mit Touch. Das sind alle Interaktionen, die wir evaluieren wollen.

6.1.2 Anwendungsbeispiele der Validierungsstudie

Nach diesem Übungsdurchgang werden die fünf Anwendungsbeispiele getestet.

Im ersten Anwendungsbeispiel soll die Lüftung auf Stufe 3 gestellt werden, siehe Abbildung 6.1.1 auf der nächsten Seite. Dafür muss zuerst der Lüftungsbutton mit dem Sprachbefehl „Lüftung“ selektiert werden. Anschließend wird mit drei Swipe-Touchgesten die Lüftung von null auf drei gestellt. Der Wert wird nach einer kurzen Pause von selbst einprotokolliert, es muss also nicht mit einem Tap bestätigt werden. Mit unserem Modell wird die Interaktionsdauer folgendermaßen berechnet:

$$\begin{aligned}\text{Lüftung} &= \text{DA}_S + R(\text{DA}) + V(\text{Inkr.}(s))_T + 3 * \text{Inkr.}(s)_T + W_{\text{ST}} + R(\text{Screen}) = \\ &= 2,795\text{s} + 0,5\text{s} + 0,906\text{s} + 3 * 0,547\text{s} + 0,273\text{s} + 0,016\text{s} = 6,133\text{s}\end{aligned}$$

Das nächste Anwendungsbeispiel kennen wir auch schon aus der ersten Studie. Es soll die Lautstärke von 50% auf 80% erhöht werden, mit den Modalitäten Sprache und Geste, siehe Abbildung 6.1.2 auf der nächsten Seite. Zuerst wird mit den Sprachbefehlen „Medien“ und „Lautstärke“

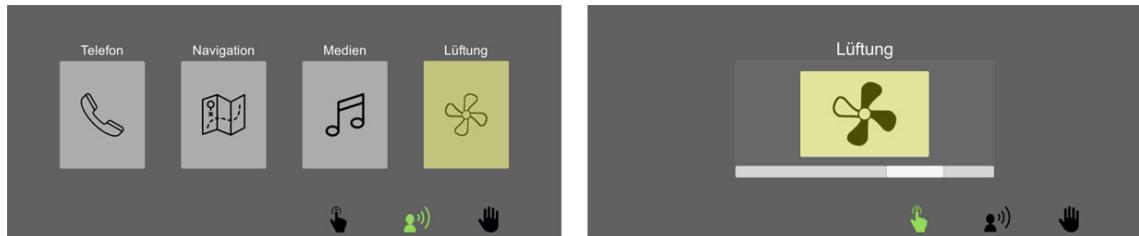


Abbildung 6.1.1: Anwendungsbeispiel Lüftung mit den Modalitäten Sprache und Touch

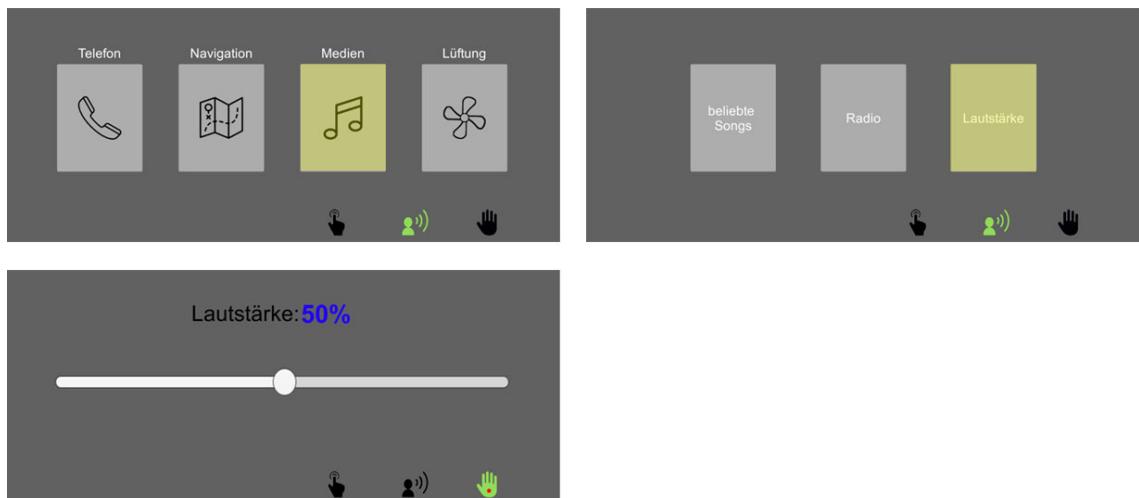


Abbildung 6.1.2: Anwendungsbeispiel Lautstärke mit den Modalitäten Sprache und Touch

zweimal die Aktion DA ausgeführt. Anschließend wird mit der Slidegeste die Lautstärke verstellt. Die Berechnung zur Vorhersage der Interaktionsdauer mit unserem Modell sieht wie folgt aus:

$$\begin{aligned}\textbf{Medien Sound} &= 2 * (\text{DA}_S + R(\text{DA})) + \text{Inkr.}(d)_G + W_{SG} + 3 * R(\text{Screen}) = \\ &= 2 * (2,795s + 0,5s) + 4,593s + 0,127s + 3 * 0,016s = 11,358s\end{aligned}$$

Im Anwendungsbeispiel Medien, soll zunächst per Geste die DA der Button (Medien) und (beliebte Songs) ausgewählt werden. Es erscheint eine Liste mit beliebten Songs. Mit dem Sprachbefehl „Happy“ wird das Lied Happy abgespielt. Auf dem nächsten Screen soll mit Touch die Lautstärke geändert werden. Dafür wird per Touch der Lautstärke Button aktiviert. Es erscheint ein Slider, der von 20% auf 80% ebenfalls per Touch geändert werden soll. Wird der Slider im Bereich von 75 und 85% losgelassen kommt ein Popup, dass noch per Touch bestätigt werden muss, siehe Abbildung 6.1.3 auf der nächsten Seite. Die Interaktionszeit besteht aus diesen Aktionen und Wechselkosten:

$$\begin{aligned}\textbf{Medien} &= 2 * (\text{DA}_G) + \text{Wort(m)}_S + W_{GS} + V(1. \text{ Bu.})_T + 1-3 \text{ Bu.} + W_{ST} + \\ &\quad + \text{Inkr.}(d)_T + B_T + 6 * R(\text{Screen}) = \\ &= 2 * (2,436s) + 2,839s + 0,105s + 1,123s + 1,495s + 0,218s + \\ &\quad + 2,795s + 0,829s + 6 * (0,016) = 15,014s\end{aligned}$$

Im nächsten Anwendungsbeispiel wollen wir mit den Modalitäten Geste, Sprache und Touch den nächsten Coffeeshop finden. Dazu werden per Geste die Buttons „Navigation“ und „Points of Interest“ ausgewählt. Dann kommt ein Texteingabefeld mit einem Inputfeld zur Eingabe des POI.

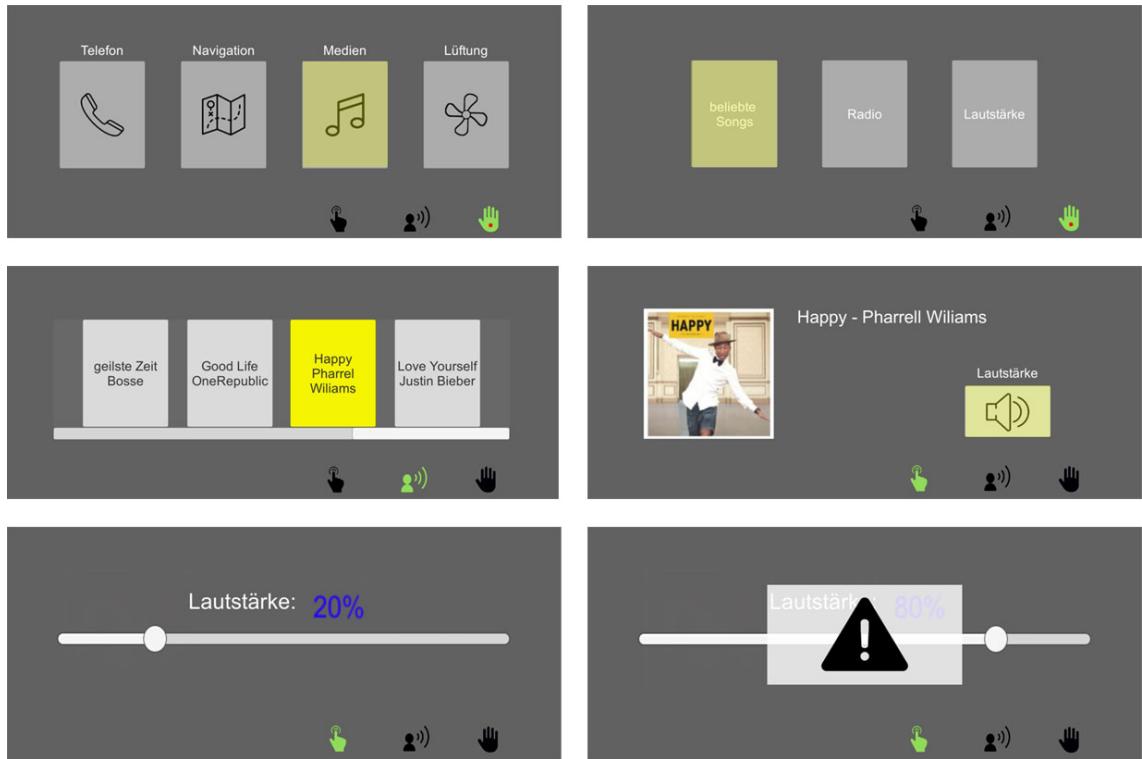


Abbildung 6.1.3: Anwendungsbeispiel Lautstärke mit den Modalitäten Geste, Sprache und Touch

Mit dem Sprachbefehl „Fastfood“ wird Fastffod im Inputfeld eingetragen. Erneut mit Spracheingabe wird die Eingabe mit „OK“ bestätigt. Jetzt kommt eine Liste von Fastfood Geschäften. Es muss zum Starbucks auf der zweiten Seite mit Touch geswiped werden. Auf der richtigen Seite angekommen muss noch der Starbucksbutton ausgewählt werden, siehe Abbildung 6.1.4 auf der nächsten Seite.

In der ersten Studie haben wir festgestellt, dass die Positionierung des Inputfeldes oben besser geeignet ist. Auch wenn in unseren Beispielen kein Touch verwendet wird, haben wir diese Anpassung bei den beiden Screens der Texteingabe vorgenommen.

Die Aktionen des Anwendungsbeispiels werden folgendermaßen zusammengesetzt und berechnet:

$$\begin{aligned}
 \text{Navi POI} &= 2 * (\text{DA}_G) + \text{Wort(m)}_S + W_{GS} + B_S + \\
 &+ V(L)_T + 2 * (L)_T + W_{ST}(L_T) + DA(L)_T + 4 * R(\text{Screen}) = \\
 &= 2 * (2,436s) + 2,839s + 0,105s + 1,972s + \\
 &= +0,705s + 2 * 0,813s + 1,495s + 0,375s + 1,047s + 4 * 0,016 = 13,605s
 \end{aligned}$$

Das letzte Anwendungsbeispiel heißt Navigation Nummer. Hier werden 2 Modalitäten (Touch und Sprache) mit vier Moduswechseln verwendet. Per Touch werden die Buttons Navigation und Zieleingabe gedrückt. Dann kommt ein Texteingabefeld mit zwei Inputfeldern (Straße und Nummer). Um ein Inputfeld zu aktivieren soll der Proband auf das Inputfeld Straße tapen. Dann soll die Straße „Parkring“ per Sprache eingegeben werden. Anschließend wird erneut mit einem Tap auf das Nummer-Inputfeld dieses aktiviert. Wieder mit Sprache wird die Nummer „19“ eingegeben. Als letztes wird mit einem Tap auf den OK-Button die Eingabe bestätigt, siehe (Abbildung 6.1.5 auf der nächsten Seite). Auch hier berechnen wir mit unserem erstellten Modell die potenzielle

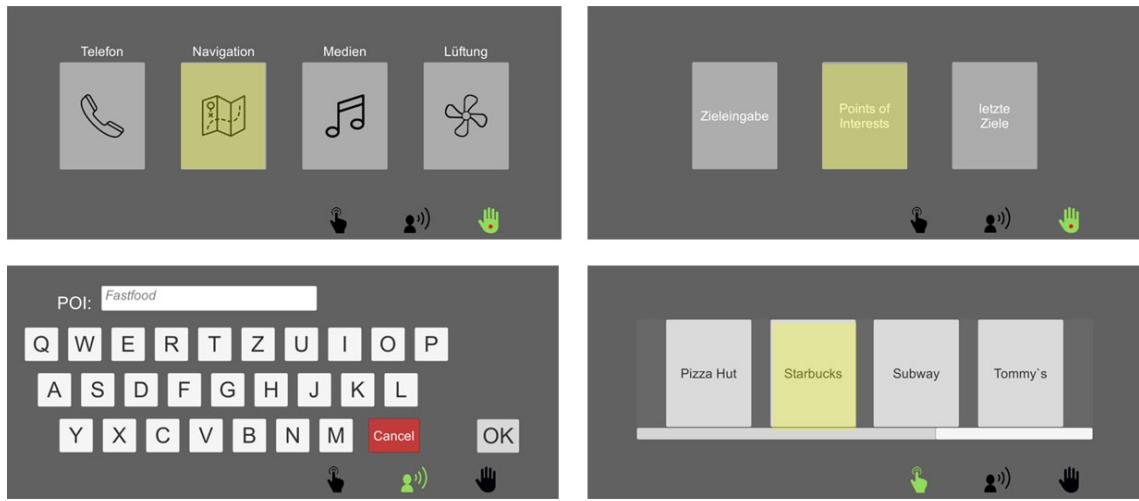


Abbildung 6.1.4: Anwendungsbeispiel Navigation POI mit den Modalitäten Geste, Sprache und Touch

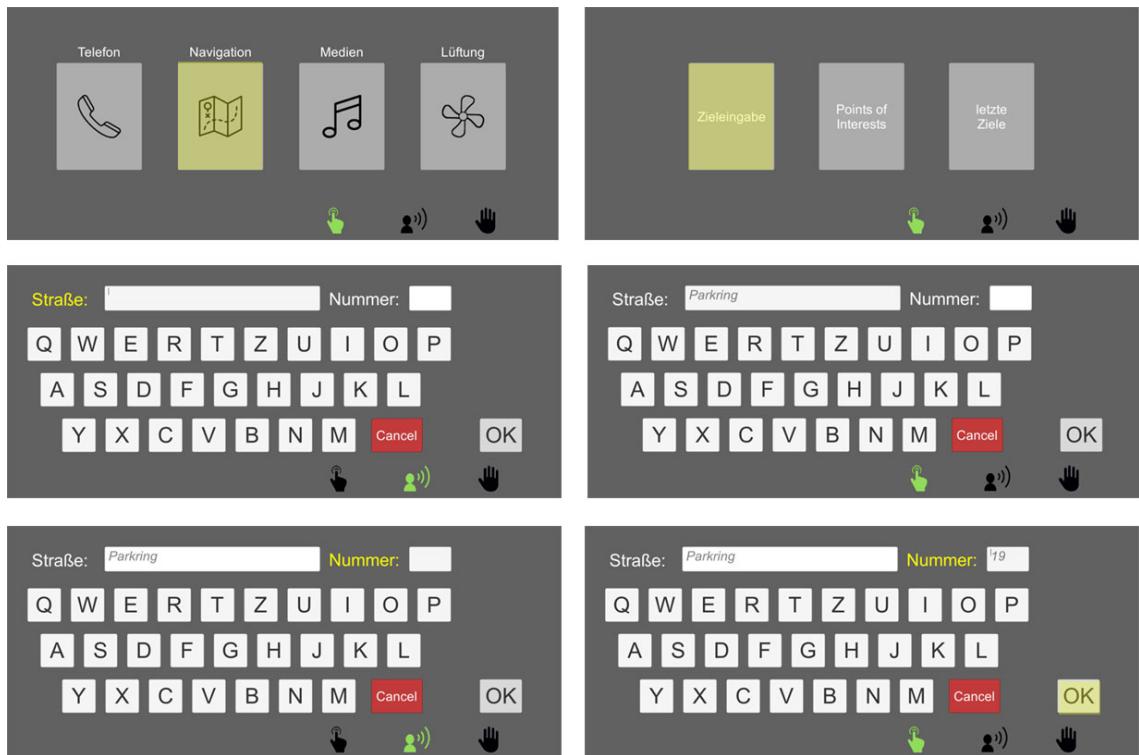


Abbildung 6.1.5: Anwendungsbeispiel Navigation POI mit den Modalitäten Touch, Sprache, Touch, Sprache, Touch

Interaktionsdauer. Sie setzt sich aus diesen Aktionen zusammen:

$$\begin{aligned}
 \text{Navi Nummer} &= 2 * (\text{DA}_T) + V(1. \text{ Bu.})_T + 1-3 \text{ Bu.}_T + \text{Wort(m)}_S + W_{\text{TS}}(\text{Wort}(m)) + \\
 &+ V(1. \text{ Bu.})_T + 1-3 \text{ Bu.}_T + W_{\text{ST}}(1. \text{ Bu.}) + \text{Wort(m)}_S + W_{\text{TS}} + B_T + W_{\text{ST}} + 4 * R(\text{Screen}) = \\
 &= 2 * (1,710s) + 1,123s + 0,642s + 2,839s + 0,079s + 1,123s + \\
 &+ 0,642s + 0,218s + 2,839s + 0,079s + 0,642s + 1,123s + 0,218s + 3 * 0,016 = 15,035s
 \end{aligned}$$

Diese Vorhersage der Interaktionsdauer für unsere 5 Anwendungsbeispiele wird die Grundlage unserer Evaluation sein. Wir werden die Anwendungsbeispiel beobachten und mit unseren Voraussagen vergleichen.

6.2 Durchführung der Studie

Die Studie fand unter gleichen Bedingungen wieder bei BMW in Garching-Hochbrück am Parkring 19 in der Parkgarage im gleichen Auto statt. Die Studie dauerte ca. 45 Minuten und wurde in einem Zeitraum von zwei Tagen durchgeführt. Die Einstellungen, der Aufbau sowie die Protokollierung wurde genau wie in der ersten Studie vorgenommen.

Die Einverständniserklärung, sowie den demographischen Fragebogen verwendeten wir gleichermaßen wie in der ersten Studie. Um dieses mal zusätzlich die Beanspruchung der Interaktionen zu messen wurde nach dem zweiten Messdurchgang jedes Anwendungsbeispiels der NASA TLX abgefragt [10](#). Nach den 5 Anwendungsbeispielen wurde der zweite Teil des NASA TLX abgefragt. Hier werden alle 6 Beanspruchungsvarianten gegeneinander verglichen. Die 15 Vergleiche wurden für jeden Probanden geshuffelt, um keinen Einfluss der Reihenfolge zu haben. Außerdem wurde der INTUI Fragebogen abgefragt, um die Usability des Prototypen einzuschätzen. sowie und noch 3 weitere Fragen siehe [10](#).

Wie schon erwähnt konnten sich die Probanden mit den Übungsbeispielen mit den Interaktionen vertraut machen. Sobald die Übungsrunde fertig war wurde die Kamera angeschaltet, um die Evaluationsbeispiele aufzunehmen, für den Fall, dass etwas nachgeprüft werden muss. Bei den Anwendungsbeispielen wurde zusätzlich wie auch bei der ersten Studie jede Aufgabe mindestens einmal geübt.

Sobald der Versuchsleiter das Gefühl hatte, dass die Aufgabe dem Probanden klar ist wurde mit dem ersten Messdurchgang begonnen. Anschließend der zweite. Traten während eines Messdurchgangs Fehler auf wurde das vom Versuchsleiter notiert ,um den Protokolleintrag im Nachhinein als fehlgeschlagen zu markieren. Dieser Messdurchgang wurde dann wiederholt. Als fehlerhaft wurde ein Messdurchgang gewertet wenn zum Beispiel die Gesten- oder Spracherkennung nicht richtig funktionierte oder wenn der Proband vergaß was er tun muss.

6.3 Auswertung der Studienergebnissen

Im folgenden werden nun die Ergebnisse der Evaluationsstudie präsentiert und diskutiert.

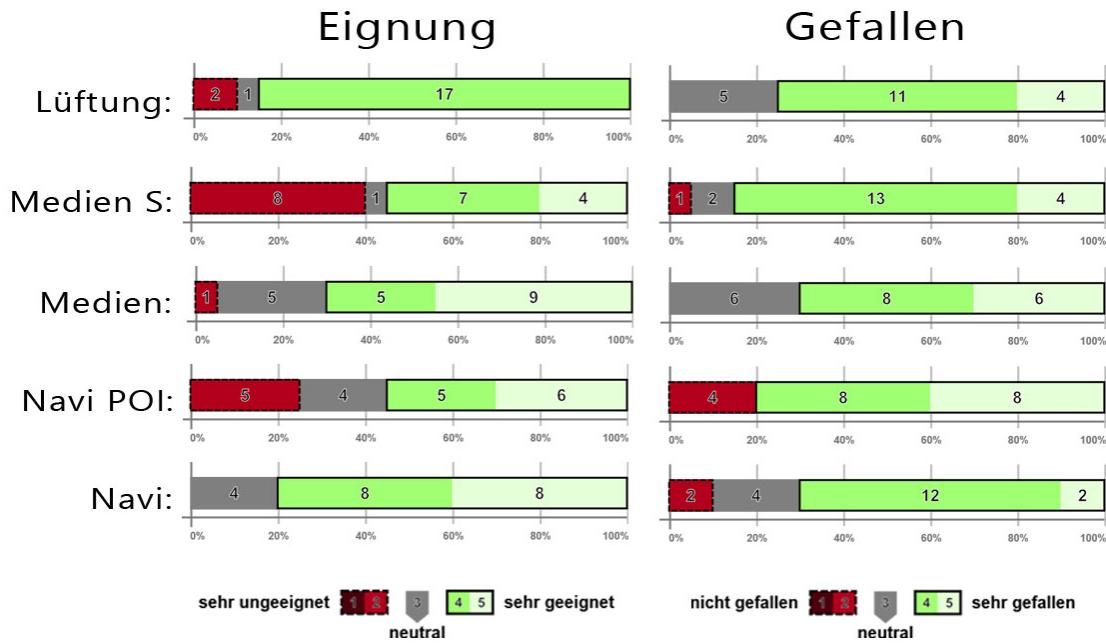


Abbildung 6.3.1: Eignung und Gefallen der 20 Messdurchgänge der 5 Aufgaben. Die Balkendiagramme zur Darstellung von Likert-Skalen wurde mit <http://likertplot.com/> generiert.

6.3.1 Teilnehmer

In der Evaluationsstudie nahmen insgesamt 10 Teilnehmer (8 männlich und 2 weiblich) teil. Das Alter erstreckte sich von 19-33 Jahren und betrug im Durchschnitt 23,8 mit einem Median von 22 Jahren, also etwas jünger als in unserer ersten Studie (Durchschnitt: 30,55 und Median: 25 Jahre). Alle 10 Teilnehmer sind Rechtshänder und sprechen deutsch als Muttersprache.

Auch in der Evaluation wurde die Vorerfahrungen mit der Bedienung von Touch, Gesten und Sprache der Teilnehmer abgefragt und von den Probanden eingeschätzt. Auf die Frage, ob sie Erfahrung mit der Bedienung von Sprache, Touch und Geste haben, konnte aus vier Optionen gewählt werden (1: nein, 2: ja, aber nur sehr wenig, 3: ja, benutze ich gelegentlich und 4: ja, benutze ich regelmäßig). Die Vorerfahrung der Bedienung von Sprache ergab im Durchschnitt 3,1, bei Touch 4 und bei Geste 2,2. Die Werte sind etwas besser als in unserer ersten Studie (Sprache: 2,32, Touch: 3,95 und Geste: 1,86), was unter anderem daran liegen könnte, dass die Teilnehmer etwas jünger waren.

6.3.2 User Experience

Im Vergleich zu Touch und Sprache war für die Teilnehmer die Gestensteuerung am ungewohntesten. Die Einschätzung der Eignung der Anwendungsbeispiele mit Geste schnitt nicht so gut ab wie die Anwendungsbeispiele ohne Geste. Allerdings bekam im Anwendungsbeispiel Medien Sound, indem der LautstärkeSlider mit einer Geste verstellt werden sollte, sehr positives Feedback siehe Abbildung 6.3.1. Es wurde auf die Frage, ob Ihnen die Interaktion gefallen würde nur eine negative Antwort gegeben. Nach jedem Messdurchgang (also 2 pro Anwendungsbeispiel) sollten die Probanden auf einer Likertskaala mit 5 Unterscheidungen angeben wie geeignet sie die Anwendung fanden und ob es Ihnen gefallen hat.

9 von 10 Teilnehmern können sich eine multimodale Interaktion während dem Fahren vorstellen. Begründungen waren die intuitive Bedienung, sowie vor allem der Vorteil sich je nach Situation die passende Modalität frei wählen zu können.

Beim NASA Task Load Index mussten 6 Beanspruchungskriterien auf einer Skala von 1-20 subjektiv von den Probanden bewertet werden. Diese Werte werden alle mit 5 multipliziert, um eine Abstufung von 1-100 zu erhalten.

Im nächsten Teil wurden alle Kriterien miteinander verglichen und mit diese 15 Vergleichen werden jetzt die Kriterien gewichtet. Wurde die Frustration 4 mal vorgezogen erhält die Frustration eine Gewichtung (G) von 4. Die Bewertung wird also mit 5 und dann mit deren Gewichtung multipliziert. All diese Werte aufsummiert durch 15 geteilt ergeben den gesamten NASA TLX. Diese Gewichtung führt dazu, dass für den Nutzer unwichtige Kriterien weniger bis gar nicht miteinbezogen werden und dafür wichtige um so mehr.

Sei B die Menge der Beanspruchungen, im speziellen Geistige Anforderung, Körperliche Anforderung, Zeitliche Anforderung, Leistung, Anstrengung und Frustration, dann sieht die Berechnung wie folgt aus:

$$\text{NASA TLX} = \frac{1}{15} \sum_{i \in B} A_i G_i$$

wobei A_i der Wert der Beanspruchung ist und G_i die Gewichtung.

Der durchschnittliche NASA TLX jedes Anwendungsbeispiels ist aus der Tabelle 6.3.1 auf Seite 79 zu entnehmen. Das Anwendungsbeispiel Lüftung verursachte mit einem Wert von 25,6 die geringste Beanspruchung. Am meisten beanspruchte das Anwendungsbeispiel Navi Nummer mit den 4 Moduswechseln mit einem Wert von 37,533. Wir stellen fest, dass mit Dauer der vorhergesagten Interaktionzeiten auch die durchschnittliche Beanspruchung zunimmt.

Im INTUI Fragebogen (siehe Anhang im Kapitel 10) werden verschiedene Komponenten der intuitiven Nutzung auf einer Skala von 1-7 abgefragt (Diefenbach and Hassenzahl, 2010), (Ullrich and Diefenbach, 2010). Die Probanden wurden hierfür darauf hingewiesen, dass sie sich dafür unseren Prototypen fehlerfrei vorstellen sollten. Denn unsere Umsetzung des Prototypen ist natürlich kein vollständig ausgearbeitetes Produkt. Die 17 Fragen mit verschiedenen Gegenüberstellungen werden vier Hauptbereichen zugewiesen (siehe (Diefenbach and Hassenzahl, 2010, Seite 24)):

- **Mühelosigkeit (M):** Je höher der Wert, um so müheloser wird die Interaktion erlebt und desto weniger Aufmerksamkeit wird erforderlich. Diese Komponente kann am ehesten mit der klassischen Usability verglichen werden.
- **Bauchgefühl (G):** Je höher der Wert, um so weniger wird die Interaktion vom Verstand, sondern vom Gefühl geleitet. Diese Komponente ist „eines der wichtigsten Merkmale intuitiver Entscheidungen in der Entscheidungspsychologie sowie im alltäglichen Sprachgebrauch“ (Diefenbach and Hassenzahl, 2010, Seite 24).
- **Magisches Erleben (X):** Je faszinierender und außergewöhnlicher die Interaktion empfunden wurde desto höher ist der Wert.
- **Verbalisierungsfähigkeit (V):** Je höher der Wert, desto besser kann die Interaktion beschrieben werden.

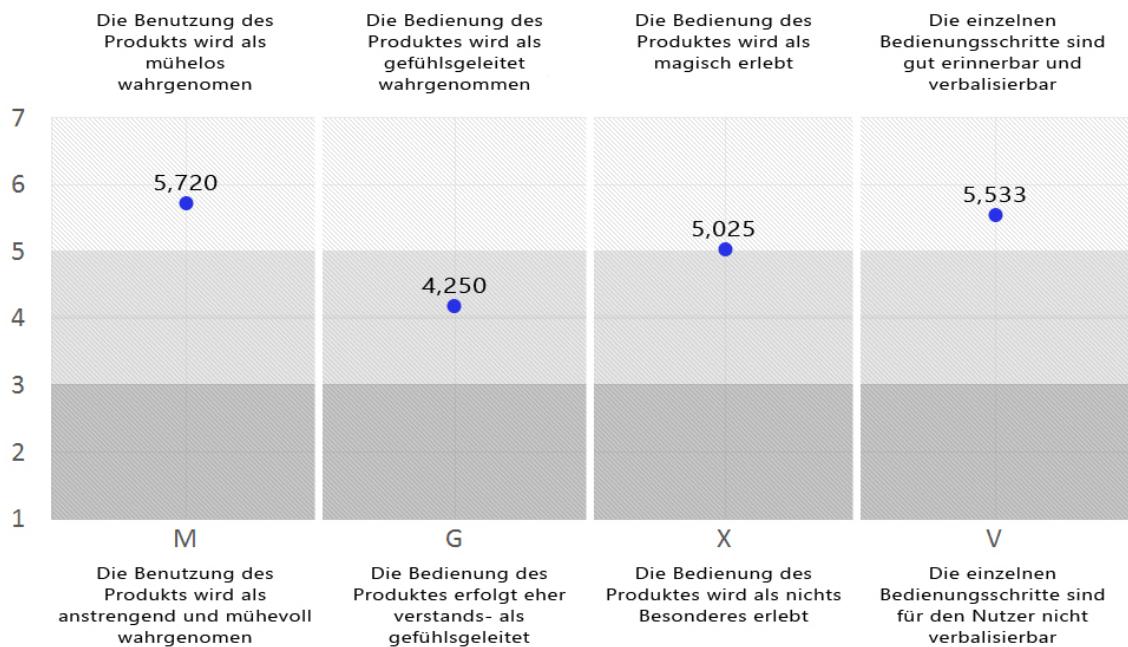


Abbildung 6.3.2: Mittelwerte der 4 Komponenten des INTUI Fragebogens

Die Ergebnisse des INTUI Fragebogens können aus Abbildung 6.3.2 auf der nächsten Seite entnommen werden. Der extremste Wert ist mit 5,72 bei der Mühelosigkeit. Das deutet auf eine gute Usability unseres Prototypen hin. Beim Bauchgefühl zeigt der Wert eine minimale Tendenz zu einer gefühlsgeleitenden Interaktion hin. Der Wert des Magischen Erlebens mit 5,025 deutet auf eine außergewöhnliche Interaktion hin. Die Abfolge unserer Bedienschritt schien unseren Probanden als einleuchtend, zumindest befindet sich der Wert der Verbalisierungsfähigkeit mit 5,533 im oberen Bereich.

6.3.3 Ergebnisse des Modells

Um unsere vorhergesagten Zeiten mit den beobachteten Zeiten vergleichen zu können berechnen wir mit dem Root Mean Square Error (RMSE). Er wird wie folgt berechnet, wobei r unsere beobachteten Zeiten pro Proband (residuals) sind und p die vorhergesagten Zeiten.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (r_i - p_i)^2}$$

Der RMSE gibt uns an wie viele Sekunden die beobachteten Zeiten durchschnittlich von unserer Vorhersage abweichen. Uns interessiert der prozentuale Fehler, also wird der RMSE noch durch unsere vorhergesagte Zeit geteilt. Wir verwenden eine logarithmische Skala, um die vorhergesagten Zeiten unseres Modells mit den beobachteten Zeiten aus der Evaluation darzustellen, da der Fehler proportional zur Dauer sein soll (Card et al., 1980) und (Schneegaß, 2009).

Die Abbildung 6.3.3 auf der nächsten Seite zeigt, dass sich unsere vorhergesagten Zeiten sehr gut mit den beobachteten Zeiten decken.

Der Durchschnittliche RMSE der 5 Anwendungsbeispiele beträgt 14,746%, was deutlich geringer ist als die maximale Erwartung von 20-30% (Card et al., 1980) und generell ganz gut im Bereich

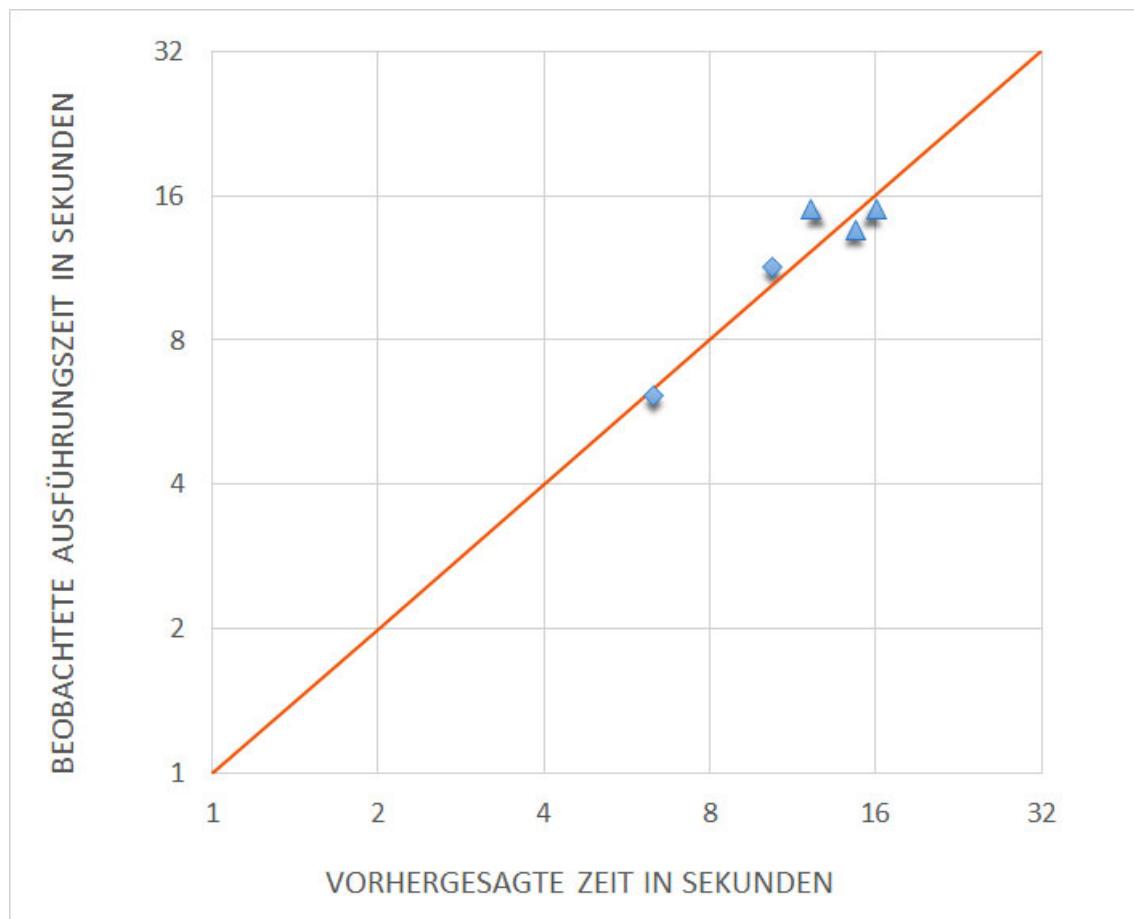


Abbildung 6.3.3: Übersicht der 5 Aufgaben im Vergleich zu den vorhergesagten und beobachteten Zeiten. Die Rauten sind Aufgaben mit nur einem Moduswechsel im Stil der ersten Studie. Die Dreiecke sind die Anwendungsbeispiele mit 2 und 4 Moduswechseln.

von 5% -20% liegt (Luo and John, 2005), (Teo and John, 2006). Da unser Modell nicht ganz so feingranulare Unterscheidungen macht wie beim KLM und deren Erweiterungen war auch ein größerer Vorhersagefehler zu erwarten. Deshalb ist der durchschnittliche Vorhersagefehler von 14,746% eine ziemlich gute Annäherung der multimodalen Interaktionszeit.

In der Tabelle 6.3.1 können die vorhergesagten Zeiten, der beobachteten Durchschnittszeiten, der RMSE in Sekunden und prozentual, sowie der durchschnittlichen Belastung des NASA TLX entnommen werden.

Aufgabe Modi (Wechsel)	vorhergesagte Zeit (s)	durchschnittliche Zeit (s)	RMSE (s)	RMSE (%)	Beanspruchung NASA TLX
Lüftung S,T (1)	6,132	6,326 (+0,194)	0,864	14,098	25,600
Medien Sound S,G (1)	11,36	10,413 (-0,947)	1,614	14,202	29,700
Medien G,S,T (2)	15,014	16,016 (+1,002)	1,860	12,389	33,100
Navi POI G,S,T (2)	13,605	14,682 (+1,077)	1,685	12,385	32,067
Navi Nummer T,S (4)	15,035	12,215 (-2,82)	3,105	20,654	37,533

Tabelle 6.3.1: Vorhergesagte Werte im Vergleich zu beobachteten Werten, deren RMSE in Sekunden und prozentual, sowie die durchschnittliche Beanspruchung des NASA TLX (Skala von 0-100)

Unser Modell scheint auch für Interaktionen mit mehr als einem Moduswechsel anwendbar zu sein. Gerade die Vorhersage der Anwendungsbeispiele Medien und Navi POI mit je 3 verschiedenen Modalitäten und somit 2 Modalitätswechsel ergeben die besten Ergebnisse mit einer Fehlerabweichung von nur 12,39%.

Die kompletten Ergebnisse aller 10 Probanden dieser beiden Anwendungsbeispiele sind in Abbildung 6.3.4 auf der nächsten Seite und Abbildung 6.3.5 auf der nächsten Seite zu sehen.

Bei dem Anwendungsbeispiel Medien liegt die Durchschnittszeit der Probanden über der vorhergesagten Zeit. Ein Grund dafür kann sein, dass für die Erhebung der Aktion Inkr. (d) eine Verschiebung von 50% auf das Interval 75-85% gemessen wurde. In dem Evaluationsbeispiel sollte die Lautstärke von 20% auf das Interval von 75-85% gestellt werden. Da hier die Spanne größer ist, könnte dieser Teil etwas länger gedauert haben.

Auch beim Anwendungsbeispiel Navi POI lag die vorhergesagte Zeit unter dem Durchschnitt.

Die Anwendungsbeispiele Lüftung und Medien Sound waren fast identisch zu Beispielen aus der ersten Studie. Die hatten beide nur einen Moduswechsel. Beim Anwendungsbeispiel Lüftung liegt unsere Vorhersage unter dem Durchschnitt. Das Anwendungsbeispiel Medien Sound liegt allerdings unsere Vorhersage über dem Durchschnitt. Das ist eventuell der Tatsache zu schulden, dass die Probanden der Evaluationsstudie insgesamt besser mit der Slidegeste für die Einstellung des Lautstärkesliders zurecht kamen. Die kompletten Ergebnisse aller 10 Probanden dieser beiden Anwendungsbeispiele sind in den 6.3.6 und 6.3.7 zu entnehmen.

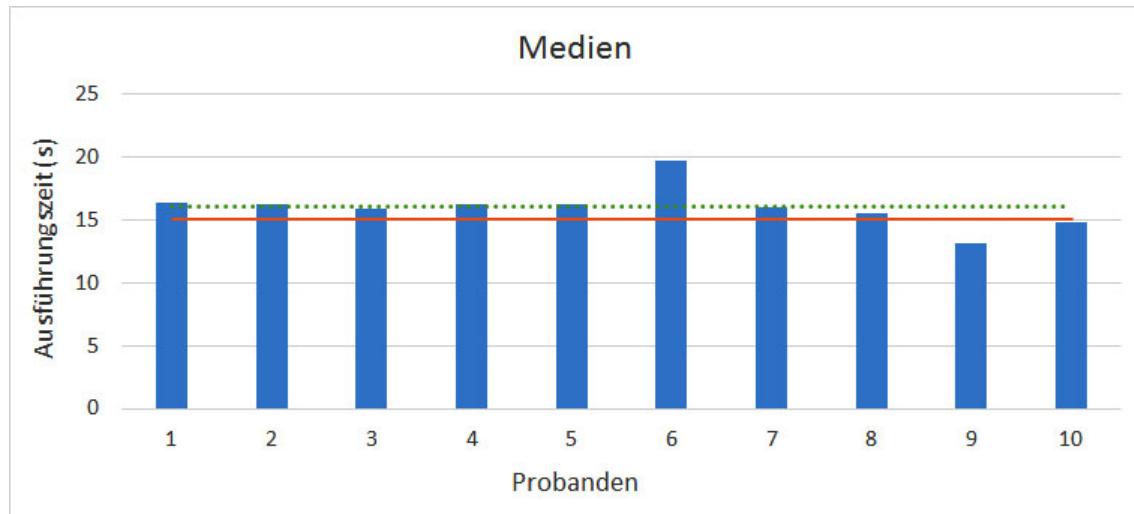


Abbildung 6.3.4: Zeiten der Probanden für das Anwendungsbeispiel Medien. Enthält Modalitäten Geste, Sprache und Touch. Die rote durchgehende Linie stellt die vorhergesagte Zeit dar, die grüne gepunktete die Durchschnittszeit aller Probanden.

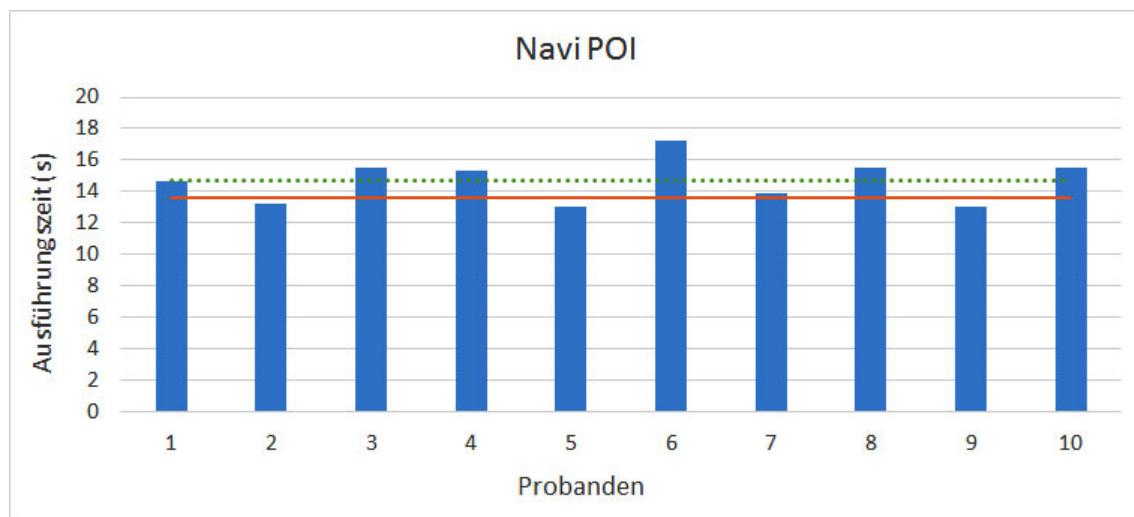


Abbildung 6.3.5: Zeiten der Probanden für das Anwendungsbeispiel Navigation zu POI. Enthält Modalitäten Geste, Sprache und Touch. Die rote durchgehende Linie stellt die vorhergesagte Zeit dar, die grüne gepunktete die Durchschnittszeit aller Probanden.

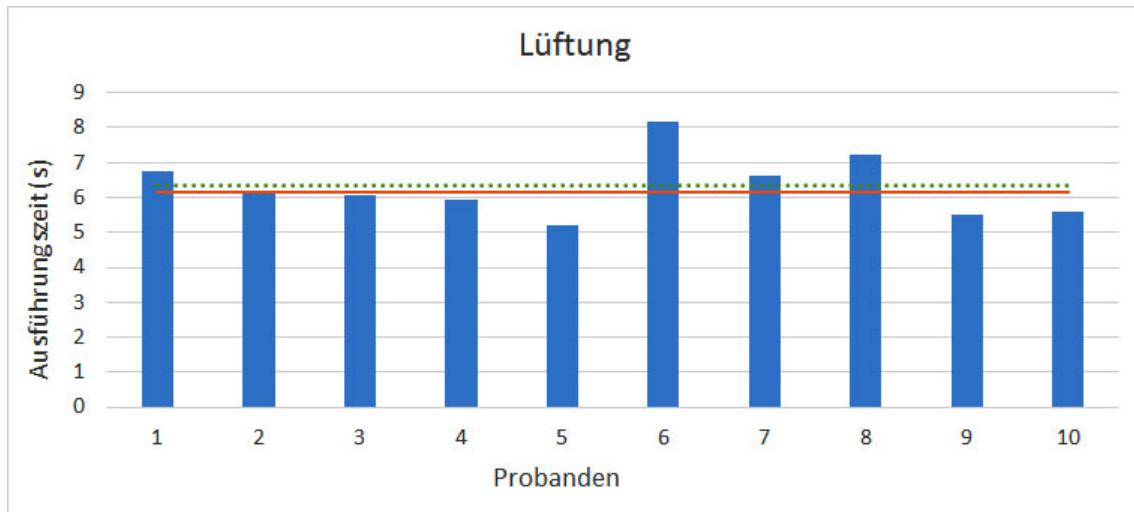


Abbildung 6.3.6: Zeiten der Probanden für das Anwendungsbeispiel Lüftung. Enthält Modalitäten Sprache und Touch. Die rote durchgehende Linie stellt die vorhergesagte Zeit dar, die grüne gepunktete die Durchschnittszeit aller Probanden.

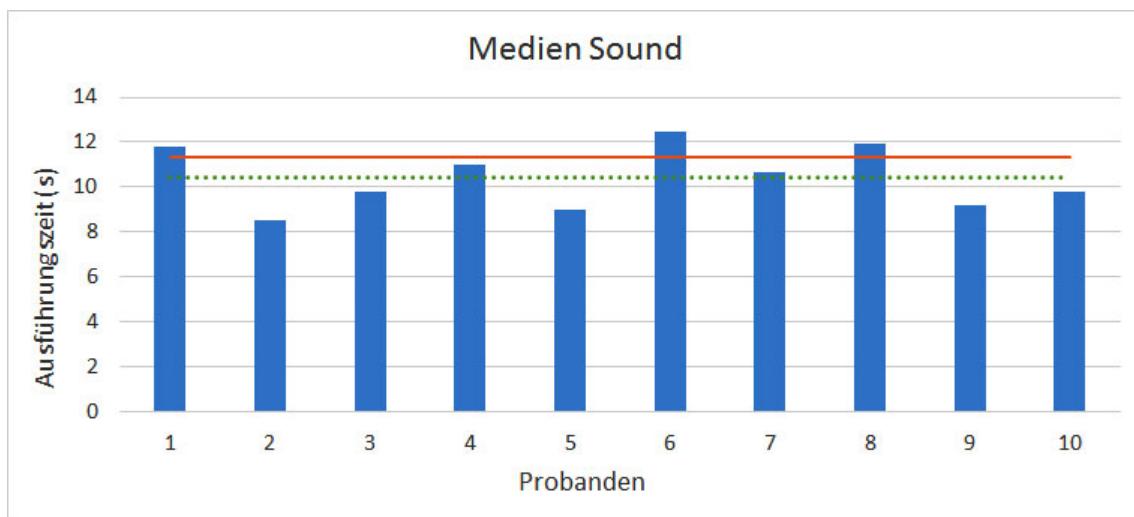


Abbildung 6.3.7: Zeiten der Probanden für das Anwendungsbeispiel Medien Sound. Enthält Modalitäten Sprache und Geste. Die rote durchgehende Linie stellt die vorhergesagte Zeit dar, die grüne gepunktete die Durchschnittszeit aller Probanden.

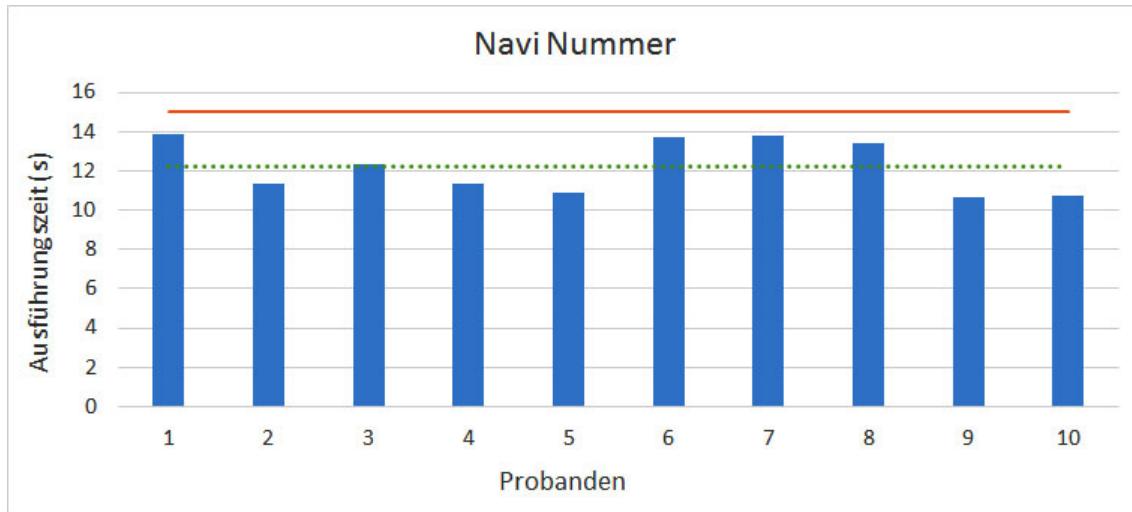


Abbildung 6.3.8: Zeiten der Probanden für das Anwendungsbeispiel Navigation. Enthält Modalitäten Touch, Sprache, Touch, Sprache und Touch. Die rote durchgehende Linie stellt die vorhergesagte Zeit dar, die grüne gepunktete die Durchschnittszeit aller Probanden.

Das Anwendungsbeispiel Navi schnitt am schlechtesten ab mit einem prozentualen RMSE von 20,654%. Mit zwei verschiedenen Modalitäten und vier Moduswechseln ist dies das aufwändigste Beispiel. Für die Aktivierung der Inputfelder verwendeten wir in unserem Modell zur Vorhersage die Zeit den ersten Buchstaben zu drücken. Im Beispiel ist die Touchfläche allerdings deutlich größer als die Touchfläche der Buchstaben aus unserer ersten Studie. Nach Fitts' Law (Fitts, 1954), (Sasangohar et al., 2009) beeinflusst die Größe der Touchfläche W die Interaktionszeit positiv. Es erschien es uns besser die Interaktionszeit zu überschätzen, als zu unterschätzen. Es lässt sich in diesem Beispiel gut erkennen, dass unsere Vorhersage für alle Probanden zu groß waren siehe Abbildung 6.3.8.

6.4 Diskussion

Die Evaluierung unseres multimodalen Modells soll nun mit Hinblick auf die Ergebnisse diskutiert und mit Ergebnissen aus anderen Arbeiten verglichen werden. Des Weiteren gehen wir auf qualitative Erkenntnisse ein, die sich aus beiden Studien durch Kommentare und Beobachtungen ergeben haben.

6.4.1 Multimodales Modell

Die fünf evaluierten Anwendungsbeispiele ließen sich mit unserem Modell mit einem durchschnittlichen RMSE von 14,746% vorhersagen. Dieser durchschnittliche RMSE befindet sich sowohl im Bereich von 20-30% den Card et al. (1980) als maximalen Fehler empfiehlt, als auch im Bereich von 5-20% (Luo and John, 2005), (Teo and John, 2006). Card et al. (1980), Luo and John (2005) und Teo and John (2006) bewerten allerdings ein Keystroke-Level-Modell oder deren Erweiterung auf Operatorebene. Davon ausgehend, dass unser Modell wesentlich grobgranularer aufgebaut ist als das Keystroke-Level-Modell, stellt sich unser Modell zur Vorhersage von multimodalen Interaktionen als vielversprechend heraus.

Die subjektive Bewertung der Nutzer durch den NASA TLX Ihrer Beanspruchung während der multimodalen Interaktion lag zwischen 25,6 und 37,533 auf einer Skala von 0-100. Den durchschnittlichen Beanspruchungswert von 25,6 hatte das Anwendungsbeispiel Lüftung mit den Modalitäten Sprache und Touch mit einem Moduswechsel. Es lässt sich erkennen, dass die Beanspruchung mit der vorhergesagten Interaktionsdauer zunahm bis zum Wert von 37,533 im Anwendungsbeispiel Navi Nummer. Dieses Anwendungsbeispiel stellte sich mit insgesamt 4 Moduswechseln der Modalitäten Touch und Sprache als das mit der größten Beanspruchung heraus. (Paper mit vergleichbaren Ergebnissen).

Diese Beanspruchungswerte scheinen in Ordnung zu sein, denn 9 von 10 Teilnehmern können sich eine multimodale Interaktion im Stile unserer Evaluationsstudie während dem Fahren vorstellen. Die intuitive Bedienung, sowie dem Vorteil sich je nach Situation die passende Modalität frei wählen zu können kommt bei den Nutzern gut an. Jedoch bleibt in diesem Feld der multimodalen Interaktion noch viel zu forschen und entwickeln bis die Erwartungen und Vorteile vollends ausgeschöpft werden können.

Beispielsweise haben wir in unserer Evaluation nicht alle Aktionen mit allen Modalitäten evaluiert. Mit 1,5 Stunden dauerte die erste Studie sehr lange und es konnte beobachtet werden, dass die Konzentration mit der Zeit etwas nach ließ. Die Evaluationsstudie wollten wir dementsprechend kürzer gestalten.

Außerdem haben wir uns speziell Anwendungsbeispiele für die Evaluationsstudie herausgesucht, die uns möglichst realistisch in der Anwendung erschienen. Die Bewertung der Nutzer auf die Eignung der Interaktionen und wie gut sie Ihnen gefallen hat schritt ganz gut. Das lässt darauf hindeuten, dass unsere Wahl zutreffend war, siehe Abbildung 6.3.1 auf Seite 75. Die Texteingabe per Touch wurde nicht evaluiert, da die Spracheignung deutlich besser eingeschätzt wurde. Ebenso die Aktion Inkr. (d) und Inkr. (s) wurde für die Gesteninteraktion nicht evaluiert.

Die beliebtesten Moduskombinationen der 22 Nutzer aus der ersten Studie waren die unimodale Variante mit Sprache. Am zweit beliebtesten war die Kombination von Touch und Sprache. Die dritt beliebteste Variante war die Kombination von Geste und Sprache.

Das Design unseres Prototypen wurde Anhand der im Workshop (Connected Minds) gewonnenen Informationen über Interaktionen im Auto modelliert. Wir mussten uns auf bestimmte Gesten einigen, die uns am sinnvollsten erschienen für unsere Aktionen erschienen. Es gibt jedoch andere Varianten, die ebenfalls als Aktionen in Frage kommen. Zum Beispiel die kreisförmige Geste mit dem Zeigefinger nach links oder rechts, um die Lautstärke zu verringern oder zu erhöhen. Eine Geste dieser Art ist bereits im BMW 7 Series eingebaut. Die Implementierung mit der Leap Motion dieser Geste war mit der neuen Version von Unity nicht so einfach umsetzbar, weshalb wir uns dort eine andere Variante überlegten.

Ein weiterer Mangel unseres Modells sind, wie schon erwähnt, die fehlenden haptischen Bedienelemente. Diese sind natürlich ein wichtiger Bestandteil und sollte als weitere Modalität in unser Modell miteinbezogen werden.

Bei der Aktion Inkr. (d) wurde in der ersten Studie zur Erhebung der Interaktionszeiten lediglich der Slider von 50% auf 75-85% verstellt. Für unsere grobe Betrachtung erschien uns diese Einstellung als ausreichend zur Bestimmung dieser Interaktionszeit. Da in dem Anwendungsbeispiel der Evaluation jedoch die Zeiten für eine Verstellung von 20 auf 75-85% länger dauerte, sollte diese Aktion eventuell genauer untersucht werden.

6.4.2 Qualitative Erkenntnisse

Während beider Studien zur Bedienung eines multimodalen Interface im Auto ist aufgefallen, dass der Bereich von Gesteninteraktionen sehr individuell von der Sitzeinstellung und Armlänge abhängt. Am ergonomisch geeignetsten hatten es die Nutzer, die ihren Arm an der Armlehne ablegen konnten und sich trotzdem mit der Hand im gewünschten Interaktionsbereich befanden.

Für zukünftige Gestensteuerung im Auto sollte das berücksichtigt werden, um verkrampte und unnatürliche Bewegungen zu reduzieren. Zum Beispiel könnte durch eine verstellbare Armlehne, sowie einen anpassbaren Interaktionsbereich zur Gestenerkennung die optimale Einstellung erreicht werden. Das ermöglicht nicht nur eine bequemere Gesteninteraktion, sondern reduziert auch mögliche Erkennungsfehler von Gesten, was wiederum zu einer besseren User Experience führt.

Der Touchbereich war ebenfalls für manche Nutzer nur gut zu erreichen, indem sie sich etwas vor lehnten. Sie wurden angewiesen sich vor der Studie den Sitz so einzustellen als würden sie fahren, doch trotz diesen Vorkehrungen war der Touchdisplay nicht für alle optimal positioniert. Auch in der Touchbedienung gibt es noch Potenzial zur Verbesserung. Die Erreichbarkeit des Touchbereichs sollte weiter vorne gelegen sein, als bei Informationssystemen im Auto ohne Touchdisplays.

Kapitel 7

Zusammenfassung und Ausblick

Motiviert durch die Vorteile verschiedener Modalitäten und die damit verbundene Möglichkeit den Nutzer zu entlasten und ihn somit weniger von der Fahraufgabe abzulenken. Der Fahrer kann selbst entscheiden in welchen Situationen welcher Modus für den jeweiligen Schritt am geeignetsten und ungefährlichsten ist, was auch laut [Muller and Weinberg \(2011\)](#) einen wesentlichen Vorteil in Autos darstellt.

Deshalb war unsere Motivation solche multimodalen Interaktionen in IVIS besser verstehen zu können und diese optimal umsetzen zu können. Mit der Entwicklung unsres Modells zur Vorhersage von multimodalen Interaktionszeiten ist es uns gelungen Designer die Möglichkeit zu geben bereits in einem frühem Stadium der Entwicklungsphase multimodale Interaktionen einzuschätzen und zu vergleichen. Damit können Designer in geeigneter Weise die besten potenziellen Varianten unterstützen, um dem Fahrer auch die besten Möglichkeiten in verständlicher Weise anzuseigen. Damit soll natürlich auch die visuelle und mentale Beanspruchung für den Fahrer so gering wie möglich gehalten werden.

Ob eine Variante besonders gut oder schlecht ist hängt natürlich von der Situation ab, Deshalb wäre ein seriell redundantes multimodalen IVIS am besten, um in jedem Schritt den Fahrer selbst die beste Modalität wählen zu lassen. Die Eingabe eines Ziels per Touch ist zum Beispiel im stehenden Auto eine gute Variante, die jedoch während der Fahrt den Fahrer durch die längere Interaktion zu sehr ablenken könnte. In diesem Fall stellt die Texteingabe per Sprache eine sehr geeignete Alternative dar.

Unser Modell zur Vorhersage von multimodalen Interaktionen im IVIS sollte in Zukunft noch dahingehend weiterentwickelt werden auch haptische Bedienelemente miteinzubeziehen. Zudem kann das Modell auf weitere Gesten erweitert werden und auch die weitere Untersuchungen von gesprochenen Sätzen könnte hilfreich sein.

Im Zuge dieser Masterarbeit wurde im Workshop Connected Minds die Bedienung im Fahrzeug beobachtet und diskutiert. Es wurde ein Konzept für ein multimodales Modell entworfen und in einem multimodalen Prototypen umgesetzt. Dieser kann mit Touch, Sprache und Geste bedient werden. Zur Erhebung der Interaktionszeiten und der Evaluierung des Modells wurde der Prototyp seriell exklusiv umgesetzt, da dem Nutzer die Modalität vorgegeben war. Unser Modell unterscheidet verschiedene Aktionen in Abhängigkeit der Modalität und enthält entstehende Wechselkosten wenn der Nutzer die Modalität ändert.

KAPITEL 7. ZUSAMMENFASSUNG UND AUSBLICK

Kapitel 8

Danksagung

Besonderer Dank geht an meine Betreuer Bastian Pfleging und Florian Roider, die mich während dieser Arbeit mit Rat und Tat unterstützt haben. Mein Dank geht natürlich auch an die insgesamt 32 Teilnehmer der zwei Nutzerstudien. Außerdem möchte ich meinem Mann (Sascha Rothe) für seine mentale Unterstützung und spontane Brainstorming Runden danken.

Für den Prototypen wurden verschiedene Icon von www.flaticon.com verwendet.

- Daumen, Musik Icon: made by Gregor Cresnar (<http://www.flaticon.com/authors/gregor-cresnar>).
- Touch Icon: made by Freepik (<http://www.flaticon.com/authors/freepik>).
- Navigation Icon: made by Madebyoliver (<http://www.flaticon.com/authors/madebyoliver>).
- Telefon, Gefahr, Touch, Geste und Sprach Icon: made by Freepik (<http://www.flaticon.com/authors/freepik>).
- Klima Icon: made by Icon Works (<http://www.flaticon.com/authors/icon-works>).

Flaticon is licensed by Creative Commons BY 3.0 (<http://creativecommons.org/licenses/by/3.0/>)

KAPITEL 8. DANKSAGUNG

Kapitel 9

Verzeichnisse

Literaturverzeichnis

Evgeniy Abdulin. Using the keystroke-level model for designing user interface on middle-sized touch screens. In *CHI '11 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '11, pages 673–686, New York, NY, USA, 2011. ACM. ISBN 978-1-4503-0268-5. doi: 10.1145/1979742.1979667. URL <http://doi.acm.org/10.1145/1979742.1979667>.

National Highway Traffic Safety Administration et al. Visual-manual nhtsa driver distraction guidelines for in-vehicle electronic devices. *Washington, DC: National Highway Traffic Safety Administration (NHTSA), Department of Transportation (DOT)*, 2012.

Eduardo Bertoldi and Lucia Filgueiras. Multimodal advanced driver assistance systems: An overview. In *Proceedings of the 2Nd International Workshop on Multimodal Interfaces for Automotive Applications*, MIAA '10, pages 2–5, New York, NY, USA, 2010. ACM. ISBN 978-1-60558-997-8. doi: 10.1145/2002368.2002370. URL <http://doi.acm.org.emedien.ub.uni-muenchen.de/10.1145/2002368.2002370>.

Gary Burnett. Designing and evaluating in-car user-interfaces. In *Handbook of Research on User Interface Design and Evaluation for Mobile Technology*, pages 218–236. IGI Global, 2008.

Andreas Butz and Antonio Krüger. *Mensch-Maschine-Interaktion*. Walter de Gruyter GmbH & Co KG, 2014.

Stuart K. Card, Thomas P. Moran, and Allen Newell. The keystroke-level model for user performance time with interactive systems. *Commun. ACM*, 23(7):396–410, July 1980. ISSN 0001-0782. doi: 10.1145/358886.358895. URL <http://doi.acm.org/10.1145/358886.358895>.

Stuart K Card, Allen Newell, and Thomas P Moran. The psychology of human-computer interaction. 1983.

S Diefenbach and M Hassenzahl. Handbuch zur fun-ni toolbox–user experience evaluation auf drei ebenen. *Online verfügbar: http://fun-ni.org/wpcontent/uploads/Diefenbach+Hassenzahl_2010_HandbuchFun-niToolbox.pdf*, 2010.

Tanja Döring, Dagmar Kern, Paul Marshall, Max Pfeiffer, Johannes Schöning, Volker Gruhn, and Albrecht Schmidt. Gestural interaction on the steering wheel: Reducing the visual demand. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, pages 483–492, New York, NY, USA, 2011. ACM. ISBN 978-1-4503-0228-9. doi: 10.1145/1978942.1979010. URL <http://doi.acm.org.emedien.ub.uni-muenchen.de/10.1145/1978942.1979010>.

Karim El Batran and Mark D. Dunlop. Enhancing klm (keystroke-level model) to fit touch screen mobile devices. In *Proceedings of the 16th International Conference on Human-computer Interaction with Mobile Devices & Services*, MobileHCI '14, pages 283–286, New York, NY, USA, 2014. ACM. ISBN 978-1-4503-3004-6. doi: 10.1145/2628363.2628385. URL <http://doi.acm.org/10.1145/2628363.2628385>.

Andy Field and Graham Hole. *How to design and report experiments*. Sage, 2002.

Paul M Fitts. The information capacity of the human motor system in controlling the amplitude of movement. *Journal of experimental psychology*, 47(6):381, 1954.

G Geiser. Man machine interaction in vehicles. *ATZ*, 87(74-77):56, 1985.

P Green. Sae j2364—navigation and route guidance function accessibility while driving (draft). *Warrendale, PA: Society of automotive engineers*, 1999a.

Paul Green. The 15-second rule for driver information systems. In *Proceedings of the ITS America Ninth Annual Meeting*. Intelligent Transportation Society of America Washington, DC, 1999b.

Paul Green. Motor vehicle driver interfaces. In Julie A. Jacko and Andrew Sears, editors, *The Human-computer Interaction Handbook*, pages 844–860. L. Erlbaum Associates Inc., Hillsdale, NJ, USA, 2003. ISBN 0-8058-3838-4. URL <http://dl.acm.org.emedien.ub.uni-muenchen.de/citation.cfm?id=772072.772126>.

Paul Green. Using standards to improve the replicability and applicability of driver interface research. In *Proceedings of the 4th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, AutomotiveUI '12, pages 15–22, New York, NY, USA, 2012. ACM. ISBN 978-1-4503-1751-1. doi: 10.1145/2390256.2390258. URL <http://doi.acm.org.emedien.ub.uni-muenchen.de/10.1145/2390256.2390258>.

Driver Focus-Telematics Working Group et al. Statement of principles, criteria and verification procedures on driver interactions with advanced in-vehicle information and communication systems. *Alliance of automotive manufacturers*, 2006.

Sandra G Hart. Nasa-task load index (nasa-tlx); 20 years later. In *Proceedings of the human factors and ergonomics society annual meeting*, volume 50, pages 904–908. Sage Publications Sage CA: Los Angeles, CA, 2006.

Sandra G Hart and Lowell E Staveland. Development of nasa-tlx (task load index): Results of empirical and theoretical research. *Advances in psychology*, 52:139–183, 1988.

Paul Holleis, Friederike Otto, Heinrich Hussmann, and Albrecht Schmidt. Keystroke-level model for advanced mobile phone interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '07, pages 1505–1514, New York, NY, USA, 2007. ACM. ISBN 978-1-59593-593-9. doi: 10.1145/1240624.1240851. URL <http://doi.acm.org/10.1145/1240624.1240851>.

Bonnie E. John, Konstantine Prevas, Dario D. Salvucci, and Ken Koedinger. Predictive human performance modeling made easy. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '04, pages 455–462, New York, NY, USA, 2004. ACM. ISBN 1-58113-702-8. doi: 10.1145/985692.985750. URL <http://doi.acm.org/10.1145/985692.985750>.

Alvin Jude, G. Michael Poor, and Darren Guinness. Evaluating multimodal interaction with gestures and speech for point and select tasks. In *Proceedings of the 8th Nordic Conference on Human-Computer Interaction: Fun, Fast, Foundational*, NordiCHI '14, pages 1027–1030, New York, NY, USA, 2014. ACM. ISBN 978-1-4503-2542-4. doi: 10.1145/2639189.2670267. URL <http://doi.acm.org.emedien.ub.uni-muenchen.de/10.1145/2639189.2670267>.

Dagmar Kern and Albrecht Schmidt. Design space for driver-based automotive user interfaces. In *Proceedings of the 1st International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, AutomotiveUI '09, pages 3–10, New York, NY, USA, 2009. ACM. ISBN 978-1-60558-571-0. doi: 10.1145/1620509.1620511. URL <http://doi.acm.org/10.1145/1620509.1620511>.

Joonbum Lee, John D Lee, and Dario D Salvucci. A saliency-based search model: Application of the saliency map for driver-vehicle interfaces. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 57, pages 1933–1937. SAGE Publications Sage CA: Los Angeles, CA, 2013.

Hui Li, Ying Liu, Jun Liu, Xia Wang, Yujiang Li, and Pei-Luen Patrick Rau. Extended klm for mobile phone interaction: A user study result. In *CHI '10 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '10, pages 3517–3522, New York, NY, USA, 2010. ACM. ISBN 978-1-60558-930-5. doi: 10.1145/1753846.1754011. URL <http://doi.acm.org.emedien.ub.uni-muenchen.de/10.1145/1753846.1754011>.

Lu Luo and Bonnie E. John. Predicting task execution time on handheld devices using the keystroke-level model. In *CHI '05 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '05, pages 1605–1608, New York, NY, USA, 2005. ACM. ISBN 1-59593-002-7. doi: 10.1145/1056808.1056977. URL <http://doi.acm.org/10.1145/1056808.1056977>.

Jannette Maciej and Mark Vollrath. Comparison of manual vs. speech-based interaction with in-vehicle information systems. *Accident Analysis & Prevention*, 41(5):924–930, 2009.

I. Scott MacKenzie. Fitts' law as a research and design tool in human-computer interaction. *Hum.-Comput. Interact.*, 7(1):91–139, March 1992. ISSN 0737-0024. doi: 10.1207/s15327051hci0701_3. URL http://dx.doi.org/10.1207/s15327051hci0701_3.

Stefan Mattes. The lane-change-task as a tool for driver distraction evaluation. *Quality of Work and Products in Enterprises of the Future*, 2003:57, 2003.

Leap Motion. Unity - game engine kernel description. URL <https://www.leapmotion.com/>.

Christian Muller and Garrett Weinberg. Multimodal input in the car, today and tomorrow. *IEEE MultiMedia*, 18(1):98–103, January 2011. ISSN 1070-986X. doi: 10.1109/MMUL.2011.14. URL <http://dx.doi.org/10.1109/MMUL.2011.14>.

Vicki L Neale, Thomas A Dingus, Sheila G Klauer, Jeremy Sudweeks, and Michael Goodman. An overview of the 100-car naturalistic study and findings. *National Highway Traffic Safety Administration, Paper*, 2005.

Robert Neuss. *Usability Engineering als Ansatz zum Multimodalen Mensch-Maschine-Dialog*. Dissertation, Technische Universität München, München, 2001.

Commission of the European Communities. Commission recommendation of 22 december 2006 on safe and efficient in-vehicle information and communication systems: update of the european statement of principles on human machine interface. *Official Journal of the European Union*, 2008.

A Pauzié. A method to assess the driver mental workload: The driving activity load index (dali). *IET Intelligent Transport Systems*, 2(4):315, 2008.

MA Pettitt, GE Burnett, S Bayer, and A Stevens. Assessment of the occlusion technique as a means for evaluating the distraction potential of driver support systems. In *IEE Proceedings. Intelligent Transport Systems*, volume 4, pages 259–266, 2006.

Michael Pettitt, Gary Burnett, and Alan Stevens. An extended keystroke level model (klm) for predicting the visual demand of in-vehicle information systems. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '07, pages 1515–1524, New York, NY, USA, 2007. ACM. ISBN 978-1-59593-593-9. doi: 10.1145/1240624.1240852. URL <http://doi.acm.org/10.1145/1240624.1240852>.

Bastian Pfleging, Michael Kienast, Albrecht Schmidt, and Tanja Döring. T.: Speet: A multimodal interaction style combining speech and touch interaction in automotive environments. In *In Adjunct Proc. AutomotiveUI 11*, 2011.

Bastian Pfleging, Stefan Schneegass, and Albrecht Schmidt. Multimodal interaction in the car: Combining speech and gestures on the steering wheel. In *Proceedings of the 4th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, AutomotiveUI '12, pages 155–162, New York, NY, USA, 2012. ACM. ISBN 978-1-4503-1751-1. doi: 10.1145/2390256.2390282. URL <http://doi.acm.org/10.1145/2390256.2390282>.

Roberto Pieraccini, Krishna Dayanidhi, Jonathan Bloom, Jean-Gui Dahan, Michael Phillips, Bryan R. Goodman, and K. Venkatesh Prasad. Multimodal conversational systems for automobiles. *Commun. ACM*, 47(1):47–49, January 2004. ISSN 0001-0782. doi: 10.1145/962081.962104. URL <http://doi.acm.org.emedien.ub.uni-muenchen.de/10.1145/962081.962104>.

Leah M. Reeves, Jennifer Lai, James A. Larson, Sharon Oviatt, T. S. Balaji, Stéphanie Buisine, Penny Collings, Phil Cohen, Ben Kraal, Jean-Claude Martin, Michael McTear, TV Raman, Kay M. Stanney, Hui Su, and Qian Ying Wang. Guidelines for multimodal user interface design. *Commun. ACM*, 47(1):57–59, January 2004. ISSN 0001-0782. doi: 10.1145/962081.962106. URL <http://doi.acm.org.emedien.ub.uni-muenchen.de/10.1145/962081.962106>.

Gary B Reid, F Thomas Eggemeier, and Clark A Shingledecker. Subjective workload assessment technique. Technical report, DTIC Document, 1982.

Andrew D. Rice and Jonathan W. Lartigue. Touch-level model (tlm): Evolving klm-goms for touchscreen and mobile devices. In *Proceedings of the 2014 ACM Southeast Regional Conference*, ACM SE '14, pages 53:1–53:6, New York, NY, USA, 2014. ACM. ISBN 978-1-4503-2923-1. doi: 10.1145/2638404.2638532. URL <http://doi.acm.org.emedien.ub.uni-muenchen.de/10.1145/2638404.2638532>.

A. Riener, A. Ferscha, F. Bachmair, P. Hagnmüller, A. Lemme, D. Muttenthaler, D. Pühringer, H. Rogner, A. Tappe, and F. Weger. Standardization of the in-car gesture interaction space. In *Proceedings of the 5th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, AutomotiveUI '13, pages 14–21, New York, NY, USA, 2013. ACM. ISBN 978-1-4503-2478-6. doi: 10.1145/2516540.2516544. URL <http://doi.acm.org.emedien.ub.uni-muenchen.de/10.1145/2516540.2516544>.

K Rumar. In-vehicle information systems. *International journal of vehicle design*, 9(4-5):548–556, 1988.

Sonja Rümelin and Andreas Butz. How to make large touch screens usable while driving. In *Proceedings of the 5th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, AutomotiveUI ’13, pages 48–55, New York, NY, USA, 2013. ACM. ISBN 978-1-4503-2478-6. doi: 10.1145/2516540.2516557. URL <http://doi.acm.org.emedien.ub.uni-muenchen.de/10.1145/2516540.2516557>.

Daniel Salber and Joëlle Coutaz. Applying the wizard of oz technique to the study of multimodal systems. In *International Conference on Human-Computer Interaction*, pages 219–230. Springer, 1993.

Farzan Sasangohar, I Scott MacKenzie, and Stacey D Scott. Evaluation of mouse and touch input for a tabletop display using fitts’ reciprocal tapping task. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 53, pages 839–843. SAGE Publications, 2009.

Stefan Schneegaß. Modeling interaction with automotive user interfaces. Master’s thesis, University Duisburg-Essen, 2009.

Stefan Schneegaß, Bastian Pfleging, Dagmar Kern, and Albrecht Schmidt. Support for modeling interaction with automotive user interfaces. In *Proceedings of the 3rd International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, AutomotiveUI ’11, pages 71–78, New York, NY, USA, 2011. ACM. ISBN 978-1-4503-1231-8. doi: 10.1145/2381416.2381428. URL <http://doi.acm.org/10.1145/2381416.2381428>.

Rajeev Sharma, Vladimir I Pavlovic, and Thomas S Huang. Toward multimodal human-computer interface. *Proceedings of the IEEE*, 86(5):853–869, 1998.

Tobias Stracke. Touch-screen-interaktion auf der mittelkonsole zur steuerung von funktionen im fahrzeug. Master’s thesis, Universität Stuttgart, 2014.

Unity Technologies. Leap motion sdk, orion beta, a. URL <https://developer.leapmotion.com/unity#100>.

Unity Technologies. Unity - game engine kernel description, b. URL <https://unity3d.com/de>.

Leonghwee Teo and Bonnie E. John. Comparisons of keystroke-level model predictions to observed data. In *CHI ’06 Extended Abstracts on Human Factors in Computing Systems*, CHI EA ’06, pages 1421–1426, New York, NY, USA, 2006. ACM. ISBN 1-59593-298-4. doi: 10.1145/1125451.1125713. URL <http://doi.acm.org/10.1145/1125451.1125713>.

Marcus Tonnis, Verena Broy, and Gudrun Klinker. A survey of challenges related to the design of 3d user interfaces for car drivers. In *Proceedings of the 3D User Interfaces*, 3DUI ’06, pages 127–134, Washington, DC, USA, 2006. IEEE Computer Society. ISBN 1-4244-0225-5. doi: 10.1109/VR.2006.19. URL <http://dx.doi.org.emedien.ub.uni-muenchen.de/10.1109/VR.2006.19>.

Daniel Ullrich and Sarah Diefenbach. Intui. exploring the facets of intuitive interaction. In *Mensch & computer*, volume 10, page 251, 2010.

Jacob O. Wobbrock, Meredith Ringel Morris, and Andrew D. Wilson. User-defined gestures for surface computing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI ’09, pages 1083–1092, New York, NY, USA, 2009. ACM. ISBN 978-1-60558-246-7. doi: 10.1145/1518701.1518866. URL <http://doi.acm.org/10.1145/1518701.1518866>.

Abbildungsverzeichnis

2.1.1 Aufteilung der Bereiche von primären, sekundären und tertiären Aufgaben. Bild entnommen aus Tonnis et al. (2006)	16
2.1.2 Evaluierungstypen und ihre Abhängigkeit zu Validität und deren Kontrollierbarkeit	17
2.2.1 Orginaler Aufbau des Versuchs zur Erstellung von Fitts' Law	20
2.2.2 Plazierungsregeln des mentalen Operators	21
2.2.3 Orginale Tabelle der Operatoren von Card et al. (1980)	22
2.2.4 Tabelle der Operatorzeiten von Schneegaß	24
2.3.1 multimodale Kommunikationskanäle	25
3.1.1 Connected Minds bei BMW	30
3.1.2 Connected Minds Ergebnisse	31
4.1.1 Übersicht der Screenabfolge, der 4 Anwendungsbeispiele.	36
4.1.2 Symbolische Anzeige des aktiven Modus	37
4.1.3 Ampeldarstellung zur Gestenerkennung	38
4.2.1 Touchgesten: Tap, Slide und Swipe	39
4.2.2 Selektionsgeste, Slidegeste und Swipegeste	40
4.2.3 Ausschnitt aus der Logdatei vom Anwendungsbeispiel Lautstärke	41
4.2.4 Einstellungen vor jedem Durchlauf	42
5.1.1 Permutation der Anwendungsbeispiel	44
5.1.2 Eignung und Gefallen eines Anwendungsbeispiels	45

ABBILDUNGSVERZEICHNIS

ABBILDUNGSVERZEICHNIS

5.2.1 BMW 6er Gran Coupé in der Parkgarage	46
5.2.2 Anbringung des Surface im Auto	46
5.2.3 Anbringung der Kamera im Studienauto	47
5.2.4 Anschluss für die Leap Motion und der externen Tastatur	47
5.3.1 Vorerfahrung der Probanden	49
5.3.2 Eignung des Screenotypen	49
5.3.3 Durchschnittszeiten in Sekunden der Direktauswahl	51
5.3.4 Durchschnittszeiten für Swipe 1, 2 und 3 in Sekunden der Liste für die Moduskombinationen von Touch und Geste	52
5.3.5 Durchschnittszeiten in Sekunden für die Direktauswahl innerhalb der Listen Aktion	52
5.3.6 Durchschnittszeiten für Touch in Sekunden der einzelnen Buchstaben	53
5.3.7 Kategorien der Buchstaben für Wörter.	53
5.3.8 Durchschnittszeiten in Sekunden der Ziele für Sprache	54
5.3.9 Durchschnittszeiten und Medianwerte in Sekunden für die Bestätigung der Texteingabe.	54
5.3.10 Durchschnittszeiten für Inkr. (d)	55
5.3.11 Durchschnittszeiten für die Bestätigung des Popups	56
5.3.12 Durchschnittszeiten für Inkr. (s)	56
5.4.1 Unterschiedliche Varianten einer Direktauswahl	58
5.6.1 Skizze des Interaktionsbereichs zur Gestenerkennung	64
6.1.1 Anwendungsbeispiel Lüftung mit den Modalitäten Sprache und Touch	71
6.1.2 Anwendungsbeispiel Lautstärke mit den Modalitäten Sprache und Touch	71
6.1.3 Anwendungsbeispiel Lautstärke mit den Modalitäten Geste, Sprache und Touch .	72
6.1.4 Anwendungsbeispiel Navihation POI mit den Modalitäten Geste, Sprache und Touch	73
6.1.5 Anwendungsbeispiel Navihation POI mit den Modalitäten Touch, Sprache, Touch, Sprache, Touch	73
6.3.1 Eignung und Gefallen der 5 Aufgaben	75

6.3.2 Mittelwerte der 4 Komponenten des INTUI Fragebogens	77
6.3.3 Übersicht der 5 Aufgaben im Vergleich zu den vorhergesagten und beobachteten Zeiten	78
6.3.4 Zeiten der Probanden für das Anwendungsbeispiel Medien.	80
6.3.5 Zeiten der Probanden für das Anwendungsbeispiel Navigation zu POI.	80
6.3.6 Zeiten der Probanden für das Anwendungsbeispiel Lüftung.	81
6.3.7 Zeiten der Probanden für das Anwendungsbeispiel Medien Sound.	81
6.3.8 Zeiten der Probanden für das Anwendungsbeispiel Navigation.	82

Tabellenverzeichnis

3.2.1 Ergebnisse der Umsetzungsmöglichkeiten der Modalitäten für die verschiedenen Aktionen	32
5.4.1 Durchschnittszeiten der Aktionen vor dem Moduswechsel	58
5.4.2 Durchschnittszeiten der Aktionen nach dem Moduswechsel	59
5.4.3 Erste Vereinfachung der Durchschnittszeiten der Aktionen nach dem Moduswechsel	60
5.4.4 Zusammengefasste Zeiten der Operatoren vor dem Moduswechsel	61
5.4.5 Zweite Vereinfachung Durchschnittszeiten von Touch und Geste nach dem Moduswechsel	61
5.4.6 Zweite Vereinfachung Durchschnittszeiten von Sprache nach dem Moduswechsel	62
5.5.1 Zusammengefasste Zeiten der Operatoren vor dem Moduswechsel	62
5.5.2 unimodale Interaktionszeiten.	62
5.5.3 Wechselkosten eines Moduswechsels	63
6.3.1 Vorhergesagte Werte im Vergleich zu beobachteten Werten	79

Kapitel 10

Anhang

1. Einverständniserklärung
2. demografischer Fragebogen
3. Fragebogen nach der Studie
4. Studienleitfaden
5. NASA TLX (deutsche Version)
6. INTUI

Einverständniserklärung

Ich

(Name, Vorname)

Geburtsdatum

wurde über folgende Punkte im Zusammenhang mit der Untersuchung

„Modellierung multimodaler Interaktion im Auto“

unterrichtet und erkläre mich einverstanden.

- Ich wurde für mich ausreichend mündlich und/oder schriftlich über die wissenschaftliche Untersuchung informiert.
- Ich erkläre mich bereit, dass im Rahmen der Studie Daten über mich gesammelt und anonymisiert aufgezeichnet werden. Es wird gewährleistet, dass meine personenbezogenen Daten nicht an Dritte weitergegeben werden und nur intern im Kontext der BMW Forschung verwendet werden.
Bei der Veröffentlichung in einer wissenschaftlichen Zeitung wird aus den Daten nicht hervorgehen, wer an dieser Untersuchung teilgenommen hat. Meine persönlichen Daten unterliegen dem Datenschutzgesetz.
- Ich weiß, dass ich jederzeit meine Einverständniserklärung, ohne Angabe von Gründen, widerrufen kann, ohne dass dies für mich nachteilige Folgen hat.
- Mit der vorstehend geschilderten Vorgehensweise bin ich einverstanden und bestätige dies mit meiner Unterschrift.

(Ort, Datum)

(Unterschrift)

VP ID: _____

Bitte füllen Sie den folgenden Fragebogen zu Ihrer Person aus. Die Angaben werden vertraulich und anonymisiert ausgewertet:

Alter:	_____ Jahre	
Geschlecht:	<input type="checkbox"/> männlich <input type="checkbox"/> weiblich	
Ist Deutsch Ihre Muttersprache?	<input type="checkbox"/> ja <input type="checkbox"/> nein, sondern _____	
Sind Sie Rechts- oder Linkshänder	<input type="checkbox"/> rechts <input type="checkbox"/> links <input type="checkbox"/> beidhändig	
Haben Sie Erfahrung mit der Bedienung von Geräten mit Sprachsteuerung (z.B. am Smartphone, im Auto)?	<input type="checkbox"/> ja, benutze ich regelmäßig <input type="checkbox"/> ja, benutze ich gelegentlich <input type="checkbox"/> ja, aber nur sehr wenig <input type="checkbox"/> nein	 
Haben Sie Erfahrung mit der Bedienung von Touchgeräten (Handys mit Touchscreen/Tablets)?	<input type="checkbox"/> ja, benutze ich regelmäßig <input type="checkbox"/> ja, benutze ich gelegentlich <input type="checkbox"/> ja, aber nur sehr wenig <input type="checkbox"/> nein	 
Haben Sie Erfahrung mit der Bedienung von Gestensteuerung (z.B. im Auto, Kinect)?	<input type="checkbox"/> ja, benutze ich regelmäßig <input type="checkbox"/> ja, benutze ich gelegentlich <input type="checkbox"/> ja, aber nur sehr wenig <input type="checkbox"/> nein	 

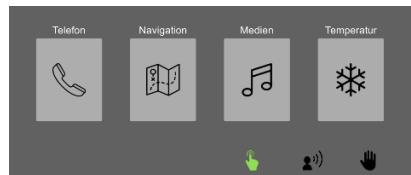
Fragebogen zur Studie

VP ID: _____

Bewerten sie die verschiedenen Screentypen nach ihrer Eignung für Touch, Geste und Sprache.

(Abstufungen sind: nicht geeignet, eher nicht geeignet, weder noch, eher geeignet, geeignet)

Screen Direktauswahl:



Ist für **Touch**
nicht geeignet **geeignet**

Ist für **Geste**
nicht geeignet **geeignet**

Ist für **Sprache**
nicht geeignet **geeignet**

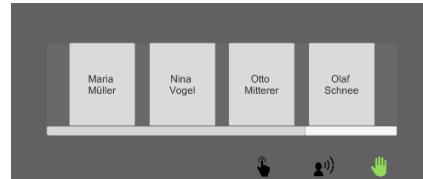
Screen Texteingabe:



Ist für **Touch**
nicht geeignet **geeignet**

Ist für **Sprache**
nicht geeignet **geeignet**

Screen Liste (4 Elemente pro Seite):



Ist für **Touch**
nicht geeignet **geeignet**

Ist für **Geste**
nicht geeignet **geeignet**

Ist für **Sprache**
nicht geeignet **geeignet**

Screen Slider:



Ist für **Touch**
nicht geeignet **geeignet**

Ist für **Geste**
nicht geeignet **geeignet**

Ist für **Sprache**
nicht geeignet **geeignet**

Screen Liste (1 Element pro Seite):



Ist für **Touch**
nicht geeignet **geeignet**

Ist für **Geste**
nicht geeignet **geeignet**

Ist für **Sprache**
nicht geeignet **geeignet**

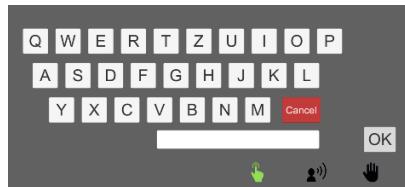
Geben Sie bitte ein **Ranking** an, welche Modi (Touch, Geste, Sprache) sie pro Screen am besten geeignet (1), zweitbesten geeignet (2) und am wenigsten geeignet (3) fanden.



Screen Direktauswahl:

(1: Am besten geeignet, 2: am zweitbesten geeignet, 3: am wenigsten geeignet)

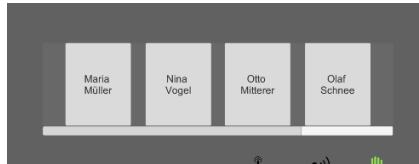
Touch: ____ Geste: ____ Sprache: ____



Screen Texteingabe:

(1: Am besten geeignet, 2: am zweitbesten geeignet)

Touch: ____ Sprache: ____



Screen Liste (4 Elemente pro Seite):

(1: Am besten geeignet, 2: am zweitbesten geeignet, 3: am wenigsten geeignet)

Touch: ____ Geste: ____ Sprache: ____



Screen Slider:

(1: Am besten geeignet, 2: am zweitbesten geeignet, 3: am wenigsten geeignet)

Touch: ____ Geste: ____ Sprache: ____



Screen Liste (1 Element pro Seite):

(1: Am besten geeignet, 2: am zweitbesten geeignet, 3: am wenigsten geeignet)

Touch: ____

Geste: ____

Sprache: ____

Sonstige Kommentare, Ideen, Anregungen?:

Vielen Dank 😊

Studienleitfaden

Mitzubringendes Material:

- Surface + Stromkabel
- Leap
- Tastatur
- Fragebogen, Permutation, UseCases, Icons, Smileys, Studienleitfaden
- Klemmbrett
- Stifte
- Autoschlüssel + Autokarte
- Ausweis

Setting im Auto einrichten:

- Surface anbringen und starten, an Strom schließen
- Leap und Tastatur an Hub anschließen, Leap aktivieren
- Hub zwischen Sicherheitsgurt einklemmen.
- Kamera anbringen (Akku und Speierkarte prüfen)
- Logfile wird geprüft. Wurde der letzte Stand als Sicherheitskopie gesichert?
Ist die Datei noch irgendwo geöffnet? Wenn ja schließen.
- Teskleiste nach oben bringen.
- Maus ausschalten

Proband abholen:

erstmal vielen Dank, dass Sie bei meiner Studie teilnimmst. Ich bin die Carina und mache in Zusammenarbeit der LMU (Bastian Pfleging) und BMW (Florian Roider) meine Masterarbeit.

Thema ist **Modellierung multimodaler Interaktion im Auto**

Ich werde während der Durchführung dieser Studie meinem Leitfaden (hier) folgen, damit die Voraussetzungen für alle Teilnehmer gleich sind. Also nicht wundern wenn ich Sätze von meinem Zettel ablese.

- Besucherausweis holen, falls nötig
- Du oder Sie?
- Die Studie wird ca. 1,5 Stunden dauern, müssen Sie/Du nochmal auf die Toilette?
- OK, dann gehen wir jetzt in die Parkgarage zum Grand Coupe, in dem die Studie stattfinden wird.

Beim Auto:

- Ich bitte dich dein Handys auf lautlos zu schalten
- Die Füße bitte nicht auf die Pedale, sonst allerdings so sitzen als würde man fahren
- Zündung starten
- Licht einschalten
- Sitzheizung nach Wunsch einschalten
- Lüftung ausschalten
- Du kannst schon einmal deinen Stitz einstellen als würdest du fahren. Ich starte in der Zeit den Prototypen und erkläre ihnen/dir dann gleich um was es geht.
- Während der Studie kann das Licht in der Garage an oder aus gehen. Davon so gut es geht nicht irritieren lassen.

Thema erklären:

- In dieser Studie geht es darum Bedienzeiten von multimodalen Interaktionen zu messen.
 - Dieser multimodale Prototyp kann mit Touch, Gestik und Sprache bedient werden. Ich habe ein Interface entworfen, dass in vereinfachter und möglichst abstrakter Form übliche Anwendungsbeispiele im Auto darstellt. Es stellt kein komplettes Interface dar.
 - In dieser Studie werden verschiedene Zeiten geloggt und wir nehmen den Verlauf der Studie auf Video auf. All diese Daten werden natürlich nur anonymisiert und vertraulich verwendet.
 - Ich bitte dich nun die **Einverständniserklärung** zu unterschreiben und einen kurzen **Fragebogen** auszufüllen.
 - Die Permutation mit Probandennummer wird geladen und angezeigt. Die Infos werden mit meinem Zettel der Permutationen auf Richtigkeit verglichen.
 - Dem Probanden wird das Erste Anwendungsbeispiel gezeigt und der zu verwendende Modus, Modi erklärt.
-
- **UseCases:** Auf diesem Blatt sehen sie 4 Anwendungsbeispiele, die sie später ausführen werden. Wie sie sehen, gibt es zwei Bereiche, zwischen denen ein Moduswechsel möglich ist. Ich erkläre jetzt erstmal das erste Anwendungsbeispiel mit den zu verwendenden Modi.

Telefon: Das ist der Hauptscreen mit dem jedes Anwendungsbeispiel beginnen wird. Dort wählen sie per Direktauswahl das Telefon aus. Dann wechselt der Screen zu einer neuen Direktauswahl mit 3 Buttons. Dort wählen sie Kontakte. Anschließend kommt ein Screen mit einer horizontalen Liste, die geswipt werden kann und soll. Der Slider unten dient nur zur Visualisierung. Ziel ist es hier den Kontakt Maria Müller zu finden, indem die Liste 3 mal geswiped wird und anschließend der Kontakt Maria Müller ausgewählt und bestätigt wird.

- **Touch:**
Tap mit dem ein Element selektiert wird. Es reicht eine leichte Berührungen
Swipegeste bei der zur nächsten Seite gewechselt wird. Hier im Touchbereich von rechts nach links gewischt.
- **Geste:**
Auswahl: Ist der gewünschte Bereich gelb gehovered kann mit einer Tipgeste des Zeigefingers der Button selektiert werden.
Swipe Geste: Bei Listen navigieren wir mit einer schnellen Wischbewegung der ganzen Hand von rechts nach links. Es soll die komplette Hand bewegt werden.
- **Sprache:**
Die Sprachbefehle sind: Telefon -> Kontakte -> Maria Müller

Navigation: Auch hier beginnen wir mit dem Hauptscreen, dort wählen sie die Navigation aus. Dann wechselt der Screen und gibt 3 Buttons, hier wählen sie die Zieleingabe aus. Jetzt kommt ein Screen mit einer Tastatur, wo das Ziel entweder per Sprache oder per Touch eingegeben werden soll. Es wird in dieser Studie 3 verschiedene Ziele geben: Rom, Dorfweg und Kirchengasse. Ist das verlangte Ziel eingegeben wird mit Ok die Eingabe bestätigt. Auch hier mit Touch oder Sprache, je nach Modus der verwendet werden soll.

- **Touch:**
Tap mit dem ein Element selektiert wird. Es reicht eine leichte Berührungen
- **Geste:**
Auswahl: Ist der gewünschte Bereich gelb gehovered kann mit einer Tipgeste des Zeigefingers der Button selektiert werden.
- **Sprache:**
Die Sprachbefehle sind: Navigation -> Zieleingabe -> Rom/Dorfweg/Kirchengasse -> OK

Lautstärke: Auch hier kommt der Hauptscreen, dort wählen sie Medien aus. Im nächsten Screen gibt es 3 Buttons von denen sie Lautstärke auswählen. Jetzt kann die Lautstärke mit einem Slider eingestellt werden. Die Aufgabe ist es die Lautstärke von 50% auf 80% zu erhöhen. Ist der Wert des Sliders zwischen 75 und 80% öffnet sich unabhängig vom Modus ein Warn-Popup. Das muss noch bestätigt werden.

- **Touch:**

Tap mit dem ein Element selektiert wird. Es reicht eine leichte Berührung
Bewegungsgeste mit der ein Slider versteilt werden soll (kein Tap). Man berrührt den Slider in der Mitte und zieht ihn nach rechts bis der Wert zwischen 75% und 85% liegt.

- **Geste:**

Auswahl: Ist der gewünschte Bereich gelb gehovered kann mit einer Tipgeste des Zeigefingers der Button selektiert werden.

Slider: nehmen sie den Slider in der Luft auf, indem sie einen Kreis horizontal mit ihrem Daumen und Zeigefinger formen. Jetzt können sie die Hand in dieser Position nach recht verschieben. Sobald sie im Intervall 75-85% sind ist es in Ordnung. Indem sie den Kreis öffnen stoppt der Slider.

- **Sprache:**

Die Sprachbefehle sind: Medien -> Lautstärke -> 80% -> Ok

Temperatur: Auch hier kommt der Hauptscreen, dort wählen sie rechts die Temperatur aus. Auf dem nächsten Screen kann die Temperatur eingestellt werden. Ziel ist es die Temperatur auf 20 Grad zu erhöhen. Anschließend muss nix weiter ausgewählt werden, die Hände können zurück ans Lenkrad.

- **Touch:**

Tap mit dem ein Element selektiert wird. Es reicht eine leichte Berührung

Swipegeste bei der zur nächsten Seite gewechselt wird. Hier im Touchbereich von rechts nach links gewischt.

- **Geste:**

Auswahl: Ist der gewünschte Bereich gelb gehovered kann mit einer Tipgeste des Zeigefingers der Button selektiert werden.

Swipe Geste: Bei Listen navigieren wir mit einer schnellen Wischbewegung der ganzen Hand von rechts nach links. Es soll die komplette Hand bewegt werden.

- **Sprache:**

Die Sprachbefehle sind: Temperatur -> 20 Grad

- Dieses Anwendungsbeispiele werden sie jetzt nacheinander mit verschiedenen Modalitäten durchspielen.
- **Icons:** Auf den jeweiligen Screens sehen sie immer den auszuführenden Modus in grün hervorgehoben. Vor jedem Beispiel werden ich auch nochmal sagen was in welchem Modus gemacht werden muss. Es wird dann immer 1-2 Probendurchläufe geben, damit keine Fehler mehr für die Messdaten entstehen. Anschließend werden wir 2 Messdurchgänge machen.

Ampelkreis bei Geste: Der Interaktionsbereich liegt zwischen dem Dashboard und der Gangschaltung. Die Hand wird allerdings auch schon am Lenkrad erkannt. Wurde deine Hand noch nicht erkannt ist der Kreis rot, wurde die Hand erkannt, aber du befindest dich nicht im Interaktionsraum ist der Kreis gelb. Der Kreis wird grün wenn deine Hand sich im Interaktionsraum befindet. Hier können die Gesten ausgeführt werden.

- **Smileys:** Nach einem Messdurchgang werden 2 Fragen zu der gerade ausgeführten Aufgabe gestellt. Sie sollen einmal beurteilen ob sie die Anwendung mit den verwendeten Modalitäten geeignet finden. Die Abstufungen gehen von ungeeignet, eher ungeeignet, neutral, eher geeignet und geeignet.
Als zweites sollen sie bewerten, ob ihnen die Interaktion mit den verwendeten Modalitäten gefallen hat. Auch hier wird abgestuft von hat mir nicht gefallen, eher nicht gefallen, neutral, eher gefallen und gefallen. Per Touch wird die Einschätzung gewählt und der nächste Durchlauf kann beginnen.
- Vor jedem Versuch befinden sich beide **Hände am Lenkrad**. Sobald ich die Anwendung starte wird der Bildschirm schwarz, sobald der Hauptscreen zu sehen ist können sie mit der Interaktion beginnen.
- Bei einer Anwendung mit Geste die Hand vor Beginn kurz flach über die Leap halten und dann die Hand ans Lenkrad.

Studie kann beginnen:

- Kameras werden gestartet:
- Maus ausschalten!
- Der erste Probendurchlauf kann starten.
- Der zweite Probendurchlauf wird gestartet
- Wenn alles gepasst hat wird das Häckchen bei Messdurchgang gesetzt.
- Auf meinem Zettel mit den Permutationen trage ich nach erfolgreichem Durchgang eine Zahl zwischen -2 und 2 ein.
- Erneuter Messdurchgang
- Das gleiche passiert mit allen Varianten und Anwendungsbeispielen.
- Anmerkungen oder Auffälligkeiten werden notiert.

So jetzt haben sie es fast geschafft. Sie müssen jetzt nur noch einen abschließenden Fragebogen ausfüllen. Fragebogen wird mit Klemmbrett und Stift an den Probanden weitergegeben.

Vielen Dank!

Anwendungsbeispiel: _____

Geistige Anforderungen:



Körperliche Anforderungen:



Zeitliche Anforderungen:



Leistung:



Anstrengung:



Frustration:



Bitte vergegenwärtigen Sie sich jetzt noch ein Mal die Nutzung des Produkts und **beschreiben Sie Ihr Erleben der Nutzung mit Hilfe der folgenden Aussagenpaare**. Die Paare stellen jeweils extreme Gegensätze dar, zwischen denen eine Abstufung möglich ist.

Vielleicht passen einige Aussagen nicht so gut, kreuzen Sie aber trotzdem bitte immer an, welcher Begriff Ihrer Meinung nach eher zutrifft. Denken Sie daran, dass es keine "richtigen" oder "falschen" Antworten gibt - nur Ihre persönliche Meinung zählt!

1 2 3 4 5 6 7

Bei der Nutzung (des Produkts)...

...handelte ich überlegt	<input type="checkbox"/>	...handelte ich spontan	G_01						
...erreichte ich mein Ziel nur mit Anstrengung	<input type="checkbox"/>	...erreichte ich mein Ziel mit Leichtigkeit	M_01						
...handelte ich unbewusst, ohne lange über die einzelnen Schritte nachzudenken	<input type="checkbox"/>	...führte ich bewusst einen Schritt nach dem anderen aus	P G_02						
...ließ ich mich von meinem Verstand leiten	<input type="checkbox"/>	...ließ ich mich von meinem Gefühl leiten	G_03						
...war ich orientierungslos	<input type="checkbox"/>	...konnte ich mich gut zurechtfinden	M_02						
...handelte ich ohne dabei nachzudenken	<input type="checkbox"/>	...konnte ich jeden Schritt genau begründen	P G_04						

Die Nutzung (des Produkts)...

...erforderte viel Aufmerksamkeit	<input type="checkbox"/>	...ging wie von selbst	M_03						
...war begeisternd	<input type="checkbox"/>	...war unbedeutend	P X_01						
...war einfach	<input type="checkbox"/>	...war schwierig	P M_04						
...war nichts Besonderes	<input type="checkbox"/>	...war ein magisches Erlebnis	X_02						
...war sehr intuitiv	<input type="checkbox"/>	...war gar nicht intuitiv	P INT_01						
...war belanglos	<input type="checkbox"/>	...war mitreißend	X_03						
...fiel mir leicht	<input type="checkbox"/>	...fiel mir schwer	P M_05						
...war faszinierend	<input type="checkbox"/>	...war trist	P X_04						

Im Nachhinein...

...fällt es mir schwer, die einzelnen Bedienschritte zu beschreiben	<input type="checkbox"/>	...ist es für mich kein Problem, die einzelnen Bedienschritte zu beschreiben	V_01						
...kann ich mich gut an die Bedienung erinnern	<input type="checkbox"/>	...fällt es mir schwer, mich zu erinnern, wie das Produkt bedient wird	P V_02						
...kann ich nicht sagen, auf welche Art und Weise ich das Produkt bedient habe	<input type="checkbox"/>	...kann ich genau sagen, auf welche Art und Weise ich das Produkt bedient habe	V_03						

Fragebogen zur Studie

VP ID: _____

Können Sie sich diese Art der Interaktion während dem Fahren vorstellen?

nein, weil _____

Ja, weil _____

Wo sehen Sie die **Vorteile** dieser Art von Interaktion bei:

Touch: _____

Sprache: _____

Geste: _____

Wo sehen Sie die **Nachteile** dieser Art von Interaktion bei:

Touch: _____

Sprache: _____

Geste: _____

Sonstige Kommentare, Ideen, Anregungen?:

Vielen Dank ☺