



Verena Rexeis

Usability Benchmark und Aktivitäts-Analyse mit Eye Tracking von Mobile Augmented Reality unterstützten Ernährungsempfehlungen

DIPLOMARBEIT

zur Erlangung des akademischen Grades

Magistra der Naturwissenschaften

Lehramtsstudium Unterrichtsfach Informatik und Informatikmanagement/Mathematik

eingereicht an der
Technischen Universität Graz

Betreuer

Univ.-Doz. Dipl.-Ing. Dr.techn. Martin Ebner

Institut für Informationssysteme und Computer Medien (IICM)

Graz, im August 2015

Kurzfassung

Die von JOANNEUM RESEARCH entwickelte App MANGO ist wegbereitend für technologieunterstützte, wissenschaftlich fundierte Ernährungsempfehlungen. Sie ermöglicht den Benutzerinnen und Benutzern den Zugang zu ihrer individuell angepassten Ernährungsform im Ernährungskonzept „functional eating“. Zentrales Element der App ist die Lebensmittelerkennung samt Ernährungsempfehlungen mittels mobiler Augmented Reality.

In der vorliegenden Arbeit wurde ein erster Usability-Test zur Benchmark-Erstellung für die beschriebene App MANGO durchgeführt. Außerdem wurde eine Aktivitäts-Analyse der Testpersonen bei der Verwendung der Lebensmittelerkennung durchgeführt, wobei Eye-Tracking-Daten erhoben und ausgewertet wurden. Nach der Testung wurde von den Testpersonen ein Fragebogen zur Bewertung der Usability ausgefüllt. Dieser bestand aus dem System-Usability-Scale-Fragebogen zur Bewertung der Usability und dem User Experience Questionnaire zur Bewertung der User Experience. Zusätzlich wurden die Testpersonen vor und nach dem Test befragt, wie wahrscheinlich sie die App herunterladen und verwenden würden.

Für die Bewertung der Anwendungs-Szenarien der App wurden messbare Kernaufgaben definiert und Zielvorgaben für deren Abschluss festgelegt. Darüber hinaus wurden drei Gesamt-Ziele in den Benchmark aufgenommen. Die Zielerreichung wurde nach der Testung durch die Analyse von quantitativen Daten, die aus den Szenen-Videos der Eye Tracking Glasses gewonnen wurden, überprüft.

Aus den Fragebogen-Bewertungen konnte eine positive Beurteilung der Testpersonen die App betreffend festgestellt werden, wobei die Steuerbarkeit, vornehmlich durch Abstürze der Beta-Version, als unterdurchschnittlich bewertet wurde. Die Gesamtziele im Usability Benchmark wurden fast vollständig erfüllt, jedoch konnten die zu den Kernaufgaben definierten Ziele zu einem Großteil nicht erfüllt werden. Auffällig war, dass die Akzeptanz der App zwischen den männlichen und weiblichen Testpersonen sehr unterschiedlich ausfiel.

Im Laufe des Tests konnten ua. durch die Aktivitäts-Analyse mehrere Punkte identifiziert werden, die an Steuerung und Feedback der App verbessert werden können und die nun in die Weiterentwicklung der App einfließen werden.

Usability Benchmark and Activity Analysis Using Eye Tracking on Nutrition Assistance Supported by Mobile Augmented Reality

Abstract

JOANNEUM RESEARCH developed the application MANGO which pioneers at the field of technology-enhanced, scientific nutrition assistance. The app allows its users to easily gain information about their individual nutrition within the “functional eating” diet. The key component is integrated in the camera mode where users get situated dietary information assistance via augmented reality functionality.

Within this thesis a first usability test of the previously described app MANGO was conducted to get benchmark information. In addition an activity analysis of the users administering the scanning functionality was performed integrating eye tracking. The usability test closed with a collection of post-test questionnaires, i.e. the System Usability Scale and the User Experience Questionnaire combined with a question, which was already rated in the pre-test, namely, how likely is it that the users will download and use the app.

Measurable core tasks to the chosen scenarios were specified and goals for the benchmark identified. The benchmark also included three overall goals. Quantitative data gathered from the eye tracking glasses' scene videos was the basis to verify the goals.

Questionnaire ratings from the users showed positive connotation to the app, whereas only dependability was rated sub-standard probably referable to crashes of the beta-version. Overall goals within the usability benchmark were accomplished satisfactorily. However, reaching the goals referring to core tasks failed mostly. Acceptance turned out to be significantly different for men and women.

Activity analysis carried out throughout this test brought up improvements regarding control and feedback which will enhance further development.

Eidesstattliche Erklärung

Ich erkläre an Eides statt, dass ich die vorliegende Arbeit selbstständig verfasst, andere als die angegebenen Quellen/Hilfsmittel nicht benutzt, und die den benutzten Quellen wörtlich und inhaltlich entnommene Stellen als solche kenntlich gemacht habe. Das in TUGRAZonline hochgeladene Textdokument ist mit der vorliegenden Diplomarbeit identisch.

Graz, im August 2015

Verena Rexeis

Statutory Declaration

I declare that I have authored this thesis independently, that I have not used other than the declared sources/resources, and that I have explicitly marked all material which has been quoted either literally or by content from the used sources. The text document uploaded to TUGRAZonline is identical to the present diploma thesis.

Graz, August 2015

Verena Rexeis

Danksagung

An dieser Stelle möchte ich mich bei allen bedanken, die mich während meines Studiums und dem Erstellen dieser Diplomarbeit unterstützt haben.

Mein besonderer Dank gilt den zwei Personen, ohne die diese Arbeit nicht dieselbe wäre. Lucas Paletta, der mir ermöglicht hat die App MANGO zu testen und somit diese Arbeit zu erstellen, und Martin Ebner, der sofort bereit war meine Betreuung zu übernehmen und sich immer für Fragen von mir Zeit genommen hat und so alle Unklarheiten beseitigen konnte.

Bedanken möchte ich mich auch bei Michael, Martin, Stefan, Anna, Patrick und Irene, ohne die mein Zeitplan nicht hätte eingehalten werden können, und bei Nina, Chrisi und allen Anderen, die sich Zeit genommen haben, an der Testung teilzunehmen.

Ein Großer Dank gilt meinen Arbeitskolleginnen vom BGF der TU Graz, die mich in letzter Zeit nicht viel gesehen haben und trotzdem für mich da waren. Lissa, Gudrun – danke! Chrisi, du warst eine sehr große Hilfe und danke auch dir Barbara, nun ist es endlich geschafft.

Am wichtigsten ist es mir jedoch, mich bei meiner Familie zu bedanken. Danke Mama und Papa für all eure Unterstützung, die mich hierhin geführt hat – finally! Danke Susi, dass du immer für mich da bist, mir immer den richtigen Weg zeigst und für hitzige Diskussionen immer Zeit übrig hast. Du siehst, du wirst für mich IMMER wichtig sein!

Stefan, ohne dich und deine deine Unterstützung hätte alles nicht geklappt. Danke, dass du mir den Rücken freigehalten hast. Ich freue mich schon so darauf, wieder High Heels zu tragen. Das wichtigste: es ist unbeschreiblich schön, dass es dich in meinem Leben gibt!

Verena Rexeis
Graz, im August 2015

Inhaltsverzeichnis

Abkürzungsverzeichnis	v
Abbildungsverzeichnis	vii
Tabellenverzeichnis	ix
1. Einleitung	1
1.1. Motivation	1
1.2. Forschungsfrage	2
1.3. Struktur der Arbeit	2
2. Usability Engineering	5
2.1. Definition von Usability	6
2.1.1. Usability-Attribute	8
2.1.2. User Experience	9
2.2. Usability Engineering Lifecycle	9
2.3. Methoden der Usability-Evaluation	13
2.3.1. Usability-Inspektion	14
2.3.2. Usability-Testungen	16
3. Messen von Usability	21
3.1. Performance-Metriken	23
3.1.1. Task Success	24
3.1.2. Time on Task	24
3.1.3. Fehler	25
3.1.4. Effizienz	25
3.1.5. Erlernbarkeit	26

3.2.	Selbst-rückgemeldete Metriken	26
3.2.1.	System Usability Scale	27
3.2.2.	User Experience Questionnaire	29
3.3.	Problem-basierende Metriken	32
3.4.	Verhaltensbezogene und physiologische Metriken	32
3.5.	Kombinierte und vergleichende Metriken	33
4.	Eye Tracking	35
4.1.	Grundlagen des Eye Tracking	35
4.1.1.	Augenbewegungen	35
4.1.2.	Eye-Tracking-Technologien	36
4.1.3.	Mögliche Problembereiche	37
4.2.	Eye Tracking in Usability-Testungen	38
4.3.	Mobiles Testen	41
4.4.	Ausgewählte Forschungsarbeiten	42
5.	Seven Stages of Action	45
5.1.	Grundlegende Prinzipien der Interaktion	45
5.1.1.	Affordanzen	46
5.1.2.	Signifiers	46
5.1.3.	Mapping	47
5.1.4.	Feedback	47
5.1.5.	Konzeptionelle Modelle	48
5.1.6.	Einschränkungen und Konventionen	48
5.2.	Seven Stages of Action	49
5.3.	SSA: Sieben grundlegende Design-Prinzipien	51
6.	Forschungsmaterialien bzw. -methoden	55
6.1.	Testumgebung „MANGO-App“	55
6.2.	Fragebogen „Vorbefragung“	57
6.3.	Aufgabenbeschreibung	61
6.4.	Fragebogen „App-Bewertung“	63
6.5.	Auszuwertende Parameter	64
6.5.1.	Usability Benchmark	65
6.5.2.	Eye-Tracking-Parameter	69

7. Empirische Studiendurchführung	73
7.1. Untersuchungskonzept	73
7.1.1. Studiendesign	73
7.1.2. Testpersonen	74
7.1.3. Hardwarebeschreibung	75
7.1.4. Räumlichkeit	76
7.1.5. Versuchsaufbau	79
7.1.6. Studienziel	80
7.2. Studienablauf	81
7.3. Pre-Test	83
8. Ergebnisse	85
8.1. Usability Benchmark	88
8.2. Post-Test-Fragebögen	96
8.3. Eye Tracking und die Seven Stages of Action	98
8.3.1. Aktionsbestimmung mit den SSA	98
8.3.2. Eye-Tracking-Ergebnisse	100
9. Diskussion und Erkenntnisse	109
9.1. Gesamtbewertung der App	109
9.2. Identifizierte Probleme	111
9.3. Eye Tracking	112
9.4. Limitation der Arbeit	113
9.5. Beantwortung der Forschungsfrage	114
10. Zusammenfassung und Ausblick	117
10.1. Ausblick	118
Literatur	119
Anhang	125
A. Unterlage „Vorbefragung“	126
B. Unterlage „Aufgaben der Eye Tracking Studie“	131
C. Unterlage „App-Bewertung“ (Bewertungsskalen)	132
D. Zustimmungserklärung	138

E. Checkliste zum Studienablauf 140

Abkürzungsverzeichnis

EOG	Elektro-OculoGraphie
EST	Institut für Ernährung und Stoffwechselerkrankungen GesmbH
ETG	Eye Tracking Glasses
FE	„functional eating“ (Ernährungsphilosophie)
FFG	Österreichische Forschungsförderungsgesellschaft
HCI	Human-Computer Interaction
HMD	Head Mounted Displays
JR	JOANNEUM RESEARCH Forschungsgesellschaft mbH
KA	Kernaufgabe
MANGO	Mobile Augmented Reality for Nutrition Guidance and Food Awareness
POG	Photo-OculoGraphie
ROI	Return on Investment
SMET	Smartphone Eye Tracking Toolbox
SSA	Sieben/Seven Stages of Action
SUM	Single Usability Metric
SUS	System-Usability-Scale-Fragebogen
UEQ	User Experience Questionnaire
UI	User Interface
UX	User Experience
VOG	Video-OculoGraphie

Abbildungsverzeichnis

2.1. Usability-Merkmale im Kontext der allgemeinen Systemakzeptanz . . .	7
2.2. User Experience und Usability aus Sicht der ISO 9241	10
2.3. Übersicht Usability-Evaluationsmethoden	14
3.1. SUS-Score Interpretation	29
3.2. Skalenstruktur des UEQ	31
4.1. Beispiel für eine Heat-Map-Darstellung	40
4.2. Beispiel eines Gaze Plots	41
5.1. Die Klüfte der Durchführung und Evaluation	50
5.2. Der Seven Stages of Action Cycle	51
5.3. Die Seven Stages of Action als Design-Hilfen	52
6.1. Die 35 Lebensmittel-Klassen des Klassifikators	56
6.2. Startbildschirm und Hauptfunktionen der MANGO-App	58
6.3. Ergebnis eines Lebensmittel-Scans	58
6.4. Detailansichten eines Lebensmittels	59
6.5. FE-Typenbeschreibungen umgesetzt in der App	59
6.6. SMET Interface mit laufender Erkennung für Daten von T04	71
7.1. SMI™ Eye-Tracking-Glasses	75
7.2. Motion Tracker und Affectiva-Armband	77
7.3. Übersicht der Einrichtung im Human Factors Labor	78
7.4. Schachbrettmuster mit Markern zur Kalibrierung der ETG	78
7.5. Typische Situation während dem praktischen User-Test	79
7.6. Fertiger Versuchsaufbau zur Präsentation der Lebensmittel	81

8.1. Verteilung der Testpersonen auf die FE-Typen	86
8.2. Ergebnis: Würden die TP die App herunterladen und verwenden? . .	87
8.3. Erfüllungsgrade der Usability-Benchmark-Zielvorgaben	91
8.4. User Performance bei Kernaufgaben und Benchmark-Zielen	91
8.5. Mittlere Dauern der Kernaufgaben	92
8.6. Anteil der Testpersonen mit Kernaufgaben-Dauer weniger einer Minute	92
8.7. Absturzhäufigkeit der App im Kameramodus	93
8.8. Anzahl Scans und Anteil der Falsch-Erkennungen in KA 3	96
8.9. SUS-Score und Einzelbewertungen	97
8.10. UEQ-Ergebnis	97
8.11. UEQ-Benchmark-Einordnung	98
8.12. Durchschnittliche Dauer der SSA	102
8.13. Durchschnittliche Blickdauern in den SSA beim Scannen	103
8.14. Durchschnittliche Fixationen/Sekunde in den SSA beim Scannen . . .	104
8.15. Lebensmittel-Erkennungen der App und Anzahl der Abstürze	105
8.16. Cluster-Einflussgrößen für die Heat-Map-Erstellung	106
8.17. Häufigkeit der Iterationen der SSA	107
8.18. Teilweise Erkennung von Daten von T04 durch SMET	108

Tabellenverzeichnis

3.1. Gegenüberstellung der originalen und verwendeten SUS Aussagen . . .	28
3.2. Beispiel eines semantischen Differentials wie im UEQ	30
5.1. Die Sieben Stages of Action	50
5.2. Sieben grundlegende Design-Prinzipien	53
6.1. Beschreibung der Hauptfunktionen der MANGO-App	57
6.2. Szenario-Entwicklung der Testung	61
6.3. Kernaufgaben-Unterteilung	66
6.4. Usability-Benchmark-Zielvorgaben	68
7.1. Auswahl der Lebensmittel für die Studie	80
8.1. Usability-Benchmark-Ergebnisse	89
8.2. Gesamtbewertung der MANGO-App durch die Testpersonen	94
8.3. Gesamt-User-Performance in der Testung	95

1. Einleitung

Es gibt verschiedenste Ernährungskonzepte und allgemeine Ernährungsempfehlungen, welche auf unterschiedliche Weise zugänglich gemacht werden. Sei es in gedruckter Form als Buch oder Information, die online zur Verfügung steht. Interessierte Personen wollen die Informationen zu ihrer Ernährung rasch aufrufen können, was trotz der vielfältigen Wege der Informationsgewinnung nicht immer einfach ist. Will man spontan einkaufen gehen hat man keine Gewähr, ob das Gekaufte auch tatsächlich empfohlen ist, außer es ist entsprechende Literatur parat, oder man findet sich in der Fülle der online verfügbaren Informationen zurecht. Zwischen Büchern und Internetrecherchen entsteht eine Kluft der praktischen Handhabung von Informationen, und kompakte Applikationen, welche mit einen wissenschaftlichen Hintergrund entwickelt wurden, gibt es kaum (Waltner et al., 2015).

1.1. Motivation

Diese zuvor beschriebene Lücke wird durch die von JOANNEUM RESEARCH (JR) entwickelte Smartphone Applikation MANGO geschlossen. Die App wurde im gleichnamigen Projekt MANGO – Mobile Augmented Reality for Nutrition Guidance and Food Awareness, gefördert durch die Österreichische Forschungsförderungsgesellschaft (FFG), entwickelt. Ihr liegt die Ernährungsphilosophie „functional eating“ (FE) zugrunde, welche wiederum von Prim. Dr. med. univ. Meinrad Lindschinger (Institut für Ernährung und Stoffwechselerkrankungen – EST) entwickelt wurde. FE beachtet geschlechterrelevante und weitere Diversitäts-Aspekte sowie differenzierte Stress-Profile (in Hinblick auf Arbeit, Familie oder Sport) und besteht aus sieben individuellen Ernährungssäulen (-typen): Balanced Food, Beauty Food, Brain Food,

Erotic Food, New Food, Power Food und Soul Food (Waltner et al., 2015). Vier davon, nämlich Beauty Food, Brain Food, Power Food und Soul Food, werden in der App abgebildet.

Damit diese App für an individueller Ernährung interessierten Personen tatsächlich den Anspruch erfüllt, rasch und einfach bedient werden zu können, soll die App auf ihre Usability getestet werden. Die gewonnen Erkenntnisse werden in die weitere Entwicklung der App miteinbezogen. Da es sich derzeit um eine Beta-Version der App handelt, werden quantitative Daten erhoben, um in folgenden Testungen die Entwicklung der User Experience (UX) und der Usability der App erheben zu können.

1.2. Forschungsfrage

Die in dieser empirischen Studie erhobenen quantitativen Daten werden mit Usability-Benchmark-Zielvorgaben bewertet und zusammengefasst. Außerdem wird eine Aktivitäts-Analyse mit Eye-Tracking-Unterstützung während dem Verwenden der Lebensmittelerkennung durchgeführt. Die daraus abgeleitete Forschungsfrage lautet:

Können mit Usability Benchmark und Aktivitäts-Analyse durch Eye-Tracking-Unterstützung vertiefende Kenntnisse für App-Entwicklung im Bereich Ernährungsempfehlung gewonnen werden?

1.3. Struktur der Arbeit

Die vorliegende Diplomarbeit beginnt mit den dazugehörigen theoretischen Grundlagen, die in Kapitel 2 bis 5 mit Usability Engineering, dem Messen von Usability, Grundlagen zu Eye Tracking und aktuellen Forschungsarbeiten, sowie der Erarbeitung der sieben Stages of Action dargelegt werden. Die in dieser Arbeit durchgeführten Leistungen beginnen ab Kapitel 6 mit der Auswahl der Methoden und der Beschreibung der verwendeten Materialien und in Kapitel 7 wird die eigentliche Durchführung der empirischen Studie beschrieben. Alle Ergebnisse dieser Studie

werden in Kapitel 8 präsentiert und in Kapitel 9 interpretiert. Schließlich wird diese Diplomarbeit noch einmal zusammengefasst (Kapitel 10) und ein Ausblick auf künftige Arbeiten gegeben.

2. Usability Engineering

„Usability ist der Grad an Qualität, in welchem der Benutzer/die Benutzerin die Interaktion mit etwas erlebt.“

— Jakob Nielsen, vgl. Rampl (2007)

Als Usability Engineering bezeichnet man den iterativen Prozess zur Verbesserung eines Systems, genauer gesagt die Verbesserung der Usability einer Mensch–Computer-Schnittstelle (Andrews, 2015) wie dem User Interface (UI). Usability Engineering ist fachlich der seit rund 1970 bestehenden wissenschaftlichen Disziplin der Human–Computer Interaction (HCI) zuzuschreiben und hat zum Ziel Lösungen zu erstellen, die auf die Benutzerinnen und Benutzer (User) zugeschnitten sind (Richter et al., 2013). Dazu müssen

- Informationen über die Benutzerinnen und Benutzer, deren Arbeitsabläufe, Bedürfnisse, Anforderungen, Aufgaben und Umgebung systematisch in die Software- oder Produktentwicklung einfließen und
- die technischen Möglichkeiten, Grenzen und Rahmenbedingungen in einer verständlichen Form an die Benutzerinnen und Benutzer zurückfließen (Richter et al., 2013).

Mayhew (1999) hebt hervor, dass für Benutzerinnen und Benutzer das User Interface für *das Produkt* steht, Usability eben eine messbare Eigenschaft eines UI ist, die stärker oder weniger stark ausgeprägt ist, und Usability Engineering geeignete Methoden bereitstellt, um Usability im UI-Design während der Produktentwicklung zu erreichen. Usability Engineering entstammt einigen anderen wissenschaftlichen Disziplinen, darunter kognitive Psychologie, experimentelle Psychologie, Ethnografie und Softwaretechnologie. Dies wird deutlich, wenn man die Faktoren betrachtet, die beim UI-Design u. a. berücksichtigt werden müssen:

- kognitive, motorische und wahrnehmende Fähigkeiten und Einschränkungen von Menschen allgemein
- besondere und einzigartige Charakteristika der angestrebten User-Population im Speziellen
- einzigartige Charakteristika der physischen und sozialen Umwelt der Benutzerinnen und Benutzer
- einzigartige Charakteristika und Anforderungen der User-Anwendungen, die vom Produkt unterstützt werden
- einzigartige Möglichkeiten und Einschränkungen der gewählten Software und/oder Hardware und Plattform des Produktes (Mayhew, 1999)

Bevor auf die Umsetzung im Usability Engineering Lifecycle (siehe Kapitel 2.2) eingegangen wird, vorher noch eine Zusammenfassung von Definitionen von Usability.

2.1. Definition von Usability

Mit zunehmender Multifunktionalität von technischen Systemen, bei denen man früher der Meinung war, dass lediglich die Art und die Anzahl der zur Verfügung stehenden Funktionen entscheidend sei, ob ein System genutzt werde oder nicht, kam die Frage der „Benutzerfreundlichkeit“ auf und bald wurde dieses Konstrukt zugunsten der Usability (dt. „Gebrauchstauglichkeit“) fallen gelassen, da hier neben komfortabler Benutzung auch die Zielerreichung durch die Benutzerinnen und Benutzer im Vordergrund steht (Sarodnick et al., 2011).

Seit 1997 ist in der ISO 9241, die 2006 von „Ergonomische Anforderungen für Bürotätigkeiten mit Bildschirmgeräten“ in „Ergonomie der Mensch-System-Interaktion“ umbenannt und um einige Teile erweitert wurde, die Usability eines Produktes im Teil 11 „Anforderungen an die Gebrauchstauglichkeit“ als „das Ausmaß, in dem ein Produkt, System oder ein Dienst durch bestimmte Benutzer in einem bestimmten Anwendungskontext genutzt werden kann, um bestimmte Ziele effektiv, effizient und zufriedenstellend zu erreichen“ definiert (DIN EN ISO 9241-11, 1999).

Die drei, durch die DIN EN ISO 9241-11 (1999) definierten, messbaren Usability-Merkmale sind (Andrews, 2015):

- *Effektivität*: Die Genauigkeit und Vollständigkeit mit der ein Benutzer/eine Benutzerin festgelegte Ziele erreicht.
- *Effizienz*: Die verwendeten Ressourcen in Relation zur Genauigkeit und Vollständigkeit mit der ein Benutzer/eine Benutzerin festgelegte Ziele erreicht.
- *Zufriedenheit*: Fehlen von Unbehagen und positive Haltung gegenüber der Verwendung des Produkts.

Nielsen (1993) definiert Usability als ein qualitatives Merkmal, also wie leicht oder schwer eine Benutzeroberfläche oder ein System zu verwenden ist, und zeigt in einem Modell, wie in Abbildung 2.1 veranschaulicht, die Merkmale der Usability im Kontext der allgemeinen Systemakzeptanz: Unter der Annahme, dass ein System sozial akzeptiert ist kann man die praktische Akzeptanz mit mehreren klassischen Kategorien wie Kosten, Kompatibilität, Zuverlässigkeit, sowie der Nützlichkeit (Usefulness) analysieren. Bei genauerer Betrachtung sieht man, dass man die Nützlichkeit in zwei weitere Begriffe spalten muss, Funktionalität und Usability. Wobei die Funktionalität meint, ob das System dazu verwendet werden kann das festgelegte Ziel zu erreichen, und Usability dafür steht, wie gut Benutzerinnen und Benutzer diese Funktionalität dann tatsächlich nutzen können.

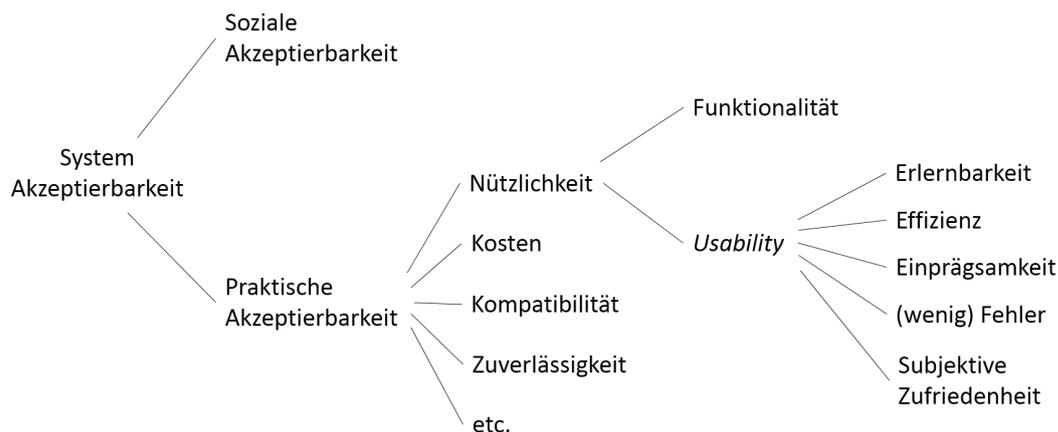


Abbildung 2.1.: Usability-Merkmale im Kontext der allgemeinen Systemakzeptanz, nach Nielsen (1993, S. 25)

Bei Nielsen (1993) bedingen sich Funktionalität und Usability gegenseitig, da es beiderseitig problematisch ist, wenn einerseits einfache Verwendung nicht zum gewünschten Ziel führt, also ein Problem bei der Funktionalität vorliegt, andererseits aber das System zum gewünschten Ziel führen kann, aber z. B. das Interface zu kompliziert ist, also ein Problem im Bereich der Usability vorliegt. Usability wird laut Nielsen (1993) in fünf messbare Attribute zerlegt (siehe Kapitel 2.1.1): Effektivität, Erlernbarkeit, Effizienz, Einprägsamkeit, Fehler und Zufriedenheit.

2.1.1. Usability-Attribute

Kombiniert man die drei in der DIN EN ISO 9241-11 (1999) definierten Merkmale mit den fünf von Nielsen (1993) (siehe Abbildung 2.1) erhält man sechs messbare Usability-Merkmale (Andrews, 2015):

- *Effektivität (Effectiveness)*: Die Vollständigkeit mit der Benutzerinnen und Benutzer festgelegte Ziele erreichen.
- *Erlernbarkeit (Learnability)*: Die Leichtigkeit des Erlernens eines *unerfahrenen* Benutzers/einer *unerfahrenen* Benutzerin, um grundlegende Aufgaben zu erfüllen.
- *Effizienz (Efficiency)*: Die Performance auf hohem Niveau eines *erfahrenen* Benutzers/einer *erfahrenen* Benutzerin.
- *Einprägsamkeit (Memorability)*: Die problemlose Wiederverwendung nach längerer Nichtverwendung eines *gelegentlichen* Benutzers/einer *gelegentlichen* Benutzerin.
- *Fehler (Errors)*: Die Fehlerrate bei der Benutzung des Systems, von leichten bis zu schwerwiegenden Fehlern.
- *Zufriedenheit (Satisfaction)*: Wie zufriedenstellend die Benutzung des Systems für Benutzerinnen und Benutzer ist.

Auf das Messen von Usability wird im gleichnamigen Kapitel 3 detaillierter eingegangen.

2.1.2. User Experience

Ein eng verwandter Begriff zur Usability ist die *User Experience* (UX) (Sarodnick et al., 2011). Wird zwischen den Begriffen Usability und UX unterschieden, gilt Usability meist als die Befähigung eines Benutzers/einer Benutzerin eine Aufgabe erfolgreich durchführen zu können und UX betrachtet hingegen die gesamte Interaktion eines Benutzers/einer Benutzerin mit dem System, dies beinhaltet entstandene positive wie negative Gefühle, Gedanken, Meinungen, Vorlieben, Sinneswahrnehmungen und physische sowie psychologische Reaktionen (Tullis et al., 2013; Sarodnick et al., 2011). UX beinhaltet für Tullis et al. (2013) drei definierende Charakteristiken:

- Ein Benutzer/eine Benutzerin ist beteiligt.
- Dieser Benutzer/diese Benutzerin interagiert mit einem Produkt, System oder irgendwie mit einem User Interface.
- Das Erlebnis der Benutzerinnen und Benutzer ist von Interesse und beobachtbar oder messbar.

UX betrachtet laut der DIN EN ISO 9241-210 (2011) nicht nur die tatsächliche Nutzung eines Systems, sondern auch die antizipierte Nutzung sowie die Verarbeitung der Nutzungssituation (Identifikation oder Distanzbildung) im Anschluss an die vollzogene Nutzung (siehe Abbildung 2.2). Arbeitssysteme werden meist nach Maßstäben der Usability bewertet, wohingegen die Bewertung von UX eher für Produkte für Endverbraucher von Bedeutung ist (Sarodnick et al., 2011).

2.2. Usability Engineering Lifecycle

Produkte können im Usability Engineering einen Lebenszyklus durchlaufen, der als Usability Engineering Lifecycle zusammengefasst werden kann, welcher laut Nielsen (1993) angepasst von Andrews (2015) wie nachfolgend aufgebaut sein kann:

1. Kennen des Benutzers/der Benutzerin
2. Usability-Benchmark-Erstellung
3. Zielorientiertes Interaktions-Design

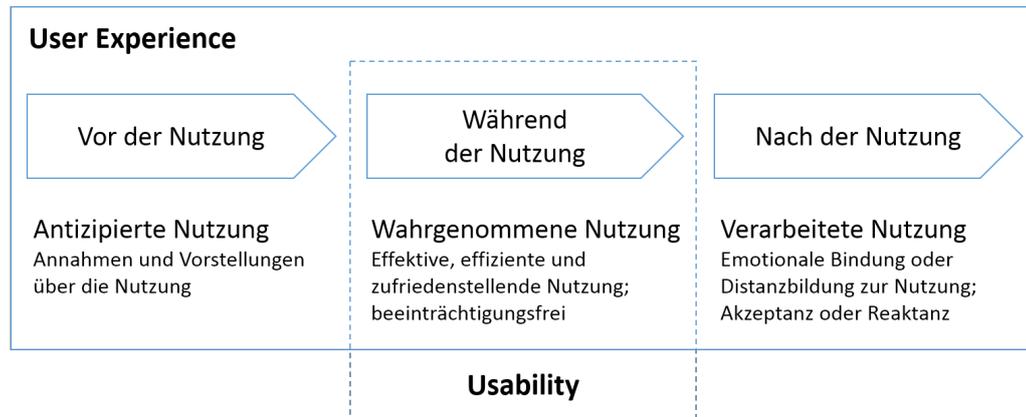


Abbildung 2.2.: User Experience und Usability aus Sicht der ISO 9241, nach Sarodnick et al. (2011, S. 22)

4. Iteratives Design:

- a) Prototyping
- b) Formative Usability Evaluation (Inspektion und/oder Tests)

5. Summative Usability Evaluation

6. Follow-up-Studien

Die Phasen im Usability Engineering Lifecycle werden nun überblicksartig nach Andrews (2015) beschrieben, wobei auf Usability Benchmarking näher eingegangen wird und Usability-Evaluationsmethoden im nachfolgenden Kapitel beschrieben werden.

Zu Beginn steht mit *Kennen des Benutzers/der Benutzerin* der Prozess des genauen Identifizierens des zukünftigen Benutzers/der zukünftigen Benutzerin. In der Praxis bedeutet dies qualitative Recherche, die das Beobachten sowie das Befragen der Benutzerinnen und Benutzer beinhaltet, um aus den so gewonnenen Informationen eine Unterteilung z. B. aufgrund von Erfahrung, vorzunehmen. Für jede potentielle Kategorie von Benutzerinnen und Benutzern wird ein zugehöriges Profil entwickelt, das im Allgemeinen auf demografischen und verhaltensbezogenen Variablen basiert. Wurden die Ziele und Verhaltensweisen der Benutzerinnen und Benutzer identifiziert können typische User-Szenarien erstellt werden.

Als zweites folgt im Usability Engineering Lifecycle das *Usability Benchmarking*, bei dem, nach Andrews (2015), einige grundlegende Fragen beantwortet werden müssen. Zur ersten Frage, „Wie nützlich ist der Vergleich mit konkurrierenden Systemen?“, muss der derzeitige „State of the Art“ eruiert und die Entscheidung getroffen werden, wie hoch im Vergleich dazu die Ziele gesetzt werden. Die zweite Frage, „Wie viel besser sollte das eigene Interface sein?“, bedeutet, dass im Vorhinein über Usability-Metriken und die angestrebten Level an messbarer Usability entschieden werden. Zur Beantwortung der dritten Frage, „Was ist der wahrscheinliche Ertrag des Aufwandes?“, wird der Return on Investment (ROI) errechnet, indem eine Finanzfolgeabschätzung (engl. finance impact analysis) durchgeführt wird.

Usability Benchmarks können aber auch für den Vergleich von fertig entwickelten Produkten herangezogen werden, da sie quantitative Daten der User Performance liefern. Von entscheidender Bedeutung ist die Auswahl der Kernaufgaben, zumal nur bei sinnvoller Wahl, repräsentative Ergebnisse erzielt werden können. Nach dem Festlegen der Liste mit Kernaufgaben muss diese noch einmal sorgfältig geprüft werden, diese sollte maximal zehn bis fünfzehn Kernaufgaben umfassen, die innerhalb von maximal zwei Stunden erfüllt werden können sollten. Eine gute Auswahl der Kernaufgaben hat auch eine positive Auswirkung auf zukünftige Usability Benchmarks, da man die Bereiche mit erhöhtem Zeitbedarf bei Laborversuchen schon identifiziert hat. Des Weiteren ermöglicht dies bei weiteren Entwicklungen zu fragen: „Welche Kernaufgabe wird dieses Feature berühren/verbessern?“. Soll keine Kernaufgabe von der Neuentwicklung betroffen sein, ist zu überlegen, ob sich der Aufwand des Implementierens überhaupt lohnt. Im nächsten Schritt muss definiert werden, was gemessen werden soll. Im Allgemeinen gibt es drei wichtige Metriken: Erfolg/Misserfolg mit Zeitangabe, Dauer der Aufgabe, Anzahl der Fehler bis zum Abschluss der Aufgabe. Oft wird ein überraschend hoher Prozentsatz der Kernaufgaben nicht oder nur teilweise erfüllt. Bei der erstmaligen Bestimmung eines Usability Benchmarks zu einem gegebenen Produkt scheinen die gesetzten Ziele oft willkürlich, zumal keine Referenzwerte existieren, auf die Bezug genommen werden kann. Abhilfe schaffen kann ein Referenzwert, erzeugt aus den erzielten Ergebnissen der entwickelnden Personen, der natürlich um einen Faktor angepasst werden muss. Alternativ kann als Referenz die Performance eines Konkurrenzproduktes herangezogen werden. Typische Ziele sehen wie folgt aus: *Kernaufgabe: Finden sie Lebensmittel*

2. Usability Engineering

X in der Anwendung; Ziel: 75 % aller Testpersonen werden die Kernaufgabe innerhalb von 3 Minuten abschließen können. Es können des Weiteren verfeinerte Ziele definiert werden. Beim Auswerten der Daten ist es wichtig, sich hauptsächlich auf die Ergebnisse der definierten Ziele zu konzentrieren, da sowohl Management als auch das Entwickler-Team daran das größte Interesse haben. (Berkun, 2003)

Darauf folgt laut Andrews (2015) das *Goal-Oriented Interaction Design*. Interaction Design ist vom, rein durch Coding charakterisiertem, Interaction Design zu unterscheiden, da beim Interaction Design die Funktionalität, das Verhalten und die End-Präsentation von hohem Stellenwert sind. Goal-Oriented bezieht sich auf das Berücksichtigen der Ziele der Menschen, die durch konkrete Schritte erreicht werden können. Die notwendigen Schritte können sich mit der Technologie verändern, die übergeordneten Ziele bleiben jedoch bestehen. Der Interaction-Design-Prozess beginnt wieder mit User-Interviews aus denen konkrete Personas entwickelt werden, worauf deren Ziele definiert und konkrete Szenarien entwickelt werden. Darauf basierend wird eine Design-Lösung entwickelt, die beispielsweise aus parallelem Design entsteht, bei dem mehrere Design-Teams unabhängig von einander Entwürfe erstellen. User-Meinungen können sehr hilfreich sein, wenn es darum geht Bezeichnungen für Funktionen auszuwählen.

Nachfolgend setzt der *iterative Design-Prozess* auf, also das *Prototyping*, worauf immer eine *formative Usability-Evaluation* folgt. Der Kerngedanke kann mit „Design, Test, Re-Design“ zusammengefasst werden. Die verschiedenen, in ihrer Komplexität zunehmenden, aufeinander aufbauenden Prototypen sind: Verbale Prototypen – mündliche Beschreibungen von Wahlmöglichkeiten und deren Ergebnissen. Papier-Prototypen – handgeschriebene/gezeichnete Konzepte mit niedriger Komplexität bzw. komplexere Entwürfe, die bereits in digitaler Form erstellt wurden. Interaktive Sketches – eine Zusammenstellung von handgeschriebenen Skizzen interaktiv verbunden. Abschließend dann die ersten funktionsfähigen Prototypen, die skelett-artig die wichtigsten Punkte enthalten, bis hin zum finalen Design. Zu jedem spezifischen Prototyp werden mit formativen Usability-Evaluationsmethoden, welche in Kapitel 2.3 näher beschrieben werden, Verbesserungsvorschläge erarbeitet.

Die anschließende *summative Usability-Evaluation* beinhaltet das Evaluieren des fertig entwickelten Produktes und wird ebenfalls näher in Kapitel 2.3 erläutert.

Den Abschluss des Usability Engineering Lifecycles bilden *Follow-Up-Studien*, denn, nach dem Release ist vor dem Release. Durch unterschiedliche Möglichkeiten können nach dem Release eines Produktes Usability-relevante Daten gesammelt werden. Beispielsweise durch spezielle Feld-Studien, wie Interviews, Fragebögen und Beobachtungen, standardisierte Marketing-Studien oder durch die Analyse der eingegangenen Beschwerden, Bug Reports und Veränderungswünsche der Benutzerinnen und Benutzer. (Andrews, 2015)

2.3. Methoden der Usability-Evaluation

Nachfolgend soll ein Überblick über Usability-Evaluationsmethoden gegeben werden, woraufhin Methoden der Usability-Inspektion (siehe Kapitel 2.3.1) und Usability-Test-Methoden (siehe Kapitel 2.3.1) beschrieben werden.

Abbildung 2.3 zeigt, dass Usability-Evaluationsmethoden auf zwei unterschiedliche Weisen eingeteilt werden können. Einerseits abhängig von wem die Evaluation durchgeführt wird, andererseits in Abhängigkeit des Zwecks der Testung, wie nachfolgend beschrieben (Andrews, 2015):

- *Explorativ* (engl. exploratory): Nachweisen wie und wozu ein Produkt verwendet wird
- *Prädiktiv* (engl. predictive): Abschätzung darüber, wie gut die User Performance basierend auf einem Interface-Design sein wird
- *Formativ* (engl. formative): qualitatives Design-Feedback zur laufenden Entwicklung in Form von Listen von Problemen und empfohlenen Lösungen
- *Summativ* (engl. summative): quantitative Gesamt-Bewertung eines Produktes oder eines Vergleichs mehrerer Produkte bzw. Lösungen

Sehr gut abgegrenzt werden können diese Begriffe untereinander durch Robert Stake's „Suppen-Analogie“, die von Andrews (2008) folgend erweitert wurde:

“When the cook tastes other cooks' soups, that's exploratory.

When the cook predicts the quality of a soup from a recipe, that's predictive.

When the cook tastes his own soup while making it, that's formative.
When the guests (or food critics) taste the soup, that's summative."

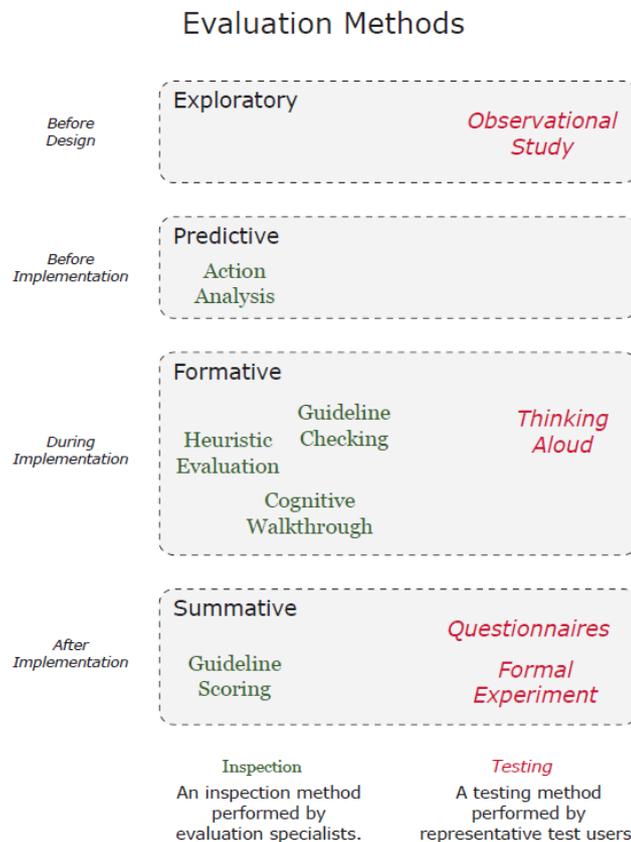


Abbildung 2.3.: Übersicht von Neun Usability-Evaluationsmethoden, welche nach der ausführenden Gruppe und dem Zweck unterteilt sind, entnommen aus Andrews (2008)

Usability-Evaluationsmethoden können, wie in Abbildung 2.3 dargestellt, auch in Inspektions- und Testmethoden unterteilt werden, welche nachfolgend beschrieben werden.

2.3.1. Usability-Inspektion

Usability-Inspektionsmethoden werden von Usability Professionals oder repräsentativen Anwenderinnen und Anwendern durchgeführt. Sie prüfen ein Produkt basie-

rend auf ihrer Erfahrung und Expertise, um es zu bewerten. Diese Überprüfung ist heuristisch, da sie auf Analyse und Bewertung beruht, und nicht auf einem Experiment (Andrews, 2015). Die fünf in Abbildung 2.3 gelisteten Inspektionsmethoden werden nachfolgend beschrieben.

Heuristische-Evaluation

Bei der heuristischen Evaluation prüft ein kleines Team von Professionals ein Interface unter Verwendung einer kurzen Checkliste von grundlegenden Prinzipien, den Heuristiken. Die Checkliste setzt sich aus den zehn heuristischen Prinzipien zusammen, der Sichtbarkeit des System-Status, einer Verbindung zwischen System und echter Welt, der Kontrolle bzw. Freiheit des Benutzers/der Benutzerin (z. B. ein schneller Exit nach Fehleingaben), einheitliche Standards bzw. Konventionen, Fehler-Vermeidung, Sichtbarkeit anstatt Erinnerung, Flexibilität und Effizienz (im Sinne von Beschleunigern), ästhetisches und minimalistisches Design, Hilfe bei der Analyse von Fehlern (gute Fehlermeldungen) und als letztes, Hilfe bzw. Dokumentation. Als Ergebnisse werden meist Listen von potentiellen Problemen erreicht.

Guideline-Checking

Beim Guideline Checking überprüfen die Usability Professionals ein Interface anhand einer detaillierten Guideline. Dies liefert eine Liste von Abweichungen von eben dieser Guideline. Im Gegensatz zur heuristischen Evaluation, bei der zehn grundlegende Prinzipien überprüft werden, können mittels Guideline Checking dutzende oder auch hunderte, oft sehr spezifische Objekte einer Checkliste, abgearbeitet werden. (Andrews, 2015).

Cognitive Walkthrough

Ein kleines Team von Professionals führt bei einem Cognitive Walkthrough typische Aufgaben mit der Denkart eines ungeübten Benutzers/einer ungeübten Benutzerin durch, und erzeugt eine Erfolgs- oder Misserfolgs-Geschichte an jedem Punkt

des korrekten Weges. Der Fokus liegt auf dem Testen der Erlernbarkeit eines Systems, da es auf dem kognitiven Model CE+ des erforschenden, menschlichen Lernens beruht. Das CE+ Modell besteht aus einer Problemlösungskomponente, einer Lernkomponente und einer Ausführungskomponente. (Andrews, 2015).

Guideline Scoring

Usability Professionals bewerten beim Guideline Scoring ein Interface bezüglich einer speziellen Guideline und liefern ein Gesamtergebnis, das repräsentiert in welchem Grad das Interface der Guideline folgt. Vorteile sind die billige und intuitive Umsetzbarkeit. (Andrews, 2015).

Action Analysis

Ein Usability Professional erstellt für die Action Analysis eine zeitliche Schätzung, wie lange ein erfahrener Benutzer/eine erfahrene Benutzerin für eine Aufgabe benötigt, indem die Aufgabe in immer kleinere Teilschritte zerlegt wird und dann die Mini-Einzelzeiten wieder summiert werden, wobei der Fokus, wie erwähnt, auf der Performance von erfahrenen Benutzerinnen und Benutzern liegt. Es kommen hier zwei Varianten zum Einsatz, die „Keystroke-Level-Analyse“, eine formale Methode auf hohem Niveau, bei der alle Aufgaben sehr genau zerlegt werden, sowie die informale Variante, „Back of the Envelope“, was zu Deutsch „Milchmädchenrechnung“ genannt wird, und die Aufgaben weit weniger detailliert vorgeschrieben werden. (Andrews, 2015).

2.3.2. Usability-Testungen

„Would you fly in an airplane that hasn't been flight tested? Of course not. So you shouldn't be using software that hasn't been usability tested.“

Ben Shneiderman, The Front Desk, BBC Video, 1995, vgl. Andrews (2015)

Anhand des angeführten Zitats wird verdeutlicht, dass Usability-Testing-Methoden einen hohen Stellenwert bei jeder Software-Neuentwicklung einnehmen müssen. Im

Speziellen bedeutet das, dass empirische Testmethoden entwickelt werden, die darauf beruhen, dass repräsentative Testpersonen das Produkt verwenden, wobei Messungen und Beobachtungen durchgeführt werden. Die wichtigsten, nachfolgend näher ausgeführten, Methoden sind: Thinking Aloud, Co-Discovery, Formales Experiment, A/B-Test, Fragebögen und Benutzungs-Studien.

Am Beginn steht jedoch die korrekte Vorbereitung für Usability Tests, wobei der Testumgebung besondere Beachtung geschenkt werden sollte sowie dem Test-Equipment. Die Testung sollte in einem ruhigen Raum stattfinden. Weiters sollte ein Schild an der Tür angebracht werden mit „User-Test im Gange – bitte nicht stören“. Alle mobilen wie stationären Telefone sind zu deaktivieren und es ist für adäquate Beleuchtung sowie (nicht-alkoholische) Erfrischungsgetränke zu sorgen. (Andrews, 2015).

Thinking Aloud

Testpersonen werden bei Thinking-Aloud-Tests aufgefordert laut zu denken, während sie eine Anzahl an typischen Aufgaben ausführen. Man spricht hier vom aufgabenbegleitenden Lauten Denken oder „concurrent thinking aloud – CTA“. Beim retrospektiven Lauten Denken (engl. retrospective thinking aloud – RTA) werden die Aufgaben zunächst ohne Verbalisierung durchgeführt und im Anschluss sollen anhand der aufgenommenen Videoaufzeichnungen die Gedanken zu den Aufgaben im Nachhinein formuliert werden. Die gewonnene Erkenntnis, warum Probleme auftauchten, wird dazu verwendet eine Liste mit Empfehlungen zu erstellen. (Andrews, 2015).

Formale Experimente

Formale Experimente sind kontrollierte Experimente, bei denen Messungen vorgenommen bzw. quantitative Daten gesammelt werden. Es werden sowohl objektive als auch subjektive Daten gesammelt, sowie summative Evaluationen durchgeführt, gefolgt von einer statistischen Analyse. Die beiden Hauptverwendungen sind die Bestimmung einer absoluten Performance eines Produktes und der objektive Vergleich von zwei oder mehreren alternativen Designs. (Andrews, 2015).

Co-Discovery

Bei der Co-Discovery entdecken zwei Testpersonen das Interface gemeinsam. Der interessante Aspekt dabei ist, die sich natürlich entwickelnde Interaktion bzw. Kommunikation zwischen den beiden Personen beobachten zu können. Das Verständnis über auftauchende Probleme wird durch diese beiden Aspekte gewonnen. Man benötigt aber auch doppelt so viele Testpersonen und man hat das grundlegende Problem der Validierung, da man davon ausgehen muss das das Produkt im echten Betrieb nicht von zwei Benutzerinnen und Benutzern gleichzeitig verwendet wird. (Andrews, 2015).

A/B-Test

Der A/B-Test ist ein kontrolliertes Experiment, meist in Distanz zu tatsächlichen Live-Usern, typischerweise bezüglich Webseiten, von denen es ein leicht verändertes Design B und die Standard-Seite A gibt, auf die zufällig verwiesen wird. Im Hintergrund werden laufend Messungen für das generelle Evaluationskriterium für beide Varianten durchgeführt und diese Unterschiede werden auf ihre statistischen Unterschiede untersucht. (Andrews, 2015).

Fragebögen

Die Einsatzmöglichkeiten von Fragebögen sind, samt Beispielen für frei erhältliche Fragebögen, in Kapitel 3.2 näher beschrieben.

Benutzungsstudien

Benutzungsstudien beruhen auf der Sammlung von Benutzungsdaten von einer kleinen Anzahl an Testpersonen, die über einen längeren Zeitraum in ihrer normalen Umgebung ihre spezifischen Aufgaben ausführen. Der Zweck ist herauszufinden, wie viel Zeit für jede Aktivität aufgewendet wird, sowie festzustellen, welche Software für welchen Zweck verwendet wird, als auch Erkenntnis zu gewinnen, welche Features einer Software am meisten verwendet werden, und welche am wenigsten. Die drei

wichtigen Varianten sind: Diary-Studies, bei denen die Benutzerinnen und Benutzer Tagebuch-artig die Art der Verwendung des Systems festhalten, sowie Software Logging, wo eine Software alle Interaktionen zum Zweck der späteren statistischen Analyse aufzeichnet, und Beobachtungsstudien, bei denen ein oder mehrere typische Tage der Verwendung des Systems aufgezeichnet und als Timeline später manuell ausgewertet werden, um die resultierenden Daten statistisch analysieren zu können. (Andrews, 2015).

3. Messen von Usability

Wie bereits in Kapitel 2.1.1 zusammengefasst, gibt es mehrere Attribute die Usability näher charakterisieren, welche auf unterschiedliche Weise gemessen werden können:

Effektivität: Kriterien des erfolgreichen Erreichens des Ziels werden definiert. Man kann also qualitativ bestimmen, inwieweit das Ergebnis mit den definierten Zielen übereinstimmt, und quantitativ messen wie oft das Ergebnis die definierten Ziele erfüllt (Andrews, 2015).

Laut Nielsen (1993) können die von ihm definierten Merkmale u. a. folgendermaßen gemessen werden:

Erlernbarkeit: Man wählt unerfahrene Benutzerinnen und Benutzer des Systems aus und misst die Zeit die sie zum Erfüllen festgelegter Ziele benötigen. Es kann beispielsweise zwischen Benutzerinnen und Benutzern mit keiner Erfahrung im Umgang mit Computern/einem bestimmten System und solchen mit allgemeiner Erfahrung unterschieden werden. Betrachtet man den Prozess des Erlernens genauer, kann festgestellt werden, dass Benutzerinnen und Benutzer ein UI nicht vollständig erlernen bevor sie anfangen es zu verwenden und vice versa es beginnen zu verwenden, sobald sie ein Teil davon erlernt haben. Da Benutzerinnen und Benutzer oft ungeübt beginnen ein System zu nutzen, sollte nicht nur gemessen werden wie lange es dauert eine hervorragende Beherrschung des Systems zu erlangen, sondern auch wie lange es dauert annehmbare Ergebnisse bei relevanter Arbeit zu erreichen.

Effizienz: Man definiert das Level an Performance eines erfahrenen Benutzers/einer erfahrenen Benutzerin indem man z. B. eine gewisse Zeit des Umgangs mit dem System definiert. Diese Definition wird oft beim Test neuer Systeme verwendet, da es noch keinen ausreichenden Pool an erfahrenen Benutzerinnen und Benutzern gibt.

Alternativ definiert man ein Level an Expertise und sucht sich dementsprechende, repräsentative Probanden, was nicht immer leicht ist, und misst die Zeit die diese zum Erfüllen festgelegter Ziele benötigen.

Einprägsamkeit: Die dritte Gruppe neben unerfahrenen und erfahrenen Benutzerinnen und Benutzern ist die der gelegentlichen Benutzerinnen und Benutzer, die sich dadurch definieren, dass sie einigen Umgang mit dem System hatten, es aber für einige Zeit nicht verwendet haben. Wichtig für deren Wiedereinstieg ist hierbei ein leicht zu merkendes Interface. Man sucht also repräsentative Benutzerinnen und Benutzer dieser Gruppe und misst die Zeit, die diese zum Erfüllen festgelegter Ziele benötigen. Man kann alternativ auch einen Erinnerungstest durchführen, indem man die Benutzerinnen und Benutzer bittet, diverse Kommandos zu erklären oder deren Namen wiederzugeben. Erstere Methode ist die repräsentativere, da moderne Interfaces darauf basieren, dem Benutzerinnen und Benutzer so viel wie möglich sichtbar zu machen. Die Benutzerinnen und Benutzer müssen sich also nicht spezifisch an alle Funktionen erinnern, da sie bei Bedarf vom System aufgezeigt werden. Eine Studie hat gezeigt, Benutzerinnen und Benutzer können ohne sich an die Namen oder Icons des Interfaces zu erinnern problemlos damit umgehen, wenn sie das Interface wieder verwenden. (Nielsen, 1993)

Fehler: Ein Fehler ist definiert als eine Aktion, die nicht zum Ziel führt und kann in leichte und schwerwiegende Fehler unterteilt werden. Man misst also die Anzahl der Abweichungen vom optimalen Weg. Sofort korrigierte Fehler werden hier nicht eingerechnet, da sie bei der Effizienz eingerechnet werden, schließlich verlangsamen sie nur ein wenig den Prozess.

Zufriedenheit: Die subjektive Zufriedenheit wird gemessen, indem man die Benutzerinnen und Benutzer befragt, nachdem sie das System zur Erfüllung festgelegter Ziele verwendet haben. Einzelantworten bleiben subjektiv, aber zusammengefasst ergeben sie ein durchschnittliches Level an Zufriedenheit mit dem System. In der Mehrzahl der Studien werden Benutzerinnen und Benutzer direkt befragt, typischerweise werden kurze Fragebögen verwendet, in denen die Benutzerinnen und Benutzer aufgefordert werden das System auf Bewertungsskalen von 1-5 oder 1-7 zu beurteilen. Die Skalen sind meist Likert-Skalen oder es werden semantische Differentiale verwendet (für Beispiele von Skalen siehe Kapitel 3.2).

Diese Attribute können durch verschiedene Maße gemessen werden. Konkret teilen Tullis et al. (2013) das Messen von Usability in mehrere Arten von Metriken ein, Performance-Metriken (Kapitel 3.1) und selbst-rückgemeldete Metriken (Kapitel 3.2), die hier genauer beschrieben werden, aber auch Problem-basierende Metriken (Kapitel 3.3), verhaltensbezogene und physiologische Metriken (Kapitel 3.4), sowie kombinierte und vergleichende Metriken (Kapitel 3.5), die in ihren Grundzügen beschrieben werden.

Ob die UX von einem Produkt zum Folgenden tatsächlich verbessert wurde, zeigen Usability-Metriken durch das Messen und Vergleichen der beiden Produkte untereinander, sowie durch die Auswertung der möglichen Verbesserung auf, was zu einer Win-win-Situation führt. Usability-Metriken liefern Informationen für die Berechnung des ROI. Ohne Metriken ist es nicht möglich zu bestimmen, wie viel Geld durch ein neues Produktdesign gespart werden kann. Eine leichte Veränderung eines Eingabefeldes einer internen Webseite könnte eine Vielzahl von Verbesserungen nach sich ziehen, was zu erhöhter Kunden-Zufriedenheit, einem Anstieg der Bestellungen und schließlich zu einer generellen Steigerung des Umsatzes des Unternehmens führen kann. Beispiele für eine Verkettung von Veränderungen, die die oben beschriebenen positiven Effekte hätten, wären: verringerte Dateneintrags-Fehler um 75 %, verkürzter Zeitaufwand zum Abschließen eines Falles im Kundendienst, Erhöhen der abgearbeiteten Aufträge pro Tag, Senken des Rückstaus der Bestellungen und Reduzieren der Verspätungen bei den Kundenlieferungen. (Tullis et al., 2013)

3.1. Performance-Metriken

Performance-Metriken gehören für alle Usability Professionals zu den nützlichsten Instrumenten, denn sie untermauern wesentliche Entscheidungen, wie den Launch eines neuen Produktes. Sie sind die beste Methode um, basierend auf User-Verhalten und nicht aufgrund subjektiver Aussagen, die Effektivität und Effizienz verschiedenster Produkte zu beurteilen (Tullis et al., 2013). Es können fünf wesentliche Arten von Performance-Metriken unterschieden werden, die nachfolgend beschrieben werden.

3.1.1. Task Success

Task Success ist ein weit verbreitetes Usability-Maß, da für beinahe jedes Produkt für das Aufgaben definiert werden können, auch die Erfolgsrate der Benutzerinnen und Benutzer angegeben werden kann. Der Benutzer/die Benutzerin gibt an, wie effektiv eine Reihe von Aufgaben mit klarem Ziel bzw. Endzustand abgeschlossen werden kann.

Die einfachste Möglichkeit den Task Success zu ermitteln, ist mittels *binärer Bewertung*: 1 steht für Erfolg, 0 für Misserfolg. Daraus können sehr einfach die durchschnittliche Erfolgsrate pro Task und die durchschnittliche Erfolgsrate pro Benutzerin/pro Benutzer ermittelt werden, wobei typischerweise die Erfolgsraten pro Task als Ergebnis mit Hilfe von Balkendiagrammen und Konfidenzintervallen präsentiert werden. Es ist auch möglich die Ergebnisse nach Nutzergruppen (z. B. nach Gebrauchshäufigkeit oder Vorerfahrung mit dem Produkt) zusammenzufassen oder eine Häufigkeitsverteilung der Erfolgsraten anzufertigen, zweiteres jedoch nur bei einer relativ großen Anzahl von Personen in der Studie (mindestens 12, besser mehr als 20 Personen).

Eine andere Möglichkeit ist, *Erfolgsstufen* basierend auf dem Erfüllungsgrad einer Aufgabe einzuführen, diese Erfüllungsgrade haben meist drei bis sechs Level, z. B. „1 = Erfolg“, „0,5 = teilweiser Erfolg“, „0 = Misserfolg“, und können auch aufgrund der UX definiert werden, z. B. aufgrund des beobachteten Auftretens von Problemen, die den Erfolg beeinflussen können, durch die Bewertung „1 = kein Problem“, „2 = geringes Problem“, „3 = bedeutendes Problem“, „4 = Misserfolg/Aufgabe“. Diese Level müssen gegenseitig sehr genau abgegrenzt werden, damit eine eindeutige Zuordnung möglich ist. Präsentiert werden können Erfolgsstufen meist als Häufigkeitsverteilung pro Task durch gestapelte 100 %-Säulen. (Tullis et al., 2013)

3.1.2. Time on Task

Die Time on Task zu messen ist eine gute Art und Weise um die *Effizienz* eines Produktes zu bestimmen, sofern man daran interessiert ist wie schnell Benutzerinnen und Benutzer eine Aufgabe abschließen können. Dies könnte nicht der Fall sein,

wenn man Spiele testet, wo eigentlich bezweckt wird, dass Benutzerinnen und Benutzer nicht allzu schnell zu spielen aufhören. Die Zeit die benötigt wird, um eine Aufgabe abzuschließen kann für alle Benutzerinnen und Benutzer, eine Teilmenge dieser Personen oder einen Anteil von Benutzerinnen und Benutzer, die die Aufgaben in einer vorgegebenen Zeit erfüllen konnten, betrachtet werden. Präsentiert werden die durchschnittlichen Zeiten (Mittelwert, Median oder geometrisches Mittel) durch Balkendiagramme mit entsprechenden Konfidenzintervallen. Eine Gruppierung der Zeiten und Darstellung mittels Häufigkeitsverteilung ist auch denkbar. (Tullis et al., 2013)

3.1.3. Fehler

Fehler (engl.: Errors) die von Benutzerinnen und Benutzern gemacht werden, sind nicht mit Usability-Problemen (engl.: usability issues) gleichzusetzen, denn ein Usability-Problem ist die zugrunde liegende Ursache für ein Problem und eine gewisse Anzahl von Fehlern ist ein daraus resultierendes Ergebnis, also sind Fehler falsche Handlungen, die zum Misserfolg bezüglich einer Aufgabe führen können. Nachdem bestimmt wurde, wie viele Möglichkeiten es in einem Task gibt Fehler zu begehen, können die Fehler bestimmt und auf vielfältige Weise präsentiert werden, beispielsweise absolut als Fehlerrate oder Tasks, deren Fehlerrate über einer bestimmten Schwelle liegen, je nachdem, was Grundlage der Erfassung war und was durch das Aufzeichnen der Fehler ausgedrückt werden will. (Tullis et al., 2013)

3.1.4. Effizienz

Neben der Zeit die für das Abschließen einer Aufgabe benötigt wird, kann Effizienz auch durch den Umfang des (physischen und kognitiven) Aufwands ausgedrückt werden. Typischerweise wird dieser Aufwand durch die Anzahl der Aktionen oder Schritte, die für den Aufgaben-Erfolg benötigt werden, ausgedrückt. Physischer Aufwand ist beispielsweise jedes Klicken, Ausfüllen oder Betätigen eines Schalters, kognitiver Aufwand beinhaltet das Suchen von Informationen, Entscheiden für Aktionen oder Interpretieren von Handlungsergebnissen. Präsentiert werden die durchschnittliche Anzahl der Aktionen mit Konfidenzintervallen pro Task oder pro Benutzerin/pro

Benutzer. Ein Vergleich der Tasks untereinander ist jedoch nur dann sinnvoll, wenn jeder Task ungefähr gleich viele Schritte benötigt.

Das Protokollieren der Schritte kann manuell schnell sehr aufwändig werden oder aufgrund der Geschwindigkeit der Aktionen nicht möglich sein, darum ist es auch möglich, Effizienz als Kombination von Aufgaben-Erfolg und -Dauer zu betrachten, was meist durch das Verhältnis der Erfolgsrate pro Task zur durchschnittlichen Time on Task, also dem Task Success pro Zeiteinheit, ausgedrückt wird. Weiters kann die Effizienz von Seiten der Benutzerinnen und Benutzer durch das Verhältnis der Anzahl der erfolgreich abgeschlossenen Aufgaben zur Gesamtzeit aller Aufgaben bestimmt werden. (Tullis et al., 2013)

3.1.5. Erlernbarkeit

Erlernbarkeit bezieht sich auf das effiziente Erlernen von etwas Bestimmten und dabei wird betrachtet, wie sich beliebige Effizienz-Metriken über die Zeit verändern. Das heißt, dass Aufzeichnungen mehrerer Versuche von der selben Session, der selben Session mit Pausen zwischen den Aufgaben oder mehreren Sessions ausgewertet werden. Die Zeit zwischen den Versuchen oder Sessions ist abhängig von der zu erwarteten Häufigkeit der Verwendung des getesteten Systems. Dargestellt wird ein Effizienz-Maß als Funktion der Erfahrung über die Versuche. Es können Werte von Aufgaben für sich oder von allen Aufgaben summiert betrachtet werden. (Tullis et al., 2013)

3.2. Selbst-rückgemeldete Metriken

Selbst-rückgemeldete Metriken (Englisch: self-reported metrics), wie sie Tullis et al. (2013) nennen, werden oft auch als subjektive Daten oder Präferenz-Daten bezeichnet, was aber einen Mangel an Objektivität der Daten implizieren oder die Wahl einer Möglichkeit voraussetzen könnte. Diese Metriken sind dem Usability-Attribut Zufriedenheit zuzuschreiben, denn Benutzerinnen und Benutzer werden gefragt ihre Erfahrung mit einem System bewertend wiederzugeben. Effizient können diese Daten durch bestimmte Bewertungsskalen aufgezeichnet bzw. ausgewertet werden.

Items von Likert-Skalen sind positiv oder negativ formulierte Aussagen, die üblicherweise auf einer 5-stufigen Skala der Zustimmung von 1 bis 5 bewertet werden, was verbalisiert „stimme gar nicht zu“, „stimme nicht zu“, „stimme weder zu noch nicht zu“, „stimme zu“ und „stimme voll zu“ gleich kommt. Es können auch ebenso 7-stufige Skalen verwendet werden, bei denen dann aber meist nur die beiden Enden verbal als Ankerpunkte beschrieben werden. Durch die ungerade Anzahl der Bewertungsmöglichkeiten wird es den Testpersonen ermöglicht neutrale Antworten zu geben. Bei semantischen Differentialen werden bipolare Adjektivpaare mit Hilfe einer 5- oder 7-stufigen Skala bewertet, wie am Beispiel von Items des User Experience Questionnaire in Tabelle 3.2 dargestellt. Am besten werden diese Formate bei Befragungen nach jeder Aufgabe (Post-Task-Bewertung) und/oder nach Beendigung der gesamten Session (Post-Test-Bewertung) eingesetzt. Für beide Anwendungen gibt es zahlreiche standardisierte und ausreichend validierte Fragebögen-Formate, was es bei den meisten Testobjekten und Studien nicht notwendig machen wird, eigene Fragebögen zu entwickeln, denen leicht ein Bias anheften kann (Tullis et al., 2013). Zwei unterschiedliche Post-Test Fragebögen werden nachfolgend näher beschrieben.

3.2.1. System-Usability-Scale-Fragebogen

Der System-Usability-Scale-Fragebogen (SUS) wurde bereits 1986 von John Brooke bei Digital Equipment Corporation entwickelt, besteht aus zehn abwechselnd positiv und negativ formulierten Aussagen (siehe Tabelle 3.1), die jeweils auf einer 5-stufigen Likert-Skala von „1-strongly disagree“ bis „5-strongly agree“ zu bewertet sind, und ist ein gängiger Post-Test-Fragebogen zur Bewertung der Usability eines Systems durch die Testpersonen (Barnum, 2010; Tullis et al., 2013). Mit der bereitgestellten Auswertungsvorgabe, die besagt, dass positive Aussagen als Wert ihre Bewertung minus 1 und negative Aussagen als Wert 5 minus ihre Bewertung haben, ergibt sich für den SUS durch die Multiplikation der Summe aller Werte mit 2,5 ein eindimensionaler Gesamtwert für die Usability des getesteten Systems von 0 bis 100, wobei ein höheres Ergebnis einer besseren Usability zuzuschreiben ist. Die Auswertung einzelner SUS-Aussagen hat laut Brooke (1996) keine Bedeutung.

Tatsächlich haben Lewis et al. (2009) zwei unterschiedliche Dimensionen zur Bewertung ausgemacht, nämlich die (neue) *Usability* (SUS Aussagen 1, 2, 3, 5, 6, 7, 8

Originale SUS-Aussagen	Angepasste deutsche SUS-Aussagen
I think that I would like to use this system frequently	Ich kann mir sehr gut vorstellen, die App regelmäßig zu nutzen
I found the system unnecessarily complex	Ich empfinde die App als unnötig komplex
I thought the system was easy to use	Ich empfinde die App als einfach zu nutzen
I think that I would need the support of a technical person to be able to use this system	Ich denke, dass ich technischen Support brauchen würde, um die App zu nutzen
I found the various functions in this system were well integrated	Ich finde, dass die verschiedenen Funktionen der App gut integriert sind
I thought there was too much inconsistency in this system	Ich finde, dass es in der App zu viele Inkonsistenzen gibt
I would imagine that most people would learn to use this system very quickly	Ich kann mir vorstellen, dass die meisten Leute die App schnell zu beherrschen lernen
I found the system very cumbersome to use	Ich empfinde die Bedienung als sehr umständlich
I felt very confident using the system	Ich habe mich bei der Nutzung der App sehr sicher gefühlt
I needed to learn a lot of things before I could get going with this system	Ich musste eine Menge Dinge lernen, bevor ich mit der App arbeiten konnte

Tabelle 3.1.: Gegenüberstellung der zehn originalen SUS Aussagen (Brooke, 1996) und der verwendeten deutschen Aussagen (Rauer, 2011) mit der Modifikation, anstatt *System* immer die Bezeichnung *App* zu verwenden.

und 9) und die *Erlernbarkeit* (SUS Aussagen 4 und 10). Multipliziert man die Ergebnisse der neuen Usability mit 3,125 und die der Lernfähigkeit mit 12,5 erhält man wieder jeweils einen Wert zwischen 0 und 100, kann die Ergebnisse mit dem Gesamt-SUS-Ergebnis vergleichen und hat dadurch eventuell eine feinere Abstufung des Ergebnisses.

Zur besseren Interpretation der SUS-Ergebnisse und der leichteren Beantwortung der Frage, was denn nun ein „guter“ SUS-Wert sei, stellen Bangor et al. (2008) einen Vergleich von verschiedenen Ratings (Abbildung 3.1) bereit. Die einzelne qualitative Bewertung der Aussage „Overall, I would rate the user-friendliness of this product as:“ von „1“ Worst Imaginable“ bis „7“ Best Imaginable“, steht in hoher Korrelation ($r = 0,806$) zum SUS-Ergebnis, und konnte somit gut im zuvor genannten Vergleich (Abbildung 3.1) integriert werden. Der SUS hat den Vorteil, dass er sich bereits ab einer relativ kleinen Anzahl von Testpersonen (z. B. 8-10) als robust erwiesen hat (Tullis et al., 2013).

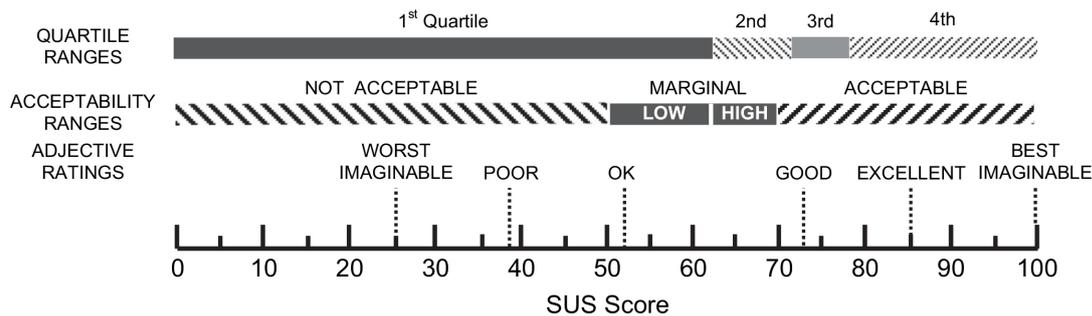


Abbildung 3.1.: SUS-Score Interpretation nach Quartilen, Akzeptanz und qualitativer Bewertung, entnommen aus Bangor et al. (2008).

3.2.2. User Experience Questionnaire

Zur einfachen und kostengünstigen Messung der User Experience sind im deutschsprachigen Raum die Fragebögen UEQ – User Experience Questionnaire (Laugwitz, Schrepp et al., 2006) und AttrakDiff (Hassenzahl et al., 2003) weit verbreitet (Schrepp et al., 2013). Der UEQ ist ebenfalls ein Post-Test-Fragebogen, der sich aus 26 bipolaren Items, die die Form eines 7-stufigen semantischen Differentials haben,

3. Messen von Usability

zusammensetzt, ein solches semantisches Differential ist in Tabelle 3.2 veranschaulicht.

	1	2	3	4	5	6	7		
unverständlich	<input type="radio"/>	verständlich	2						
übersichtlich	<input type="radio"/>	verwirrend	21						
kompliziert	<input type="radio"/>	einfach	13						
attraktiv	<input type="radio"/>	unattraktiv	24						

Tabelle 3.2.: Beispiel eines 7-stufigen semantischen Differentials mit bipolaren Items, wie im UEQ (Laugwitz, Schrepp et al., 2006)

Die einzelnen Items sind einer von sechs Dimensionen zugeordnet: *Durchschaubarkeit*, *Effizienz*, *Originalität*, *Steuerbarkeit*, *Stimulation* (jeweils 4 Items) sowie *Attraktivität* (6 Items). Darunter sind mit Durchschaubarkeit, Effizienz und Steuerbarkeit „harte“ Usability-Kriterien, die aufgabenbezogene (pragmatische) Qualitätsmerkmale eines Produktes beschreiben, mit Originalität und Stimulation „weiche“ Usability-Kriterien, die nicht aufgabenbezogene (hedonische) Qualitätsmerkmale eines Produktes beschreiben, und mit Attraktivität eine globale Bewertung auf einer Zustimmungs- bzw. Ablehnungsdimension, wie in Abbildung 3.2 dargestellt. Hassenzahl (2001) (Schrepp et al., 2013; Laugwitz, Schrepp et al., 2006). Mit diesen sechs Dimensionen soll der UEQ den Gesamteindruck eines Nutzers/einer Nutzerin bezüglich eines Systems messen, die Definition von UX erfolgte aber beispielsweise nicht gleich wie in der Normenreihe 9241 (vgl. Kapitel 2.1.2), das heißt, der UEQ kann keine normgerechte Messung der UX liefern, jedoch muss dies in der Praxis nicht immer notwendig sein (Sarodnick et al., 2011).

Der UEQ steht in mehreren Sprachversionen kostenlos online¹, samt einem Excel Auswertungstool, zur Verfügung. In der Auswertung haben die einzelnen Items einen Wert von -3 bis $+3$, wobei jeweils die negativste Bewertung eines Adjektivpaares -3 bzw. die positivste Bewertung $+3$ als Wert hat. Die Bildung der Mittelwerte aller Items in den einzelnen Dimensionen liefert das Gesamtergebnis für den UEQ, das wiederum Werte zwischen -3 und $+3$ annehmen kann, wobei ein Wert $> 0,8$ einer

¹www.ueq-online.org (zuletzt besucht am 21.8.2015)

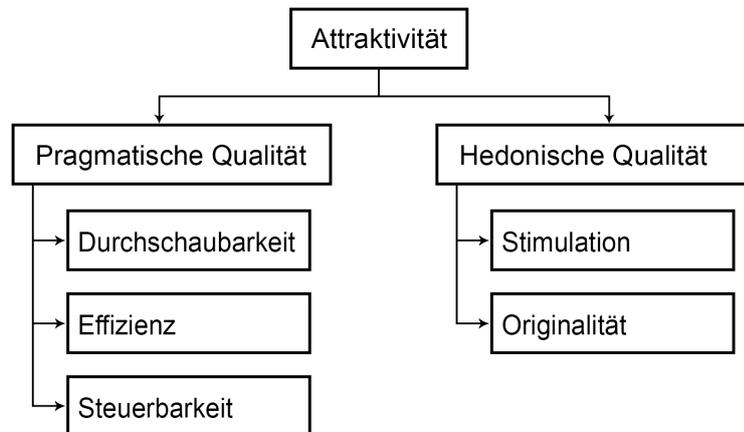


Abbildung 3.2.: Skalenstruktur des User Experience Questionnaire (UEQ), nach Schrepp et al. (2013).

positiven Bewertung der jeweiligen Dimension durch die Benutzerinnen und Benutzer entspricht, ein Wert $< -0,8$ für eine negative und ein Wert dazwischen für eine neutrale Beurteilung steht, in der Praxis jedoch Werte $> 2,0$ bzw. $< -2,0$ aufgrund generell bestehender Antworttendenzen, z. B. der Vermeidung extremer Antwortkategorien durch viele Personen, nur selten auftreten.² Die gesamte Konstruktion und Validierung des UEQ sind in Laugwitz, Schrepp et al. (2006) und Laugwitz, Held et al. (2008) beschrieben.

Mittlerweile wurde zur besseren Interpretation der erhaltenen Ergebnisse, auch ein Benchmark für den UEQ veröffentlicht (Schrepp et al., 2013) und die entsprechenden Intervalle im Auswertungstool zum UEQ integriert. Der UEQ Benchmark fasst hierbei Daten von 4818 Personen aus 163 Studien zusammen und erlaubt Rückschlüsse zur relativen Qualität des getesteten Produkts im Vergleich zu anderen Produkten. Jede Dimension hat ihre eigenen Grenzen für exzellente, gute, überdurchschnittliche, unterdurchschnittliche und schlechte Bewertungen. Ein exzellentes Ergebnis liegt unter den besten 10% der bisher beobachteten Ergebnisse, ein schlechtes Ergebnis unter den schlechtesten 25% der bisher beobachteten Ergebnisse, für eine gute Bewertung einer Dimension gilt, 10% der Ergebnisse sind besser und 75% der

²Informationen entnommen aus dem UEQ-Excel-Auswertungstool, welches unter www.ueq-online.org verfügbar ist (zuletzt besucht am 21.8.2015)

Ergebnisse schlechter, ein überdurchschnittliches Ergebnis charakterisiert sich dadurch, dass 25 % der Ergebnisse besser und 50 % der Ergebnisse schlechter sind und ein unterdurchschnittliches Ergebnis dadurch, dass 50 % der Ergebnisse besser und 25 % der Ergebnisse schlechter sind, als das beobachtete.

3.3. Problem-basierende Metriken

Usability-Probleme (Englisch: usability issues) aufzudecken und Verbesserungsvorschläge zu machen ist für viele Personen, die im Feld der Usability tätig sind, das wichtigste. Usability-Probleme werden qualitativ beschrieben, aber auch diese, einmal identifiziert, können gemessen werden. Sie werden nach Schweregrad kategorisiert, der immer die Auswirkung auf die UX berücksichtigen sollte. Meist werden Usability-Probleme in Form der Häufigkeit einzelner Probleme, dem Anteil der Benutzerinnen und Benutzer, bei denen ein bestimmtes Problem aufgetreten ist, und der Häufigkeit von Problemen bei unterschiedlichen Tasks oder in unterschiedlichen Kategorien von Problemen gemessen. Usability-Probleme zu identifizieren wird von einer Reihe von Faktoren beeinflusst, sodass die Gefahr besteht, dass die Ergebnisse mit einem Bias behaftet sind. (Tullis et al., 2013)

3.4. Verhaltensbezogene und physiologische Metriken

Es gibt zahlreiche verhaltensbezogene und physiologische Maße, die bei Usability-Tests hilfreich sein können, außerdem erhält man einen anderen Blickwinkel auf die Ergebnisse, denn man vertraut nicht ausschließlich auf das, was von Testpersonen gesagt bzw. wie gemacht wird. Ohne zusätzliches Equipment kann offensichtliches verbales und nonverbales Verhalten in Form von Protokollen oder Notizen von erfahrenen Beobachtern dokumentiert werden, dies sollte jedoch für eine einfache anschließende Auswertung sehr strukturiert erfolgen. Weitere Maße, wie Gesichtsausdrücke oder Hautleitwert und Herzrate, können nur durch entsprechendes Equipment gewonnen werden. Gesichtsausdrücke während des Usability-Tests zu bestimmen hilft über das Gesagte hinaus einen Einblick in das Gefühlte und Gedachte zu bekommen. Der Hautleitwert und die Herzrate lassen Rückschlüsse auf Stress

zu und können eingesetzt werden, um Teile des UI auszumachen, die Personen als besonders frustrierend empfinden. Eye-Tracking-Daten können in Usability-Tests außergewöhnlich nützlich sein, alleine, dass die Beobachter in Echtzeit oder bei der Nachbereitung sehen können wohin Benutzerinnen und Benutzer blicken, ist äußerst hilfreich, da bestimmt werden kann, ob ein Nutzer/eine Nutzerin ein bestimmtes Element des UI überhaupt angesehen hat. Mit einem Eye Tracker können auch Daten über die Pupillen-Durchmesser gewonnen werden. Pupillen neigen dazu, sich bei geistiger Anstrengung und allgemeiner Erregung, zu erweitern. (Tullis et al., 2013)

3.5. Kombinierte und vergleichende Metriken

Man kann aus bestehenden Daten leicht neue Usability-Metriken ableiten, durch Kombination von mehr als einer Metrik zu einem einzelnen Usability-Maß oder durch Vergleich von den gewonnen Usability-Daten zu Experten- oder idealen Ergebnissen. Eine praktische Möglichkeit verschiedene Usability-Metriken zu kombinieren ist, das Bestimmen des Anteils der Testpersonen, die eine Kombination von Zielsetzungen erreicht haben. So erhält man den Gesamtanteil der Testpersonen, die eine gute Erfahrung mit dem getesteten Produkt hatten. Jede Metrik in Anteile umzuwandeln und dann zusammen zu mitteln ist eine andere Möglichkeit, die man anwenden kann, wenn für jede Metrik ein dazugehöriges Minimum und Maximum angegeben werden kann. Weitere Möglichkeiten zur Kombination basieren auf dem Einsatz von Z-Scores oder der SUM – Single Usability Metric, die Task-basierend zusammenfasst und auch die Fehler und Post-Task-Bewertungen beinhaltet. Grafisch können viele Ergebnisse in Usability Scorecards zusammengefasst werden. Linien- und Balkendiagramme eignen sich für die Darstellung von Ergebnissen von zwei Metriken auf einer Task-Ebene, Netzdiagramme können eingesetzt werden, um Ergebnisse von drei oder mehr Metriken gesammelt darzustellen, und mit Harvey Balls, Symbolen für verschiedene Level von Metriken, in einer Vergleichstabelle können effektiv die Ergebnisse von drei oder mehr Metriken auf Task-Ebene zusammengefasst werden.

Der wahrscheinlich beste Weg um den Erfolg eines Usability-Tests zu bestimmen, ist das Vergleichen der Ergebnisse mit einer Reihe von vordefinierten Usability-Zielen, die sich meist auf den Task Success, die -Dauer oder die User-Zufriedenheit beziehen,

3. Messen von Usability

um so die Anteile der Testpersonen, die die Ziele erfüllt haben, als Zusammenfassung zu präsentieren. Gerade für Task-Dauer-Daten ist es üblich das Ergebnis des Usability-Tests mit den Ergebnissen von Expertinnen und Experten zu vergleichen. Die Daten der Expertinnen und Experten können Aufschluss geben, ob ein Task nicht sowieso verhältnismäßig lange dauert. (Tullis et al., 2013)

4. Eye Tracking

Beim Eye Tracking, zu Deutsch auch manchmal Blickregistrierung, -erfassung oder Augentracking genannt, werden quantitativ Blickbewegungen erfasst, die Aufschluss über Reihenfolgen der Wahrnehmung oder Auffälligkeiten im Benutzerverhalten geben können. In diesem Zusammenhang werden zunächst einige Grundlagen erläutert, wo etwa Augenbewegungen im Zusammenhang der Messung mit Eye-Tracking-Systemen beschrieben werden. Im darauf folgenden Kapitel wird die Relevanz von Eye Tracking in klassischen Usability-Testungen dargelegt und das abschließende Kapitel zeigt wesentliche Unterschiede zum Testen von mobilen Anwendungen auf.

4.1. Grundlagen des Eye Tracking

Zu Beginn wird erklärt welche Augenbewegungen für das Eye Tracking von Bedeutung sind und welche Technologien es gibt. Abschließend werden mögliche Probleme beim Einsatz von Eye Tracking beschrieben.

4.1.1. Augenbewegungen

Der Mensch hat zwar ein großes Blickfeld, der Bereich in dem scharf gesehen wird, der Bereich des „gelben Fleckes“ mit der Sehgrube (der Fovea) auf der Netzhaut (Retina), ist jedoch sehr klein, sodass Augenbewegungen notwendig sind, um nacheinander kleine Teile eines größeren Bereichs scharf zu sehen. Diese Augenbewegungen werden durch drei Muskelpaare gesteuert, die horizontale, vertikale und rotierende Bewegungen des Augapfels ermöglichen (Holmqvist et al., 2011). Diese so entstehenden Bewegungsmuster, die verschiedenen Zwecken dienen, werden nachfolgend bezugnehmend auf die Messung mit Eye-Tracking-Systemen erläutert.

Die am meisten beachteten Ereignisse beim Eye Tracking sind kurze Momente, in denen sich das Auge still verhält, bzw. relative Ruhephasen aufweist, sogenannte *Fixationen*, die Zehntel von Millisekunden bis hin zu einigen Sekunden dauern können. Eine Fixation weist typischerweise eine Dauer zwischen 250 und 300 ms auf (Goldberg et al., 1999). Bei einer Fixation steht das Auge jedoch nicht komplett still, sondern es wird durch die Mikro-Bewegungen Tremor, Drift und Minisakkaden stabilisiert. Mit dem Messen von Fixationen wird meist assoziiert, dass Aufmerksamkeit gemessen werden kann. (Holmqvist et al., 2011)

Die Bewegung zwischen zwei Fixationen wird *Sakkade* genannt, diese ist die schnellst mögliche Bewegung des menschlichen Körpers, die gesamt typischerweise 30 bis 80 ms andauert. Eine Sakkade endet meist nicht direkt am beabsichtigten Ziel, sondern ist gefolgt von einer anderen sehr kurzen Bewegung, die von Holmqvist et al. (2011) Glissade genannt wird. Sakkaden sind sprunghaft und können sowohl bewusst, als auch reflexartig ausgeführt werden (Duchowski, 2007).

Das Auge kann auch gleitende, langsame Bewegungen machen, um etwas Bewegtem zu folgen, dies sind sogenannte *Folgebewegungen* (engl. smooth pursuits), die von einem anderen Teil des Gehirns durchgeführt werden als Sakkaden. (Holmqvist et al., 2011)

Die am häufigsten verwendete Video-basierende Messung von Augenbewegungen basiert auf der Erfassung der Reflexionen der Hornhaut (Cornea), erzeugt durch infrarote Lichtquellen, und der Pupillenbestimmung (Holmqvist et al., 2011). Nachfolgend werden die Verfahren und Geräte des Eye Trackings kompakt beschrieben.

4.1.2. Eye-Tracking-Technologien

Bereits zum Ende des 19. Jahrhunderts wurden Eye-Tracking-Geräte gebaut, die jedoch meist noch sehr invasive Methoden aufwiesen, die für die Testpersonen unangenehm und störend waren. Das erste Verfahren, dass das Messen von Reflexionen von externen Lichtquellen durch die Fovea (durch Fotografie) etablierte, wurde Anfang des 20. Jahrhunderts veröffentlicht, und diese viel weniger invasive Methode hat sich bald durchgesetzt und ist über die letzten Jahre in einer moderneren Form zur dominierenden Technik herangewachsen. Seit damals haben sich zahlreiche weitere

Verfahren entwickelt. Die Elektro-OculoGraphie (EOG) war in den 1970er Jahren die weit verbreitetste Möglichkeit zur Erfassung von Augenbewegungen, bei der mit Elektroden Spannungsdifferenzen zwischen den Bezugspunkten gemessen werden. Eine weitere Art Augenbewegungen zu messen ist mit Hilfe von skleralen Kontaktlinsen, die meist eine Drahtspule integriert haben, direkt ins Auge eingesetzt und in einem elektromagnetischen Feld gemessen werden. (Holmqvist et al., 2011; Duchowski, 2007)

Photo-OculoGraphie (POG) oder Video-OculoGraphie (VOG) beinhalten alle auf Bild und/oder Video basierenden Methoden zur Augenbewegungs-Aufzeichnung, die verschiedene Merkmale, wie Position und Form der Pupille, erkennen und in den meisten Fällen über die bereits erwähnten Reflexionen der Hornhaut und Pupillen, welche durch gerichtete Lichtquellen entstehen, berechnet und erfasst werden. Hierzu können zwei Video-basierte Eye Tracker unterschieden werden: *stationäre Eye Tracker* und *Head Mounted Displays* (HMD), welche beide vor der Verwendung auf die Testpersonen kalibriert werden müssen. (Holmqvist et al., 2011)

Bei stationären Eye Trackern können solche Modelle unterschieden werden, bei denen der Kopf der Testpersonen während dem Versuch fixiert wird, beispielsweise durch Kinn- und Stirn-Auflageflächen – ähnlich wie bei einem Optiker, und sogenannte Remote Eye Tracker, die ohne Fixierung des Kopfes auskommen, aber etwas ungenauere Ergebnisse liefern können. Bei beiden Methoden sind die Lichtquelle und die Augen-Kamera beim stationären Gerät angebracht. (Holmqvist et al., 2011)

HMD sind mittlerweile meist Eye Tracking Glasses (ETG), können aber auch andere Formen aufweisen. Gemeinsam haben alle HMD die beim Kopf der Testpersonen angebrachte Lichtquelle und Augen-Kamera, und dass die Sicht der Testperson mit einer Szenen-Kamera aufgenommen wird. (Holmqvist et al., 2011)

4.1.3. Mögliche Problembereiche

Es kann zu Problemen bei der Kalibrierung der Eye Tracker durch offensichtliches, wie Brillen oder Kontaktlinsen, kommen, jedoch können dabei auch dicke, lange Wimpern oder stark aufgetragene Mascara stören. Auch permanent erweiterte Pupillen machen eine Kalibrierung unmöglich.

Weiters können äußere Einflüsse einen wesentlichen Einfluss auf die Eye-Tracking-Ergebnisse haben. So sollte der Versuchsraum gut, aber nicht sehr hell ausgeleuchtet sein und auch wechselnde Sonneneinstrahlung kann zu unberechenbaren Ergebnissen führen. Falls die Testpersonen während des Versuchs sitzen, sollte ihnen ein stationärer Sessel und kein klassischer Bürostuhl zur Verfügung gestellt werden, da die Testpersonen sonst zu Bewegungen verleitet werden können. Auch der „Arbeitsplatz“ der Testpersonen sollte, wegen der vorherigen Begründung, frei von Ablenkungen sein. (Holmqvist et al., 2011; Pernice et al., 2015)

4.2. Eye Tracking in Usability-Testungen

Wie bereits in Kapitel 3.4 beschrieben, hat sich Eye Tracking als weitere Methode zur Messung von Usability etabliert, mit der meist neue Designvorschläge für UI evaluiert werden. Für Entwicklerinnen und Entwickler kann es sehr aufschlussreich sein zu wissen, welche Elemente von den Benutzerinnen und Benutzern gesehen und welche übersehen werden. Es ist möglich, Eye-Tracking-Daten mit bereits bekannten Usability-Metriken zu vergleichen, um neue Aspekte darin zu entdecken. Eye Tracking gibt Auskunft über Objekte, die von Usern zuerst angesehen werden, bevor noch die ersten Handlungen durchgeführt werden, oder über Verweildauern in bestimmten Gebieten auf Bildschirmen und zeigt so Informationen auf, die mit traditionellen UX-Methoden nicht zu erfassen gewesen wären. (Olmsted-Hawala et al., 2014)

Das Verwenden von nicht invasiven Eye Trackern (z. B. ohne Kinn-Stütze und mit Infrarot-Kameras) in Usability-Untersuchungen lässt sich meist gut in ein vorhandenes Studiendesign eingliedern, da sich die Testpersonen trotzdem gut auf ihre Aufgaben konzentrieren können und nur eine einfache Kalibrierung vor Beginn der Studie notwendig ist, um starten zu können (Olmsted-Hawala et al., 2014). Während der Studie muss berücksichtigt werden, dass Interaktionen mit den Testpersonen sowie auch Dinge in Reichweite dieser (wie Getränkeflaschen etc.), oder sich bewegende Stühle zu Datenverlusten führen können. (Pernice et al., 2015)

Darüber hinaus gibt es einige weitere Nachteile, die mit der Nutzung von Eye Tracking in Usability-Testungen verbunden sein können. Auch wenn nicht viel Zeit

für das zusätzliche Kalibrieren aufgewendet werden muss, steigen der zeitliche Aufwand und die damit verbundenen Kosten entsprechend der Anzahl der Testpersonen. Die eigentliche Anschaffung der Eye-Tracking-Geräte ist sogar mit signifikanten Kosten verbunden, obendrein muss die entsprechende Software zur Analyse angekauft werden. Zusätzlich kann es der Fall sein, dass für verschiedene getestete Produkte, unterschiedliche Software angeschafft werden muss, da das sonst eventuell notwendige manuelle Auswerten von Fixations-Daten zu ungenau und zu zeitaufwändig wäre. Beim geplanten Einsatz von Eye Tracking neben Thinking Aloud als Testmethode muss beachtet werden, dass es immer noch strittig ist, ob dies ohne Bedenken gemeinsam eingesetzt werden kann. In diesem Zusammenhang wurden unterschiedliche Ergebnisse in Bezug auf Altersgruppen und begleitendem oder retrospektivem Einsatz von Thinking Aloud festgestellt. (Olmsted-Hawala et al., 2014)

Eye Tracking bietet mit einer reichen Vielfalt an Daten die erfasst werden können, wie gesagt, auch Vorteile. Augenbewegungen können emotionale Zustände und kognitive Prozesse widerspiegeln, diese werden beim Eye Tracking meist als Fixationen und Sakkaden erfasst. In bisherigen Untersuchungen wurde festgestellt, dass, abhängig vom Kontext, eine höhere Anzahl von Fixationen auf eine weniger effiziente Such-Strategie und längere Fixationen auf Schwierigkeiten mit der Wahrnehmung des Displays hinweisen (Goldberg et al., 1999; Bojko, 2006; Olmsted-Hawala et al., 2014; Poole et al., 2006). Jedoch können längere Fixationen auch auf ein größeres Interesse am inspierten Objekt hindeuten (Poole et al., 2006). Eye Tracking ist bei einem Vergleich zweier verschiedener Designs ein geeignetes Mittel um zu bestimmen, ob ein Design effizienter ist (Bojko, 2006). Durch die Definition von Areas of Interest (AOIs), spezifische Gebiete oder Inhalte des UI, die für die Usability relevant sind, können Daten speziell zu diesen AOIs zugeordnet und ausgewertet werden. Abhängig vom Studiendesign kann erfasst werden, wohin User als erstes auf einer Seite geblickt haben, wie lange wohin geblickt wurde und ob der Inhalt fokussiert wurde oder schnell zu einem anderen Gebiet am Bildschirm geblickt wurde. (Olmsted-Hawala et al., 2014)

Verschiedene Visualisierungen, wie *Heat Maps*, helfen dabei zu verstehen, womit Interaktionen der Testpersonen stattgefunden haben. In Heat Maps werden mit unterschiedlichen Farben die Menge der Fixationen der Testpersonen oder Dauern, wie lange Gebiete fixiert wurden, visualisiert. Wie in Abbildung 4.1 werden

4. Eye Tracking

in rot meist hohe Anzahlen von Fixationen oder lange Blickdauern dargestellt und die Visualisierung reicht dann über verschiedene Abstufungen bis hin zu grün oder blau für die niedrigsten Werte. Gebiete, die nicht eingefärbt sind wurden nicht fixiert, was aber nicht heißt, dass diese nicht gesehen wurden, sie können genauso sehr kurz oder peripher gesehen worden sein, ohne dass es der Eye Tracker registriert hat. (Olmsted-Hawala et al., 2014)

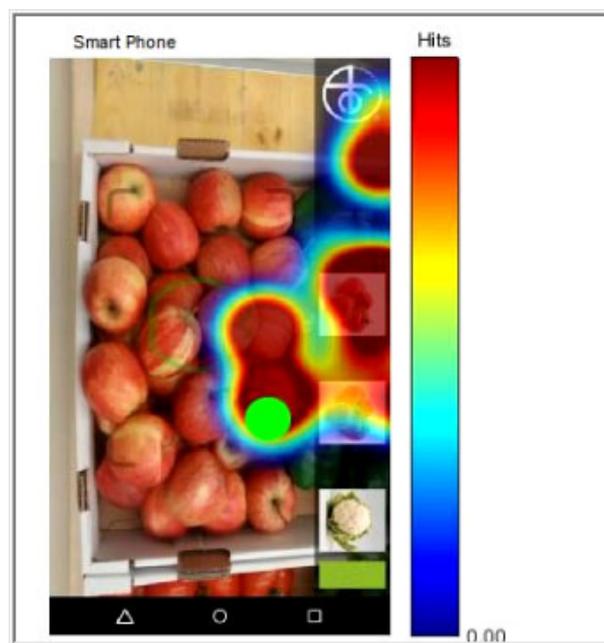


Abbildung 4.1.: Beispiel für eine Heat-Map-Darstellung

Mit *Gaze Plots* bzw. *Scan Paths* werden Fixationen und Sakkaden in einem Zeitabschnitt visuell repräsentiert. Die häufigste Darstellung für Fixationen sind dabei nummerierte Punkte und die Sakkaden dazwischen werden als Linien dargestellt (Abbildung 4.2). Fixationen können auch, je nach Dauer als, entsprechend größere und kleinere Punkte dargestellt werden. Mit nummerierten Gaze-Plots kann mehr oder weniger effizientes Suchen im Inhalt nachvollzogen werden. (Olmsted-Hawala et al., 2014)

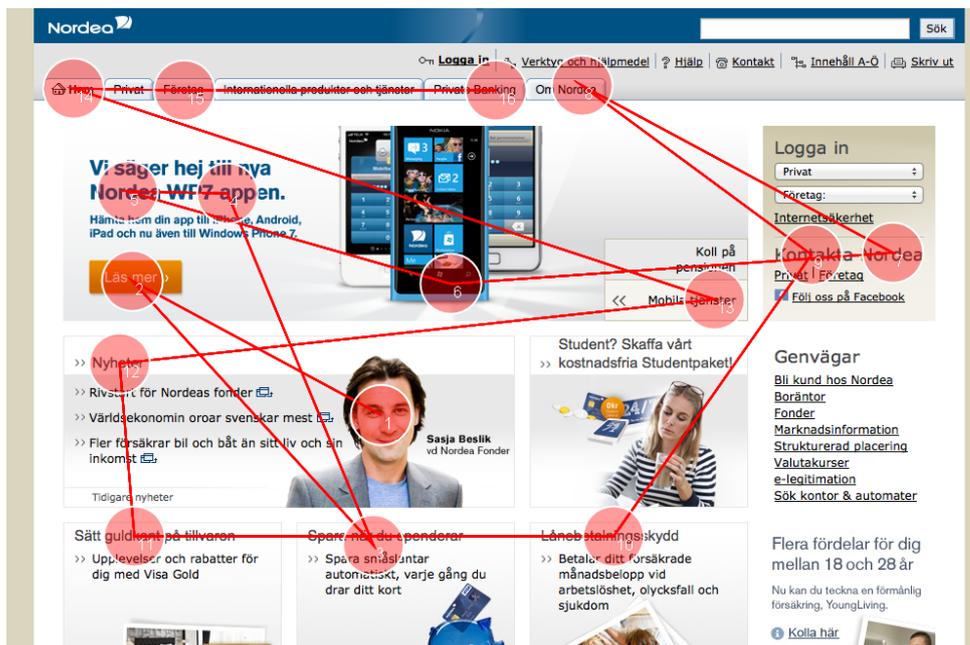


Abbildung 4.2.: Beispiel eines Gaze Plots. Quelle: Per Axbom, CC BY-SA 2.0

4.3. Mobiles Testen

Mobile Endgeräte halten immer mehr Einzug in den Alltag und Firmen versuchen immer stärker eine solide UX mit solchen Geräten zu vermitteln. Daher steigt auch der Bedarf an effektiven Methoden zur Evaluierung der mobilen UX. Durch Usability-Tests mit summativer Evaluation können Informationen zur User Performance, Effizienz und Zufriedenheit mit einem mobilen Produkt wohl am effektivsten bestimmt werden. Eye Tracking mit all den zusätzlichen, nützlichen Informationen zu Abläufen, Fixationen und Dauern, die zuvor beschrieben wurden, kann auch gut eingesetzt werden. Jedoch unterscheiden sich mobile Geräte von stationären Geräten sehr: sie haben viel kleinere Displays, sind im Gegensatz zu den meisten stationären Geräten über Touch Displays steuerbar und werden überall, besonders unterwegs, verwendet. Abgesehen von Designüberlegungen müssen grundlegende Überlegungen zu Darstellungsformen auf unterschiedlich großen mobilen Geräten ebenso bedacht werden, wie zunächst einfach erscheinende Entscheidungen zu Schriftgrößen oder der Einbettung von Navigationselementen unter Berücksichtigung von verschiedenst möglichen Bedienungsformen. (He et al., 2014)

Die Technologie die für mobile Studien eingesetzt werden kann, hat sich in den letzten Jahren enorm entwickelt und es wurden auch bessere Eye Tracker entworfen um mobile Geräte zu testen. Im Vergleich mit Eye Tracking mit Desktop PCs oder Laptops, wobei Daten sehr präzise ausgewertet werden können, steckt die Technologie für mobile Anwendungen jedoch noch in ihren Kinderschuhen. Heute übliche HMD, die diverse ETG einschließen, erlauben die volle Freiheit der Testpersonen in der Verwendung der mobilen Geräte, musste früher das Gerät doch immer fixiert werden, was die UX erheblich einschränkte. Diese völlige Freiheit kann aber auch zu Schwierigkeiten in der Auswertung führen, denn das Bezugssystem der User ändert sich laufend. Daher werden ETG oft für qualitative Datenerhebungen, z. B. bei Shopping Research, eingesetzt und die Anwendungsbereiche im Usability Testing sind limitiert, da für vergleichbare Ergebnisse die Geräte immer noch fixiert werden müssen. Es wird daran gearbeitet Eye-Tracking-Daten künftig direkt über die mobilen Geräte zu sammeln, was, wie bei stationären Geräten, vergleichbare Ergebnisse auch im mobilen Kontext liefern könnte. (He et al., 2014)

Da derzeit diese direkte Verbindung zwischen mobilem Gerät und Eye Tracker noch fehlt, können keine automatischen Auswertungen über die besuchten Seiten/Screens erfolgen, was die manuelle Auswertung und zuvor Segmentierung der Daten erfordert. Mit der künftigen direkten Verbindung wird auch die Auswertung schnell und einfach. Effiziente Analysen setzen derzeit ausgereifte, unterstützende Auswertungs-Software voraus. (He et al., 2014)

4.4. Ausgewählte Forschungsarbeiten

Es ist weit verbreitet Usability-Methoden in verschiedenen Kombinationen einzusetzen, ähnlich wie in dieser Arbeit.

So haben Çöltekin et al. (2009) in ihrer Evaluation von zwei unterschiedlichen Designs von interaktiven online Karten mit 30 Testpersonen einen User-Test mit Fokus der Gewinnung von Usability-Performance-Metriken, Zufriedenheit, Effizienz (Time on Task) und Effektivität (Genauigkeit), in Ergänzung des Einsatzes von Eye Tracking zur Interface-Evaluation durchgeführt. Durch die Auswertung der Performance-Metriken und des verwendeten SUS konnte ein besseres und von den

Testpersonen präferiertes Design identifiziert werden. Jedoch konnten durch die gewonnenen Eye-Tracking-Daten für beide Webseiten gleichermaßen Design-Probleme identifiziert werden. Sie konnten also erfolgreich zeigen, dass traditionelle Usability-Ergebnisse vom Einsatz von Eye-Tracking-Auswertungen profitieren.

Ebenfalls Webseiten wurden von Romano Bergstrom et al. (2013) getestet. Sie führten fünf Usability-Studien zu Webseiten mit der Erhebung von Performance- und Eye-Tracking-Daten bei Personen unterschiedlichen Alters durch, da auf Bedürfnisse älterer Erwachsener beim Design von Webseiten meist nicht eingegangen wird, obwohl die Zahl der älteren Internet-Benutzerinnen und -Benutzer steigt. Für ältere Personen konnten sowohl in den Performance-Messungen als auch bei den Eye-Tracking-Daten alters-bezogene Unterschiede während der Webseiten-Navigation festgestellt werden. Beispielsweise brauchten ältere Testpersonen in zwei der Studien länger als jüngere Testpersonen, um Aufgaben abzuschließen, und blickten öfter auf das Zentrum des Bildschirms. In einer Studie blickten die älteren Testpersonen weniger oft auf den peripheren linken Teil des Bildschirms und brauchten länger, bis sie das erste mal auf den peripheren oberen Teil des Bildschirms blickten. Sie konnten damit das Potential für alters-spezifische Entwicklungen durch die gewählten Methoden aufzeigen und schlossen aufgrund der gemessenen Unterschiede von älteren zu jüngeren Testpersonen, dass die Entwicklung und Forschung in diesem Bereich, auch aufgrund der älter werdenden Gesellschaftsstrukturen, erweitert werden sollte.

Bei beiden genannten Studien wurden Websites auf Desktop PCs getestet, doch wie im vorherigen Kapitel beschrieben, wird auch dem Testen von mobilen Anwendungen immer mehr Aufmerksamkeit geschenkt.

Chynał et al. (2012) testeten die mobile Website des Social-Network-Service facebook.com, einmal emuliert auf einen Desktop PC und einmal regulär auf einem Smartphone, um mit Eye-Tracking-Daten Aussagen zur Genauigkeit und der dementsprechend nützlicheren Aufzeichnungs-Form treffen zu können. Beide Darstellungsformen wurden von je fünf Testpersonen getestet, die die mobile Version der Facebook Website nur teilweise kannten. Chynał et al. (2012) empfehlen Applikationen zusätzlich zu mehreren verschiedenen Smartphones auch emuliert auf PCs zu testen, da dort genauere Ergebnisse durch bessere Eye-Tracking-Analyse-Software

erzielt werden können. Sie stellten fest, dass aufgrund der unterschiedlichen Größe der Darstellung und durch die Verwendung der Computer-Maus zur Auswahl, bei der emulierten Seite die Erfüllung der gestellten Aufgaben länger dauerte als bei der Verwendung des Smartphones. Es wurden keine quantitativen Eye-Tracking-Daten präsentiert, es wurde nur angemerkt, dass durch die Aufnahme des Blicks der Testpersonen Usability-Probleme identifiziert werden konnten.

He et al. (2014) untersuchten eine Anwendung zum Erfassen von Fernsehgewohnheiten auf vom Betriebssystem her unterschiedlichen mobilen Geräten, um Inkonsistenzen zwischen den Plattformen aufzudecken. Eye-Tracking-Daten konnten einige Probleme identifizieren, die ohne den Einsatz von Eye Tracking nicht gefunden worden wären. Deren Testpersonen lasen zum Großteil auf keiner der Plattformen den Einstiegstext auf der Login-Seite, was nach dem Test jedoch nur von einigen Testpersonen ebenfalls so geschildert wurde. Zwischen Android und iOS gab es größtenteils keine Design-Unterschiede, doch unterschied sich die Darstellung der Eingabeseiten für das Aufzeichnen von Beginn und Ende von Fernsehprogrammen. Unter iOS war die Startzeit bei der Eingabe der Endzeit sichtbar und die Eye-Tracking-Daten suggerieren, dass die Testpersonen ihre zuvor eingegebenen Daten verfügbar haben wollen. Diese Annahme bestätigend konnte unter Android das häufige Wechseln zur vorherigen Seite mit den Informationen zur eingegebenen Startzeit beobachtet werden. Eye Tracking hat geholfen Usability-Probleme objektiv zu beurteilen.

In dieser Arbeit wird ebenso erwartet, dass durch die zusätzlichen Daten aus der Eye-Tracking-Auswertung wertvolle Erkenntnisse über klassische Usability-Daten hinaus gewonnen werden können.

5. Seven Stages of Action oder die Psychologie von Alltagshandlungen

Oft scheitert man an der Bedienung eines unbekanntes Gerätes, wobei Norman (2013) das plakativ schlechtem Design zuschreibt, denn gutes Design ignoriert die Bedürfnisse und Psychologie von Menschen nicht (siehe auch Kapitel 2) und braucht keine zusätzlichen Beschriftungen und Beschreibungen. Ursprünglich erschien *The Design of Everyday Things* 1988 als *The Psychology of Everyday Things*, wofür Norman seine Inspiration aus einem England-Aufenthalt und den damit verbunden Problemen mit englischen Wasserhähnen oder Türen erhielt (Norman, 1999). Benötigen einfache Gegenstände Beschriftungen, dann hat das Design-Team versagt.

Um zu sieben grundlegenden Design-Prinzipien zu gelangen (siehe Kapitel 5.3) werden zunächst Prinzipien der Interaktion näher behandelt (siehe Kapitel 5.1) und die sieben Stages of Action eingeführt (siehe Kapitel 5.2).

5.1. Grundlegende Prinzipien der Interaktion

Um mit einem Produkt interagieren zu können, muss man dessen Funktionalität verstehen bzw. entdecken was es macht, wie es funktioniert und welche Operationen möglich sind – wie Norman (2013) summiert: Entdeckbarkeit. Diese Entdeckbarkeit, man könnte wohl auch Sichtbarkeit in diesem Zusammenhang sagen, wird durch

entsprechende Anwendung von fünf fundamentalen psychologischen Konzepten gewährleistet, diese sind im Englischen: *affordances*, *signifiers*, *constraints*, *mappings*, *feedback* und laut Norman (2013) das wichtigste: *the conceptual model*.

5.1.1. Affordanzen

Der Wahrnehmungspsychologe Gibson (1979) prägte das Wort *Affordanz*, das für alle umsetzbaren Eigenschaften zwischen der Welt und einem Agierenden (einer Person, einem Tier, oder sogar einer Maschine oder einem Roboter) steht, somit stellen Affordanzen Beziehungen dar, die natürlich existieren, die aber nicht sichtbar, bekannt oder wünschenswert sein müssen (Norman, 1999; Norman, 2008). Norman (1988) führte den Begriff Affordanz in der Design-Welt ein. Ein Sessel bietet Auflage und ermöglicht daher Sitzen, die meisten Sessel können von einer Person aufgehoben werden (sie ermöglichen Heben), aber manche können nur von sehr starken Personen oder einer Gruppe von Personen aufgehoben werden. Die Existenz einer Affordanz ist also von den Eigenschaften des Objektes und den Fähigkeiten des Agierenden abhängig (Norman, 2013). Im Deutschen sind für Affordanz auch noch die Übersetzungen Angebotscharakter oder Aufforderungscharakter gebräuchlich.

Norman unterschätzte den Anklang dieses Konzeptes – viele Designerinnen und Designer behaupteten durch das Platzieren eines Icons oder Buttons gleich eine Affordanz zu einem System hinzugefügt zu haben. Er stellte klar (Norman, 1999), dass sein Buch *The Psychology of Everyday Things* jedoch von *wahrgenommenen Affordanzen* handle, denn den Designerinnen und Designern ginge es vielmehr darum, welche Aktionen von Benutzerinnen und Benutzern als möglich wahrgenommen werden, als darum, welche Möglichkeiten physisch vorhanden sind.

5.1.2. Signifiers

Die signalisierenden Komponenten von Affordanzen sind *Signifiers* (dt. Anzeiger oder Hinweise). Affordanzen legen fest, welche Aktionen möglich sind – Signifiers präzisieren wo eine Aktion stattfinden soll. Ein Signifier ist also jegliche Markierung, jeglicher Klang oder wahrnehmbarer Hinweis der einer Person geeignetes Han-

deln nahe legt. Manche Signifier sind Kennzeichnungen, wie die Beschriftungen „drücken“, „ziehen“ und „Ausgang“ auf Türen, andere sind wahrgenommene Affordancen, wie ein Türgriff, diese sind bewusst und intentional, Signifier können aber auch zufällig und unabsichtlich sein, wie die Anzahl der Personen auf Bahnsteigen, um festzustellen bzw. abzuschätzen, ob ein Zug verpasst wurde oder nicht (Norman, 2013). Er beschreibt *soziale Signifier* als die Hinweise aus der sozialisierten Welt, die aus Menschen und Technologien besteht. Soziale Signifier ersetzen Affordancen, da sie umfassender und weitreichender sind, und sogar durch ihre Abwesenheit etwas signalisieren können, wie im zuvor genannten Beispiel, dem Fehlen von Personen an Bahnsteigen. Norman geht sogar so weit zu sagen: „*Forget affordances: What people need, and what design must provide, are signifiers*“ (Norman, 2008).

5.1.3. Mapping

Mapping spielt eine große Rolle in Design und Layout von Steuerungen und Displays, und beschreibt, wie in der ursprünglich mathematischen Bedeutung, die Beziehung zwischen den Elementen zweier Mengen von Dingen. Der Zusammenhang einer Steuerung und deren Ergebnissen ist am einfachsten zu verstehen, wenn ein verständliches Mapping zwischen der Steuerung, den Aktionen und dem beabsichtigten Ergebnis geboten wird. Gibt es eine räumliche Übereinstimmung zwischen der Anordnung der Steuerungen und dem zu bedienenden Gerät ist es leicht herauszufinden, wie das Gerät zu bedienen ist. Dies ist ein Beispiel für *natürliches Mapping*. Natürliche Mappings können aber auch biologisch oder kulturell sein, wie der universelle Standard, dass eine Handbewegung nach oben mehr anzeigt und eine Handbewegung nach unten weniger anzeigt, was in Steuerungen genauso umgesetzt werden kann, wie Gruppierung und Strukturierung von Steuerungen und Feedback, was wiederum aus Prinzipien der Wahrnehmung und der Gestalt-Psychologie folgt (Norman, 2013).

5.1.4. Feedback

Feedback ist wichtig, muss unmittelbar erfolgen und geplant werden. Feedback muss priorisiert nach jeder Aktion zurückgegeben werden, aber unwichtige Informationen

müssen unaufdringlich präsentiert werden und wichtige Signale müssen hingegen aufmerksamkeiterregend erscheinen. Viele Produkte ignorieren Feedback jedoch oder haben billige Blink- oder Tonsignale, die nicht richtig interpretiert werden können und meist nur signalisieren, dass etwas passiert ist, aber kaum Informationen darüber vermitteln, was passiert ist, und keinen Hinweis über die weitere Vorgehensweise liefern (Norman, 2013).

5.1.5. Konzeptionelle Modelle

Ein konzeptionelles Modell ist eine meist sehr vereinfachte Erklärung einer Funktionsweise von etwas. Solange es hilfreich ist, muss es weder vollständig noch korrekt sein. Konzeptionelle Modelle helfen beim Verstehen und Erahnen der Funktionsweisen von Dingen und in der Fehlerbehebung, wenn etwas nicht wie geplant verläuft. Ein gutes konzeptionelles Modell hilft die Auswirkungen von Handlungen vorherzusagen und ohne ein solches bleibt einem nichts anderes übrig als auswendig gelerntes blind anzuwenden, was nur so lange gelingt, bis etwas schief geht oder man in eine neue Situation kommt, ab dort würde man ein tieferes Verständnis, ein gutes konzeptionelles Modell, benötigen. Norman meint hier nie detaillierte und komplexe Modelle aus technischen Gebrauchsanweisungen oder Büchern für den technischen Gebrauch, sondern viel mehr *mentale Modelle/Denkmodelle*, die das Verständnis des Benutzers/der Benutzerin selbst widerspiegeln. Das heißt natürlich, dass unterschiedliche Benutzerinnen und Benutzer unterschiedliche Denkmodelle von selben Sachen haben können oder auch, dass ein Benutzer/eine Benutzerin mehrere Denkmodelle von einer Sache hat, die jeweils einen anderen Aspekt der Bedienung betreffen. (Norman, 2013)

5.1.6. Einschränkungen und Konventionen (Constraints and Conventions)

Norman unterscheidet vier Arten von Einschränkungen (Norman, 2013). *Physische Einschränkungen* beschränken die möglichen Tätigkeiten und sind effektiver und nützlicher, wenn sie gut zu sehen und einfach zu verstehen sind, da dann die Anzahl der möglichen Tätigkeiten von Beginn an eingeschränkt ist und nicht erst nach

dem erfolglosen Versuch der Ausführung erkannt wird, dass keine Reaktion folgt. Beim Designen eines graphischen Layouts wird größtenteils auf eine konventionelle Interpretation von Symbolen und Platzierungen vertraut. Die Debatte über den Einsatz von Affordanzen handelt meist von Konventionen, oder wie Norman (1999) sie noch bezeichnet: *kulturelle Einschränkungen*. Kulturelle Einschränkungen sind genauer also Konventionen, die in einer kulturellen Gruppe geteilt werden und sich über die Zeit entwickeln. Symbole und Einschränkungen sind keine Affordanzen, sie sind Beispiele der Verwendung eines gemeinsamen und sichtbaren konzeptionellen Modells, angemessenen Feedbacks und geteilten, kulturellen Konventionen (Norman, 1999). *Semantische Einschränkungen* wurden erst mit späteren Versionen von *The Psychology of Everyday Things* eigens definiert und basieren auf der Bedeutung der Situation, eine Reihe von möglichen Aktionen kontrollieren zu können, sie basieren auf unserem Wissen über die Situation und die Welt. Man setzt sich beispielsweise, aufgrund der sinnvollen Sitzrichtung, mit dem Blick nach vorne auf ein Motorrad. Denkt man an neue Technologien, die Bedeutungen von Dingen ändern, oder die sich verändernden Umgangsweisen mit unseren Technologien und untereinander, wird deutlich, dass Bedeutungen von heute nicht Bedeutungen von morgen sein müssen – diese sind mit der Zeit veränderlich wie auch kulturelle Einschränkungen. (Norman, 2013)

5.2. Seven Stages of Action

Betrachtet man genauer, wie etwas von Personen ausgeführt wird, stößt man schnell auf Fragen die geklärt werden müssen. Diese Fragen sind zwei Klüften (siehe Abbildung 5.1) zuzuordnen: der *Kluft der Durchführung* (Gulf of Execution), in der herausgefunden werden soll wie etwas funktioniert, und der *Kluft der Evaluation* (Gulf of Evaluation), in der herausgefunden werden soll was passiert ist und ob die Handlung zum Ziel geführt hat. Dabei kann die Kluft der Durchführung durch den Einsatz von Signifiers, Einschränkungen, Mappings und einem konzeptionellen Modell überbrückt werden, und die Kluft der Evaluation durch den Einsatz von Feedback und einem konzeptionellen Modell (Norman, 2013). All diese Konzepte wurden zuvor in Kapitel 5.1 beschrieben.

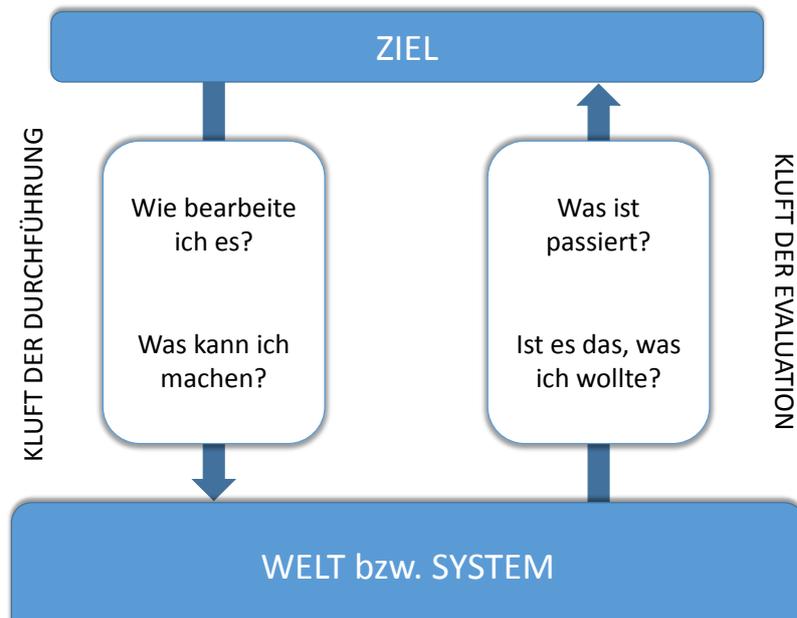


Abbildung 5.1.: Die Klüfte der Durchführung und Evaluation, nach Norman (2013, S. 39)

Das Überbrücken dieser Klüfte führt bei der ausführenden Person zu spezifischen Aktionen bzw. Handlungen, die jeweils in drei Phasen unterteilt werden können. Zusammen mit dem übergeordneten Ziel können also sieben Stages of Action (SSA) (siehe Tabelle 5.1) definiert werden: nach der Formulierung des Ziels, planen, spezifizieren und ausführen auf Seiten der Durchführung, und wahrnehmen, interpretieren und vergleichen auf Seiten der Evaluation (siehe Abbildung 5.2) (Norman, 2013).

- | | |
|--|--|
| 1. <i>Ziel</i> (formulieren) | 5. <i>Wahrnehmen</i> (des Status der Welt) |
| 2. <i>Planen</i> (der Handlung) | 6. <i>Interpretieren</i> (der Wahrnehmung) |
| 3. <i>Spezifizieren</i> (einer Handlungssequenz) | 7. <i>Vergleichen</i> (des Ergebnisses mit Ziel) |
| 4. <i>Ausführen</i> (der Handlungssequenz) | |

Tabelle 5.1.: Beschreibung der sieben Stages of Action, entnommen aus Norman (2013, S. 41)

Nicht alle Handlungen in den Phasen werden bewusst ausgeführt und bei vielen Anwendungen wird es nicht nötig sein alle Phasen der Reihe nach zu durchlaufen (Norman, 2013). Weiters sind wohl nicht alle Phasen sauber trennbar (Norman, 1986, S. 41).

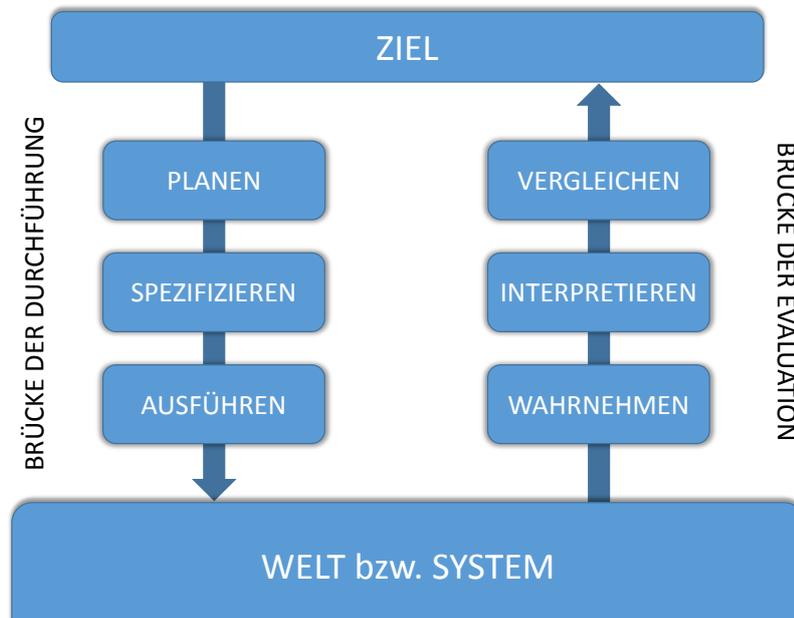


Abbildung 5.2.: Der Seven Stages of Action Cycle, nach Norman (2013, S. 41)

5.3. SSA: Sieben grundlegende Design-Prinzipien

Das Sieben-Phasen-Modell des Action Cycle liefert einen Leitfaden, um neue Produkte oder Services zu entwickeln. Aus dem Modell kann eine elementare Checkliste von Fragen abgeleitet werden, Abbildung 5.3 bietet eine Zusammenfassung der Fragen (Norman, 2013):

1. Was sind die Alternativen?
2. Was kann ich machen?
3. Wie mache ich es?
4. Was ist passiert?
5. Was heißt das?
6. Ist das okay? Habe ich mein Ziel erreicht?

Die Person die für das Design eines Systems zuständig ist, ist dafür verantwortlich, dass alle Fragen zu jeden Zeitpunkt der Systembenutzung beantwortet werden können. Die in Abbildung 5.3 eingeführten Begriffe *Feedforward* bzw. *Feedback* sind die Informationen, die helfen die Fragen der Durchführung zu beantworten bzw. die



Abbildung 5.3.: Die Seven Stages of Action als Design-Hilfen, nach Norman (2013, S. 71)

unterstützen das Geschehene zu verstehen. Feedforward wird durch den geeigneten Einsatz von Signifieren, Einschränkungen, Mappings erreicht, und Feedback durch explizite Information über die Auswirkung der Handlung. In beiden Fällen spielt das konzeptionelle Modell eine große Rolle (Norman, 2013).

Zusammenfassend führt dies zu sieben grundlegenden Design-Prinzipien, die in Tabelle 5.2 dargestellt sind (Norman, 2013).

1. **Entdeckbarkeit.** Es ist möglich herauszufinden welche Handlungen möglich sind und was der Status des Geräts ist.
2. **Feedback.** Es gibt volle und laufende Information über die Auswirkungen der Handlungen und den aktuellen Status vom Produkt oder Service. Nach dem Durchführen einer Handlung ist es einfach den neuen Status festzustellen.
3. **Konzeptionelles Modell.** Das Design bildet die gesamte Information die gebraucht wird ab, um ein gutes konzeptionelles Modell des Systems zu bekommen, was zu Verstehen und einem Gefühl der Kontrolle führt. Das konzeptionelle Modell verbessert sowohl die Entdeckbarkeit, als auch die Evaluation der Ergebnisse.
4. **Affordanzen.** Die richtigen Affordanzen zur Ermöglichung der gewünschten Handlungen existieren.
5. **Signifiers.** Effektiver Einsatz von Signifiern stellt Entdeckbarkeit und gut kommuniziertes und verständliches Feedback sicher.
6. **Mappings.** Die Beziehung zwischen Steuerungen und deren Auswirkung folgt den Grundlagen guten Mappings, so umfassend verbessert wie möglich durch räumliches Layout und zeitlichen Zusammenhang.
7. **Einschränkungen.** Das zur Verfügung stellen von physischen, logischen, semantischen und kulturellen Einschränkungen leitet Handlungen und vereinfacht die Interpretation.

Tabelle 5.2.: Beschreibung der sieben grundlegenden Design-Prinzipien, entnommen aus Norman (2013, S. 71f)

6. Forschungsmaterialien bzw. -methoden

In diesem Kapitel werden die verwendeten Materialien und Methoden in der Reihenfolge des Einsatzes in der Studie beschrieben. Vorab wird die Testumgebung, die App „MANGO“ (siehe Kapitel 6.1), mit ihren Funktionen beschrieben. Für den Usability-Test wurde eine Vorbefragung erstellt (siehe Kapitel 6.2), es wurden Aufgaben für die Testpersonen entwickelt (siehe Kapitel 6.3) und ein abschließender Fragebogen zur App-Bewertung (siehe Kapitel 6.4) erarbeitet. Zum Schluss werden die auszuwertenden Parameter seitens der Usability-Benchmark-Erstellung (siehe Kapitel 6.5.1) und des Eye Trackings in der Aktivitäts-Analyse (siehe Kapitel 6.5.2) zusammengefasst.

6.1. Testumgebung „MANGO-App“

Getestet wird eine Closed-Beta-App für Android mit dem Arbeitstitel „MANGO“ in der Version 0.8.5. Die App wurde im gleichnamigen FFG-Projekt „MANGO – Mobile Augmented Reality for Nutrition Guidance and Food Awareness“ von JOANNEUM RESEARCH (JR) entwickelt. Der App liegt die Ernährungsphilosophie „functional eating“ (FE) zugrunde, die aus sieben individuellen Ernährungssäulen oder -typen besteht, wobei vier davon, nämlich Beauty Food, Brain Food, Power Food und Soul Food, in der App umgesetzt sind.

Das technische Kernstück der MANGO-App bildet der entwickelte Prototyp eines mobilen Klassifikators zur automatischen, Video-basierten Lebensmittelerkennung, der auf einem Random-Forest-Klassifikationsverfahren basiert. Die Methode wurde für

die 35 Lebensmittel-Klassen aus Abbildung 6.1 mit 150 Bildern pro Klasse mit einem Desktop PC trainiert und durch den Random Forest mit einer 50%/50% Aufteilung evaluiert. Der Prototyp des mobilen Klassifikators wurde mit einem Smartphone auf einem reduzierten Datensatz mit 12 Lebensmittel-Klassen mit einer 80%/20% Aufteilung trainiert und evaluiert. Die Klassen des reduzierten Datensatzes sind in Abbildung 6.1 blau markiert. Mit dem mobilen Klassifikator wurde eine Erkennungsrate von 80,30 % im Erst-Vorschlag erreicht. Werden zusätzlich der zweit- bzw. drittbeste Vorschlag berücksichtigt, erhöht sich die Erkennungsrate auf 92,28 % bzw. auf 97,50 % (Waltner et al., 2015). Dieser Dreivorschlag wird als Ergebnis in der App dargestellt (siehe Abbildung 6.3).



Abbildung 6.1.: Die 35 Lebensmittel-Klassen, auf die der Klassifikator nach Random-Forest-Ansatz trainiert wurde, entnommen aus Waltner et al. (2015). Die blauen Markierungen entsprechen den 12 Lebensmittel-Klassen des Prototypen des mobilen Klassifikators.

Über den Startbildschirm (Abbildung 6.2a) der App kann auf die drei Hauptfunktionen zugegriffen werden, die in Tabelle 6.1 beschrieben sind. Die Bildschirme der Hauptfunktionen sind wie in Abbildung 6.2b gestaltet. Der Kameramodus startet im Querformat mit der Anzeige eines Kamerasuchers, der in der Mitte den Umriss eines Apfels zeigt, ein Signifier (siehe Kapitel 5.1.2), der die Testpersonen zum Verwenden mit Lebensmitteln auffordern bzw. hinweisen soll. Die Bildererkennung geschieht kontinuierlich, die Testpersonen müssen lediglich die Kamera auf die Lebensmittel richten. Nach der Klassifikation werden durch die mobile Augmented-Reality-Funktion die drei wahrscheinlichsten Ergebnisse mit entsprechenden Bildern und dazugehöri-

gen kleinen Balken rechts oben am Display (Abbildung 6.3) dargestellt. Die Balken dienen dabei der Erstinformation zur Passgenauigkeit des Lebensmittels im eigenen FE-Typ, Detailinformationen (Abbildung 6.4) zum jeweiligen Lebensmittel sind durch Auswählen des Lebensmittel-Bildes verfügbar.

Funktion	Beschreibung
Kamera	automatische Lebensmittelerkennung
Lebensmittel	Zugriff auf die Lebensmitteldatenbank, die nach FE-Ernährungstypen unterteilt ist
functional eating	Information über das Ernährungskonzept FE und die vier implementierten Ernährungstypen

Tabelle 6.1.: Beschreibung der drei Hauptfunktionen der MANGO-App.

Die Lebensmitteldatenbank ist in vier Tabellenreiter für die FE-Typen mit jeweils elf Lebensmittel-Kategorien (wie Gemüse, Obst, Milchprodukte) unterteilt (Abbildung 6.2b, links unten). Die Detailinformationen zu einzelnen Lebensmitteln (Abbildung 6.4) sind über diesen Menüpunkt oder über die Ergebnisse der Bilderkennung (Abbildung 6.3) verfügbar. Unter dem Menüpunkt FE stehen nach der Einführung zur Ernährungsphilosophie FE (Abbildung 6.2b, rechts unten) weitere Informationen zu den einzelnen FE-Typen zur Verfügung (Abbildung 6.5). Die dort verwendeten Farben spiegeln sich als Farben der Tabellenreiter der Lebensmitteldatenbank (Abbildung 6.2b, links unten) und Füllfarben der Balken seitlich der Scan-Ergebnisse (Abbildung 6.3) wider.

Die Entwicklung der App war nicht Teil dieser Diplomarbeit, die App wurde von JR in der Version 0.8.5 auf einem Motorola Moto G mit dem Betriebssystem Android 5.0 für die Testung zur Verfügung gestellt.

6.2. Fragebogen „Vorbefragung“

Vor der Teilnahme am eigentlichen Usability-Test wird eine Vorbefragung durchgeführt. Einerseits, um mehr über die Testpersonen zu erfahren und um Aspekte der App lediglich aufgrund einer Beschreibung beurteilen zu lassen, andererseits, um die den Typ nach FE zu bestimmen, da der FE-Typ der Testpersonen im praktischen Teil benötigt wird. Die Vorbefragung lässt sich also vom Aufbau her in zwei



Abbildung 6.2.: (a) Startbildschirm der MANGO-App (b) Bildschirme der drei Hauptfunktionen der MANGO-App, oben: Kamera, unten links: Lebensmittel (mit Auswahl Brain Food und geöffneter Liste Obst), unten rechts: functional eating



Abbildung 6.3.: Als Ergebnis eines Lebensmittel-Scans werden am oberen Bildschirmrand die besten drei Ergebnisse der Bilderkennung zusammen mit kleinen Balken, die die Passgenauigkeit im eigenen FE-Typ (hier Brain Food) angeblendet.

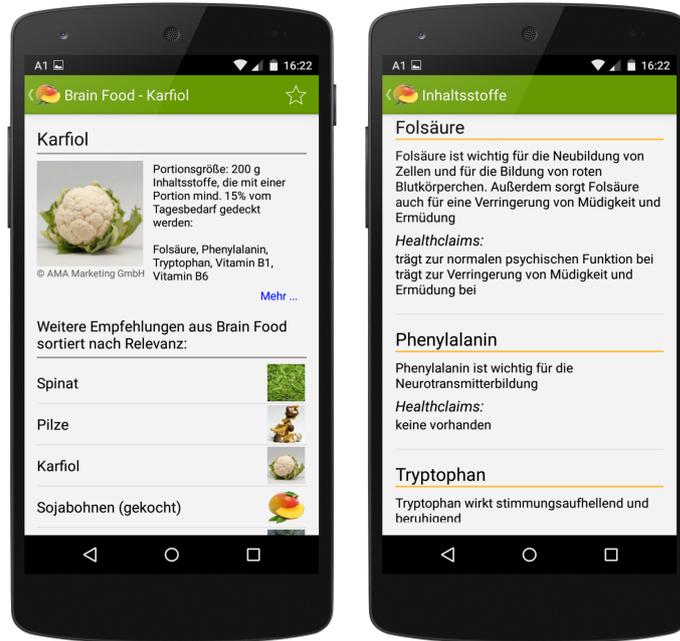


Abbildung 6.4.: links: Detailansicht eines Lebensmittels (Karfiol mit FE-Typ Brain Food), rechts: weiterführende Informationen zu den Inhaltsstoffen (durch Auswahl von „Mehr ...“ nach der Kurzinformation)

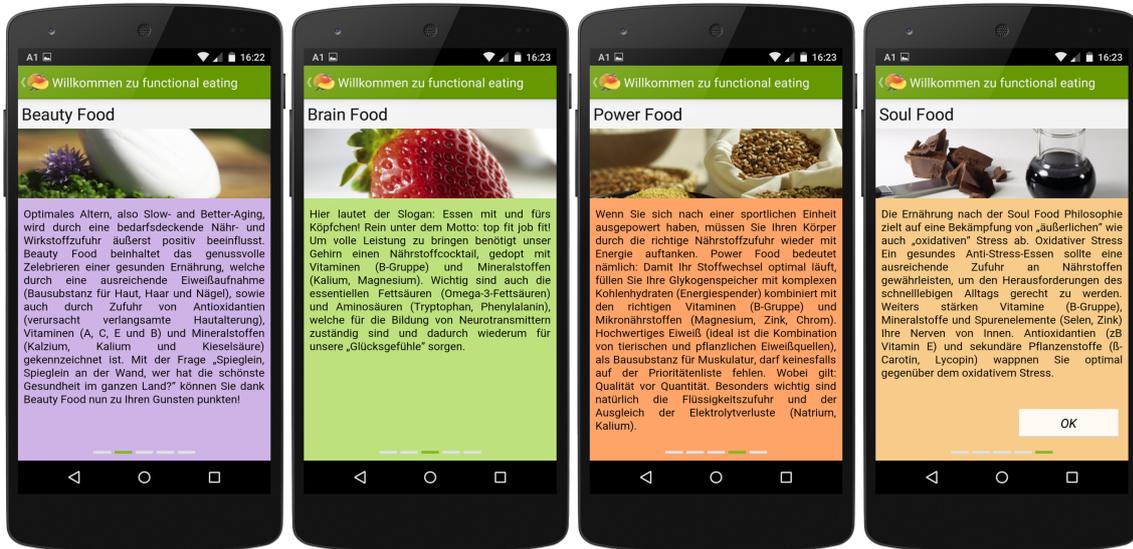


Abbildung 6.5.: FE-Typenbeschreibungen umgesetzt in der App, wobei sich die Farbgebung auch in der Lebensmitteldatenbank sowie in den Balken der Scanergebnisse im Kameramodus wieder findet.

Teile unterteilen. Der erste Teil beginnt mit einer kurzen Einleitung zu Ernährung und deren Auswirkung auf den Lebensstil, und dem Szenario, dass Informationen zur individuell angepassten Ernährungsform in einer App verfügbar wären. Zur in der Theorie beschriebenen App MANGO gibt es zwei einleitende Fragen zum Herunterladen der App, nämlich eine Bewertung zum wahrscheinlichen Herunterladen und Verwenden der App auf einer Skala von 5 bis 0 (äußerst wahrscheinlich bis gar nicht wahrscheinlich), und eine Frage nach Gründen für das Herunterladen mit acht Antwortmöglichkeiten und freien Antwortfeldern, wobei Mehrfachantworten möglich sind. Die darauf folgende Frage der Selbsteinschätzung des eigenen Lebensstils in Zusammenhang mit Ernährung hat die Antwortmöglichkeiten Beauty Food (innere Schönheit und Hautalterung), Brain Food (Konzentration, geistige Leistungsanforderungen), Power Food (Sport, körperliche Anstrengungen) und Soul Food (Stress, hektischer Alltag), genau jene FE-Typen, die auch in der App realisiert sind (siehe Kapitel 6.1). Eine letzte Frage nimmt Bezug auf die App, nämlich die Frage nach der Bereitschaft, persönliche Daten bekanntzugeben, um Informationen zur individuell angepassten Ernährungsform zu erhalten. Darauf folgt je eine Frage zur eigenen Ernährung und dem Erledigen von Einkäufen im Haushalt sowie zwei Einstufungen zur Erfahrung mit Informationstechnologie bzw. Tablets und Smartphones, die auf einer Skala von 1 bis 8 (sehr gering bis sehr groß) zu bewerten sind, gefolgt von demographischen Fragen, die den ersten Teil des Fragebogens abschließen.

Der zweite Teil, die *Typenzuordnung nach functional eating*, beinhaltet 20 Fragen zu Stress- und Aktivitätsleveln, typischen beruflichen Anstrengungen oder persönlichen Vorlieben. Diese Typenzuordnung (inklusive Auswertungsvorgabe) wurde vom EST entwickelt zur Verfügung gestellt und wird während der Studie benötigt, um den persönlichen Ernährungstyp zu ermitteln und schließlich diesen in der App hinterlegen zu können. Diese Typenzuordnung soll in späteren Versionen der App implementiert werden. Der gesamte Fragebogen ist im Anhang A ab Seite 126 zu finden.

6.3. Aufgabenbeschreibung

Die Aufgabenstellungen wurden so gewählt, dass alle drei Hauptfunktionen der App (Tabelle 6.1) verwendet werden und von den Testpersonen auch in die Tiefe navigiert werden muss. Es handelt sich um typische Verwendungsbeispiele der App, wie von Nielsen (1993) vorgeschlagen. Die Entwicklung der Szenarien (siehe Tabelle 6.2) wurde in dieser Diplomarbeit wie in Barnum (2010, S. 128ff) durchgeführt, mit der Modifikation, dass die Aufgabenstellungen für die Testpersonen bei allen Szenarien separat hervorgehoben sind.

Kategorie	Beschreibung
Frage 1	Werden User ihre Suche nach der Beschreibung zu ihren FE-Typen beim Menüpunkt <i>functional eating</i> beginnen?
Tätigkeit 1	Nachlesen der eigenen FE-Typenbeschreibung. Richtiger Weg über Menüpunkt <i>functional eating</i> , der mit einer Beschreibung der Ernährungsphilosophie beginnt und die vier FE-Typenbeschreibungen enthält, die durch Wischen von rechts nach links am Display verfügbar sind. Weiterer Weg über das FE-Symbol im Kameramodus, der zu diesem Zeitpunkt jedoch noch nicht bekannt sein sollte.
Szenario 1	Im Supermarkt wird aktuell eine App für Smartphones beworben, die helfen soll, dass Sie sich passend zu Ihrem Lebensumfeld ernähren und so ihr volles Potential, ihre volle Leistung ausschöpfen können. Sie haben sich die App heruntergeladen und installiert. Ihr Einstufungstest hat ergeben, dass Sie nach functional eating am ehesten dem Typ Beauty Food/Brain Food/Power Food/Soul Food entsprechen. In der App wurde Ihr Ernährungstyp bereits eingestellt und gespeichert.
Aufgabe 1	Verwenden Sie die App und lesen sich Ihre Typenbeschreibung, also Beauty Food/Brain Food/Power Food/Soul Food, durch!

Fortsetzung auf der nächsten Seite

Tabelle 6.2.: Szenario-Entwicklung der Testung, die mit einer Frage startet, aus der eine Tätigkeit formuliert und schließlich das Szenario in Ergänzung der Aufgabe für die Testpersonen entwickelt wird, nach Barnum (2010, S. 128ff).

Kategorie	Beschreibung
Frage 2	Kann der Kameramodus zur Bildererkennung ohne weitere Vorkenntnisse verwendet werden?
Tätigkeit 2	Verwendung des Kameramodus zur Lebensmittelerkennung. Einzige Möglichkeit über Menüpunkt <i>Kamera</i> .
Szenario 2	In der Werbung zur App haben Sie weiters gelesen, dass man mit der App Obst oder Gemüse scannen kann, um dann weitere Informationen zu erhalten. Sie wollen nun diese Funktion ausprobieren.
Aufgabe 2	Probieren Sie die Funktion „Lebensmittel scannen“ bei drei Lebensmitteln aus und lernen Sie die App kennen!
Frage 3	Verwenden die User den Kameramodus um die aufgebauten Lebensmittel auf ihre FE-Typ-Passgenauigkeit zu untersuchen? Stimmt das Ergebnis?
Tätigkeit 3	Überprüfung der sechs aufgebauten Lebensmittel auf ihre FE-Typ-Passgenauigkeit. Naheliegender Weg der Entwickler: Verwendung des Kameramodus und des Balken-Feedbacks oder der Detailansicht der Lebensmittel, bei der dann auch eine nach Relevanz sortierte Liste verfügbar ist. Weiterer Weg über Menüpunkt <i>Lebensmittel</i> , Auswahl des eigenen FE-Typs, der Lebensmittelkategorie und des Lebensmittels, und wiederum über die nach Relevanz sortierte Liste.
Szenario 3	Sie wollen nun erfahren, welche Lebensmittel am besten Ihren Ernährungstyp und somit Lebensstil unterstützen. Wie bereits gesagt wurde in der App Ihr Ernährungstyp bereits eingestellt und gespeichert.
Aufgabe 3	Finden Sie heraus welches der aufgebauten Lebensmittel am besten zu Ihrem Lebensstil passt. Nennen Sie Ihr Ergebnis wenn Sie fertig sind!

Fortsetzung auf der nächsten Seite

Tabelle 6.2.: Szenario-Entwicklung – Fortsetzung

Kategorie	Beschreibung
Frage 4	Wird auch die Lebensmittel-Datenbank verwendet? Stimmt das Ergebnis?
Tätigkeit 4	Die drei wichtigsten Gemüsesorten des eigenen FE-Typs finden. Über den Menüpunkt <i>Lebensmittel</i> und Auswahl des eigenen FE-Typs, weiter zu Gemüse. Die erste Liste ist alphabetisch sortiert. Auswahl eines Gemüses, um die nach Relevanz sortierte Liste zu erreichen. Ist noch ein Ergebnis im Kameramodus verfügbar, kann dort ein Gemüse ausgewählt werden und so die nach Relevanz sortierte Liste ebenso erreicht werden.
Szenario 3	s.o.
Aufgabe 4	Finden sie die drei wichtigsten Gemüsesorten ihres Ernährungstypes Beauty Food/Brain Food/Power Food/Soul Food. Nennen Sie Ihr Ergebnis wenn Sie fertig sind!

Tabelle 6.2.: Szenario-Entwicklung – Fortsetzung

Die Szenarien und Aufgaben aus Tabelle 6.2 bilden die Aufgabenbeschreibung wie sie in der Studie verwendet wurde, welche komplett im Anhang B auf Seite 131 zu finden ist. Alle Anweisungen wurden mit dem zuvor (aus der Vorbefragung 6.2) bestimmten FE-Typ der Testperson vorgelesen, die Einleitung „Sie bekommen einzelne Szenarien von mir beschrieben und jeweils Aufgaben dazu genannt. Bei keiner Aufgabe gibt es ein Zeitlimit.“ wurde mündlich um den Hinweis „Immer wenn Sie bereit sind, setzen wir mit der nächsten Aufgabe fort.“ ergänzt.

6.4. Fragebogen „App-Bewertung“ (Bewertungsskalen)

Für die Bewertung der Usability wurden zwei frei verfügbare Fragebögen ausgewählt. Der *User Experience Questionnaire* (UEQ) nach Laugwitz, Schrepp et al. (2006) (siehe Kapitel 3.2.2), zur Bewertung der User Experience (UX) einerseits

und der *System-Usability-Scale*-Fragebogen (SUS) nach Brooke (1996) (siehe Kapitel 3.2.1) zur Bewertung der Usability andererseits. Der SUS wurde in der deutschen Version von Rauer (2011) verwendet, wobei jeweils das Wort *System* mit *App* ersetzt wurde. Die verwendeten Aussagen finden sich als Gegenüberstellung zu den originalen englischen Aussagen in Tabelle 3.1. Da durch den SUS die Usability in einer Zahl (zwischen 1 und 100) ausgedrückt und der Wert nach Bangor et al. (2008) auch sehr gut interpretiert werden kann (siehe Abbildung 3.1), wurde der arrivierte SUS als Bewertungstool gewählt. Der UEQ wurde in Ergänzung gewählt, da eine Bewertung der UX bei Produkten für Endverbraucherinnen und -verbraucher von größerer Bedeutung ist (Sarodnick et al., 2011) und da durch den UEQ Ergebnisse auf gleich sechs Dimensionen, nämlich Durchschaubarkeit, Effizienz, Originalität, Steuerbarkeit, Stimulation sowie Attraktivität, erlangt werden. Der UEQ liefert für jede Dimension ein Ergebnis zwischen -3 und 3, wobei ein Ergebnis $> 0,8$ einer positiven Beurteilung entspricht und ein Ergebnis von 1,5 als sehr guter Wert interpretiert werden kann. Ein gutes Kriterium, um abzuschätzen wie gut ein Produkt ist, ist der Vergleich mit anderen Produkten. Durch den UEQBenchmark (Schrepp et al., 2013), der Ergebnisse aus 163 Studien zusammenfasst, können detailliertere Aussagen zur Qualität der App im Vergleich zu anderen Produkten getroffen werden.

Anschließend wird für die App-Bewertung noch einmal die Frage gestellt, wie wahrscheinlich die Testpersonen die App herunterladen und verwenden würden, und die Befragung schließt mit dem *IPIP Big Five Persönlichkeitstest*¹ ab, welcher für weitere Auswertungen durch JR angefügt wurde und dessen Ergebnisse nicht näher in dieser Arbeit behandelt werden. Der gesamte Fragebogen ist im Anhang C ab Seite 132 zu finden.

6.5. Auszuwertende Parameter

Es werden einerseits klassische Usability-Merkmale (siehe Kapitel 3) ausgewertet und andererseits wird eine Aktivitäts-Analyse nach dem Konzept der sieben Stages of Action (siehe Kapitel 5) durchgeführt, bei der Eye-Tracking-Parameter bestimmt

¹http://ipip.ori.org/New_IPIP-50-item-scale.htm (zuletzt besucht am 21.8.2015)

werden. Nachfolgend werden die Kennzahlen für den Usability Benchmark erarbeitet und beschrieben.

6.5.1. Usability Benchmark

Die vier Aufgaben der Testpersonen in der Studie (siehe Kapitel 6.3) beinhalten mehrere Kernaufgaben, die für die Benchmark-Erstellung identifiziert werden mussten, um messbare und auf sinnvolle Weise vergleichbare Ergebnisse zu erhalten. Damit ist beispielsweise gemeint, dass die erste Aufgabe, das Durchlesen der eigenen FE-Typenbeschreibung, beim Messen der Dauer dieser Aufgabe nun auch die individuelle Lesezeit der Testperson beinhalten würde. Besser geeignet ist hier das Messen der Zeit bis zum Auffinden der eigenen FE-Typenbeschreibung, was also die Kernaufgabe (KA) der ersten Aufgabe darstellt. In dieser Diplomarbeit wurden alle Aufgaben in Kernaufgaben unterteilt, wie in Tabelle 6.3 dargestellt.

Zu den einzelnen Kernaufgaben wurden weiters jeweils ein oder mehrere Benchmark-Ziele definiert, außerdem wurden Gesamt-Ziele formuliert (siehe Tabelle 6.4). Für das Setzen der Ziele wurden beispielsweise die Aufgaben selbst durchgeführt und die benötigte Zeit gestoppt. Erhöht man diese Zeit um den Faktor zwei bis drei, können für Testpersonen realistische Größen entstehen (Berkun, 2003). Ähnliche Vorgehensweisen können auch für andere Parameter angewandt werden. Der angegebene Erfüllungsgrad spiegelt eine Mischung aus dem angenommenen Schwierigkeitsgrad, einer Priorisierung und Erfahrungswerten wider.

Einige Überlegungen und Interpretationshilfen zur Zielfestlegung werden nachfolgend angeführt. Die KA 1 und 2 sollten für die meisten Testpersonen leicht zu bewältigen sein, es wurde angenommen, dass 90 % von ihnen direkt den richtigen Menüpunkt wählen, und es wurde eine Zeit von 15 bzw. 10 Sekunden als Ziel festgesetzt. Die Bewältigung der KA 3 wird mehr Zeit in Anspruch nehmen. Insgesamt wurden den Testpersonen keine Zeitvorgaben gesetzt, daher wurde die Zeit bis zum Abschluss von drei Lebensmittel-Scans betrachtet, ohne zu berücksichtigen ob die Erkennung richtig oder falsch war, und angenommen, dass 75 % der Testpersonen dies in 1 Minute schaffen werden. Die restlichen Annahmen betreffen die Funktionsweise des Kameramodus, nämlich die richtigen Scan-Ergebnisse und einen stabilen

Aufgabe 1	Verwenden Sie die App und lesen sich Ihre Typenbeschreibung, also Beauty Food/Brain Food/Power Food/Soul Food, durch!
KA 1	Auffinden der eigenen FE-Typenbeschreibung.
Aufgabe 2	Probieren Sie die Funktion „Lebensmittel scannen“ bei drei Lebensmitteln aus und lernen Sie die App kennen!
KA 2	Öffnen des Kameramodus, um Lebensmittel zu scannen.
KA 3	Durchführen von 3 (oder mehr) Lebensmittel-Scans im Kameramodus.
Aufgabe 3	Finden Sie heraus welches der aufgebauten Lebensmittel am besten zu Ihrem Lebensstil passt. Nennen Sie Ihr Ergebnis wenn Sie fertig sind!
KA 4	Ausfindig machen des am besten passenden Lebensmittels (einer Auswahl von Lebensmitteln vor Ort) im eigenen FE-Typ.
Aufgabe 4	Finden sie die drei wichtigsten Gemüsesorten ihres Ernährungstypes Beauty Food/Brain Food/Power Food/Soul Food. Nennen Sie Ihr Ergebnis wenn Sie fertig sind!
KA 5	Ausfindig machen der drei wichtigsten Gemüsesorten im eigenen FE-Typ.

Tabelle 6.3.: Unterteilung aller Aufgaben der Testpersonen in messbare Kernaufgaben.

Kameramodus. Da alle Testpersonen unterschiedlich viele Scan-Versuche machen können, wird hierfür die Gesamtanzahl der durchgeführten Scans und der Anteil der richtigen Ergebnisse pro Testperson betrachtet. Es wurde angenommen, dass der Mittelwert dieser Werte 70 % erreichen wird. Außerdem wurde davon ausgegangen, dass zumindest 55 % der Testpersonen bei den Scan-Vorgängen in KA 3 ausschließlich korrekte Ergebnisse erhalten werden und dass der Kameramodus für 90 % der Testpersonen stabil, ohne unerwartet abzustürzen, läuft. Die KA 4 und 5 ähneln sich, sodass angenommen wurde, dass 75 % der Testpersonen die KA 4 in 2 Minuten abschließen bzw. die KA 5 zumindest teilweise abschließen werden können. Es wurde definiert, dass Testpersonen die KA 5 teilweise abschließen, wenn sich zumindest zwei der genannten Lebensmittel, ohne Rücksicht auf die genannte Reihung, unter den Top drei Gemüsesorten befinden. Es wurde davon ausgegangen, dass mindestens 50 % der Testpersonen die KA 5 sogar innerhalb von 1 Minute abschließen können, da bereits Erfahrung aus der KA 4 mitgenommen wird.

Für die Gesamt-Ziele der Testung wurde angenommen, dass die Testpersonen zumindest 80 % der KA erfolgreich abschließen werden und die App mit einem durchschnittlichen SUS-Score von mindestens 80 % bewerten werden. Außerdem soll eine Performance der Testpersonen über das individuelle Erreichen der Ziel-Schwellenwerte ermittelt werden, dafür wurden die Ziele 1a, 1b, 2a, 2b, 3a, 3b, 4a, 5b ausgewählt. Das Ziel 3b wurde hier, im Gegensatz zur Bewertung im Benchmark, für jede Testperson als erreicht oder nicht erreicht bewertet. Die Ziele 3d, 3c und 5a wurden für die Performance der Testpersonen hier nicht berücksichtigt, da sich 3d (Stabilität Kameramodus) rein auf die Performance des Kameramodus bezieht und 3c bzw. 5a bereits durch die schwächer formulierten Ziele 3b bzw. 5b abgebildet sind. Die Fehlerfreiheit der Scan-Ergebnisse wurde inkludiert, da die Testpersonen diese stark beeinflussen können, beispielsweise durch die Dauer beim Wechseln zwischen den Lebensmitteln. Die wesentlichen Usability-Metriken (siehe Kapitel 6) sind bei diesem Usability Benchmark die Time on Task, der Task Success, die Bewertung der App durch die Testpersonen und die Fehlerfreiheit des Systems. Zur Gesamtbeurteilung des Systems wurden diese Metriken zusammengefasst, wofür auch die Ergebnisse des SUS und UEQ herangezogen wurden (siehe Kapitel 8).

KA 1	Auffinden der eigenen FE-Typenbeschreibung.
Ziel 1a	90 % der User werden bei der KA 1 ohne Umwege direkt den Menüpunkt functional eating ansteuern.
Ziel 1b	75 % der User werden die KA 1 innerhalb von 15 Sekunden abschließen können.
KA 2	Öffnen des Kameramodus, um Lebensmittel zu scannen.
Ziel 2a	90 % der User werden bei der KA 2 ohne Umwege direkt den Kameramodus öffnen.
Ziel 2b	75 % der User werden die KA 2 innerhalb von 10 Sekunden abschließen können.
KA 3	Durchführen von 3 (oder mehr) Lebensmittel-Scans im Kameramodus.
Ziel 3a	75 % der User werden die ersten drei Lebensmittel-Erkennungen (egal ob richtig oder falsch) in der KA 3 innerhalb von 1 Minute abschließen können.
Ziel 3b	Der Anteil der richtigen Scan-Ergebnisse pro User in der KA 3 wird im Mittel mindestens 70 % erreichen.
Ziel 3c	55 % der User werden beim Scannen in der KA 3 ausschließlich richtige Ergebnisse erhalten.
Ziel 3d	Der Kameramodus funktioniert in der KA 3 bei 90 % der User ohne abzustürzen.
KA 4	Ausfindig machen des am besten passenden Lebensmittels (einer Auswahl von Lebensmitteln vor Ort) im eigenen FE-Typ.
Ziel 4a	75 % der User werden die KA 4 innerhalb von 2 Minuten erfolgreich abschließen können.

Fortsetzung auf der nächsten Seite

Tabelle 6.4.: Die Usability-Benchmark-Zielvorgaben für die MANGO-App zu den jeweiligen Kernaufgaben bzw. die formulierten Gesamt-Ziele. Die Kernaufgabe 5 gilt als teilweise erfüllt, wenn sich zwei der drei genannten Lebensmittel, unabhängig der Reihenfolge, unter den Top 3 Gemüsesorten wiederfinden.

KA 5	Ausfindig machen der drei wichtigsten Gemüsesorten im eigenen FE-Typ.
Ziel 5a	50 % der User werden die KA 5 innerhalb von 1 Minute erfolgreich abschließen können.
Ziel 5b	75 % der User werden die KA 5 innerhalb von 2 Minuten <i>teilweise richtig</i> abschließen können.
<i>Gesamt-Ziele</i>	
Ziel 6	Die User werden zumindest 80 % der KA erfolgreich abschließen können.
Ziel 7	Die User werden zumindest 55 % der Ziele 1a, 1b, 2a, 2b, 3a, 3b, 4a, 5b entsprechend den darin definierten Schwellenwerten für sich erfolgreich abschließen können.
Ziel 8	Die User werden die App mit einen durchschnittlichen SUS-Score von mindestens 80 % bewerten.

Tabelle 6.4.: Usability-Benchmark-Zielvorgaben – Fortsetzung

In späteren Versionen der App können neue Ziele entsprechend den neuen Funktionen oder veränderten Ansprüchen von JR ergänzt und die Usability-Benchmark-Ziele aus diesem ersten Usability-Test mit anderen Anteilen bzw. Zeitvorgaben überprüft werden oder mit denselben, sollten ein oder mehrere Ziele in dieser Überprüfung nicht erreicht werden. Die mögliche Durchführung von Änderungen kann künftig auch besser abgewogen werden, sollten diese nicht die Funktionalität einer Kernaufgabe betreffen bzw. verbessern, ist fraglich, ob der Aufwand lohnend ist. Durch das Bereitstellen einer Zusammenfassung mehrerer Usability-Metriken kann auch die Gesamt-Datenlage späterer Versionen mit dieser Testung verglichen werden.

6.5.2. Eye-Tracking-Parameter

Die Aufgabe 2 des User-Tests (siehe Kapitel 6.3) wurde nach dem Konzept der sieben Stages of Action (siehe Kapitel 5.2) unterteilt und die dadurch bestimmten Aktivitäten wurden mittels Eye Tracking untersucht. Die Aufgabe 2 wurde für

die Analyse der Blickbewegungsdaten ausgewählt, da diese Aufgabe von allen Testpersonen einheitlich gelöst werden muss, nämlich durch die Verwendung der automatischen Lebensmittelerkennung, die Ergebnisse mit Hilfe einer mobilen Augmented-Reality-Funktion darstellt. Außerdem sind gerade die Ergebnisse zu dieser Funktion für JR als Entwickler von besonders großem Interesse, da in folgenden Versionen der App aufgrund der in dieser Diplomarbeit gewonnenen Ergebnisse auf das Verhalten der Testpersonen eingegangen werden kann, weil dies im Rahmen dieser Diplomarbeit analysiert wurde.

Von den Eye-Tracking-Daten werden die Anzahl der Fixationen und die durchschnittliche Blickdauer betrachtet, und bringen Aufschluss über die Verständlichkeit des Interfaces und Suchstrategien der Testpersonen (Goldberg et al., 1999; Poole et al., 2006). Diese können mit Hilfe der *Smartphone Eye Tracking Toolbox* (SMET) von JR (Paletta et al., 2014) als Heat Maps, farbcodierte Visualisierungen der Fixations-Daten, die das Smartphone-Display direkt überlagern, dargestellt werden. SMET extrahiert, aufgrund der am Smartphone angebrachten blauen Umrandung, dynamisch, aus dem Scan-Path-Video der ETG, den Bereich den das Smartphone einnimmt, um dann die Anzahl und Dauer der Fixationen farbcodiert und dynamisch in einem Video am Smartphone-Display zu visualisieren. Ein Beispiel einer Erkennung ist in Abbildung 6.6 dargestellt. Diese Farbcodierung reicht von blau (geringe Anzahl von Fixationen) über grün und gelb bis hin zu rot (hohe Anzahl von Fixationen), wenn ein Bereich nicht codiert ist wurde dieser von den Testpersonen nicht fixiert oder ist der Bereich aufgrund des dynamischen Videos bereits nicht mehr als fixiert gekennzeichnet. Die Testpersonen werden aufgrund ihres Benutzungsverhalten, also wie etwas gelöst wurde, in Gruppen unterteilt, um die Heat-Maps nach typischen Verhaltensweisen von den Testpersonen zu klassifizieren.

Die erhaltenen Eye-Tracking-Daten können, wie beim Usability Benchmark, als Baseline angesehen werden, um in späteren Versionen das bessere der Designs bestimmen zu können (Bojko, 2006). Anerkannte Thesen legen nahe, dass, abhängig vom Kontext, mehr Fixationen auf eine weniger effiziente Such-Strategie und längere Fixationen auf Schwierigkeiten mit der Wahrnehmung des Displays hinweisen (Goldberg et al., 1999; Bojko, 2006; Olmsted-Hawala et al., 2014; Poole et al., 2006). Da von den Testpersonen mehrere Male hintereinander das Scannen von Lebensmitteln durchgeführt wird, wird angenommen, dass sich die Werte der Fixationen mit

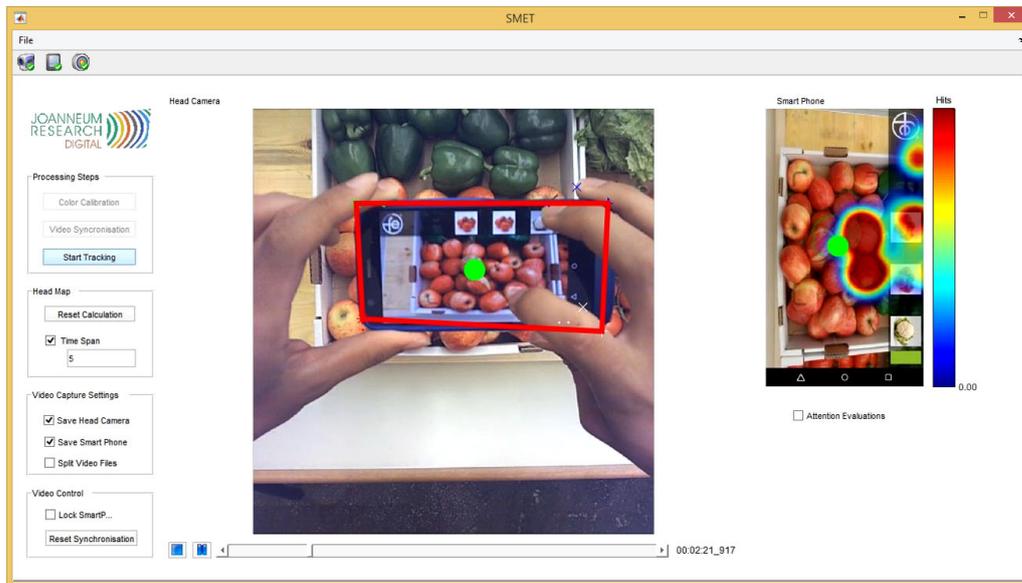


Abbildung 6.6.: SMET Interface mit laufender Erkennung für Daten von T04. Erkennt SMET das Smartphone im Video, wird dies durch einen roten Rahmen gekennzeichnet und die errechneten Blickdaten als farbcodierte Überlagerung des Screenrecordings dargestellt.

andauern der Versuche verbessern.

7. Empirische Studiendurchführung

Die Durchführung der empirischen Untersuchung zur Bewertung der Usability der App MANGO und die Analyse der Ergebnisse sind die zentralen Elemente dieser Diplomarbeit. In diesem Kapitel werden das allgemeine Untersuchungskonzept (Kapitel 7.1), der genaue Studienablauf (Kapitel 7.2) und der durchgeführte Pre-Test (Kapitel 7.3) präsentiert.

7.1. Untersuchungskonzept

Nachfolgend werden alle Facetten des zugrunde liegenden Studien-Konzepts präsentiert. Dieses umfasst das allgemeine Studiendesign und Informationen zur Rekrutierung der teilnehmenden Personen, sowie die Beschreibung der verwendeten Hardware, der Räumlichkeit der Studiendurchführung und des speziellen Versuchsaufbaus, und schließt mit einer Abgrenzung zu nicht gemachtem.

7.1.1. Studiendesign

Die Aufgabe der Testpersonen war es, sich im Rahmen der empirischen Studie mit der Smartphone-App MANGO, die bereits in Kapitel 6.1 vorgestellt wurde, auseinanderzusetzen und deren Usability zu beurteilen. Die Hauptaufgabe der Testpersonen war es, die Oberfläche der App zu explorieren und den Funktionsumfang der App anhand der konzipierten Aufgaben (siehe Kapitel 6.3) kennen zu lernen. Während der Benutzung wurden die Augenbewegungen mit Hilfe von Eye-Tracking-Glasses (ETG, Kapitel 7.1.3) aufgenommen, um anschließend Aussagen über das Benutzungsverhalten treffen zu können. Die Analyse der Szenen-Videos der ETG erfolgte rein

quantitativ anhand der in Kapitel 6.5 vorgestellten Parameter, sowie mit Hilfe der in Kapitel 3.1 beschriebenen Metriken, welche auch für die Auswertung der benutzten Fragebögen (siehe Kapitel 6.2 und 6.4) angewandt wurden. Die Zusammenfassung der Ergebnisse des Usability-Tests stellt in weiterer Folge den Usability Benchmark für die App (in der Version 0.8.5) dar und die Aktivitäts-Analyse gibt Aufschluss über das User-Verhalten beim Verwenden des Kameramodus.

Die Untersuchung war als Einzelstudie konzipiert, die eine einmalige (nicht wiederholende) Durchführung aufwies und stellt die Basis für folgende Untersuchungen der MANGO-App dar, welche eine Entwicklung der UX und Usability erfassen können. Der Zeitraum, in dem die User-Tests durchgeführt wurden, betrug drei Tage (plus ein Tag für den Pre-Test, siehe Kapitel 7.3), wobei jede Testung, inklusive der Bearbeitung der Fragebögen, etwa 50 Minuten in Anspruch nahm. Die Tests mussten in diesem komprimierten Zeitraum stattfinden, da für die realitätsnahe Verwendung der App der Aufbau von frischem Obst und Gemüse (wie im Supermarkt) notwendig war (siehe Kapitel 7.1.5). Bezüglich der zeitlichen Dimension sei noch angemerkt, dass für die Bearbeitung der Aufgaben keine Maximaldauer für die Testpersonen vorgegeben war und dies auch deutlich den Testpersonen mitgeteilt wurde (siehe Kapitel 6.3).

7.1.2. Testpersonen

Die sechzehn Testpersonen für die Studie wurden auf unterschiedliche Weise rekrutiert, wobei darauf geachtet wurde, dass acht weibliche und acht männliche Personen tatsächlich an der Studie teilnehmen. Die Möglichkeit zur Studienteilnahme wurde an der Technischen Universität Graz und der Universität Graz für Personen von 18 bis 35 Jahren, die keine Sehhilfe benötigen und Interesse an einer Eye-Tracking-Studie zu einer Ernährungsapp haben, unter Angabe des Studienzeitraums, ausgeschrieben. Des weiteren wurden einige Personen eines Testpersonen-Pools von JR für die Studienteilnahme telefonisch kontaktiert. In beiden Fällen wurden Termine im eineinhalb Stunden-Takt mit den Testpersonen an den drei Versuchstagen vereinbart. Für die Testpersonen wurde im Vorfeld eine Probandenversicherung über JR abgeschlossen und sie erhielten für ihren Aufwand eine Entschädigung von 30 Euro.

7.1.3. Hardwarebeschreibung

Im Rahmen der Studie wurde die App MANGO in der Version 0.8.5 auf einem Motorola Moto G (1. Generation) mit dem Betriebssystem Android 5.0 installiert, das über einen Quad-Core 1,2 GHz Cortex-A7 Prozessor, ein 4,5 Zoll HD-Display mit einer Auflösung von 720×1280 Pixel und 8 GB internen Speicher sowie 1 GB RAM verfügt und mit WLAN 802.11 b/g/n und Bluetooth 4.0 ausgestattet ist. Für die Heat-Map-Erstellung mit SMET (siehe Kapitel 6.5.2) war es notwendig, dass das Smartphone mit einem farbigen Rand um das Display auszustatten. Da es weder in einem Geschäft, noch online möglich war, ein vom Modell her passendes Handycover zu bekommen, wurde das Moto G mit einem blauen Isolierband entlang der Kanten versehen.

Für das Eye Tracking kamen die SMI™ Eye Tracking Glasses¹ (ETG), Version 1.0 (siehe Abbildung 7.1), aus dem Bestand vom Human-Factors-Labor von JR zum Einsatz. Dieser nicht-invasive, Video-basierende, binokulare Eye Tracker mit automatischer Parallaxenkompensation ermöglicht eine Messung der Blickbewegungen für beide Augen mit 30 Hz. Von den ETG wird ein HD-Szenen-Video (1280×960 Pixel, 24fps) mit Gaze-Annotation und einer Positionierungsgenauigkeit $0,5^\circ - 1,0^\circ$ in einem Blickfeld von 70° horizontal und 55° vertikal aufgenommen. Die ETG in dieser Version müssen während ihres Einsatzes mit dem Aufnahmelaptop über ein USB-Kabel verbunden bleiben.



Abbildung 7.1.: SMI™ Eye-Tracking-Glasses, Version 1.0, vgl. Santner et al. (2013)

¹www.eyetracking-glasses.com, zuletzt besucht am 21.8.2015

Weitere Aufzeichnungen

Zu Dokumentationszwecken wurden während der Tests zusätzlich Videoaufnahmen durchgeführt. Diese wurden mit einer handelsüblichen Kompaktkamera mit einer Auflösung von 1920×1080 Pixel aufgenommen. Platziert wurde die Kamera seitlich des Versuchsaufbaus auf einem Stativ (gut zu erkennen in Abbildung 7.5), welches während der gesamten Studie nicht verstellt wurde. Die Videoaufzeichnung beinhaltet natürlich auch eine Tonspur und beides kann zur Validierung von unklaren Szenen dienen und dokumentiert aufgrund der Platzierung auch gut, in welchem Abstand zu den Lebensmitteln die Testpersonen die App einsetzen.

Darüber hinaus wurde der Bildschirm des verwendeten Smartphones (siehe Kapitel 7.1.3) mittels der Lollipop Screen Recorder App (frei erhältlich im Google Play Store²) als Video mit einer Auflösung von 1280×720 Pixel aufgenommen. Dieses Video wird für das SMET-Auswertungstool benötigt und kann ebenfalls zur Validierung in Zweifelsfällen herangezogen werden.

Zwei weitere Aufzeichnungen wurden für aktuelle Forschungsthemen von JR geführt. Auf der rechten Seite der ETG ist am Bügel ein Akzelerometer und Gyroskop, oder Motion Tracker, befestigt, mit dem die Kopfneigung und Orientierung über eine andere App gespeichert werden kann. Diese Daten können in einem 3D-Modell des Versuchsaufbaus (vgl. Santner et al., 2013) visualisiert werden. Physiologische Messungen mit Hautleitwiderstand werden durch den Affectiva-Q-Sensor aufgenommen, der in ein Armband integriert ist und ohne die Testpersonen einzuschränken verwendet werden kann. Diese beiden Messgeräte sind im praktischen Einsatz in Abbildung 7.2 zu sehen. In der vorliegenden Studie wurden die Daten dieser Aufzeichnungen nicht weiter verarbeitet.

7.1.4. Räumlichkeit

Die Studie konnte im Human-Factors-Labor durchgeführt werden, welches freundlicherweise von JR zur Verfügung gestellt wurde. Dieses besteht aus einem kleinen

²<http://play.google.com>, zuletzt besucht am 21.8.2015



Abbildung 7.2.: Motion Tracker am Bügel der ETG und Affectivia-Armband am Handgelenk für ergänzende Untersuchungen durch JR.

Vorraum und einem großen Hauptraum mit drei Fenstern, die mit schwarzer Folie verklebt sind, um eine gleichmäßige Beleuchtung zu gewährleisten. Zusätzlich zu dem Arbeitstisch für das gesamte Equipment und das Ausfüllen der Fragebögen (siehe Abbildung 7.3), wurde ein weiterer Tisch für den Versuchsaufbau (siehe Kapitel 7.1.5) benötigt. An der Wand ist eine Ebene mit Markern zur Kalibrierung (siehe Abbildung 7.4) der ETG angebracht, zu der verschiedene Abstände, je nach Untersuchungsgegenstand, gewählt werden können.

Am Arbeitstisch erfolgte sowohl das Ausfüllen der Fragebögen vor und nach dem Benützen der App, als auch die Erklärung und das Anpassen der Hardware (siehe Kapitel 7.1.3). Die Versuchsleiterin verweilte während dem Ausfüllen der Fragebögen am Arbeitstisch, ihr Assistent blieb während des gesamten Experiments ebenfalls dort und ermittelte sofort nach dem Ausfüllen der Vorbefragung, während der Hardware-Einweisung und Kalibrierung durch die Versuchsleiterin, den FE-Typ der jeweiligen Testperson. Die Kalibrierung und Validierung der ETG erfolgt neben dem Arbeitstisch mit Hilfe der dafür vorgesehenen Ebene mit Markern an der Wand (siehe Abbildung 7.4) in einem Abstand von einem Meter. Die Testpersonen wurden gebeten ungeschminkt, im Speziellen ohne Wimperntusche, zur Untersuchung



Abbildung 7.3.: Übersicht der Einrichtung des Human Factors Labor während der Studie

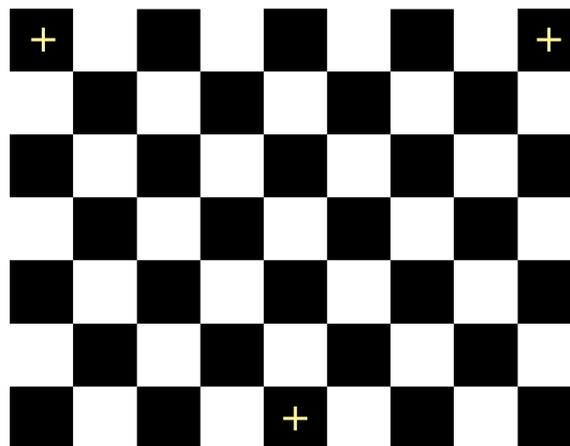


Abbildung 7.4.: Schachbrettmuster mit Markern zur Kalibrierung der ETG

zu kommen. Sollte es Probleme beim Kalibrieren wegen getragener Mascara geben, steht beim Arbeitstisch alles zum Abschminken sowie frische Mascara bereit. Außerdem wurden auch Linsenflüssigkeit und ein Linsenbehälter zur Verfügung gestellt, sollte es mit den getragenen Linsen der Testpersonen Probleme geben.

Beim Versuchsaufbau befand sich neben der Testperson auch die Versuchsleiterin, da unmittelbar vor dem praktischen User-Test in der MANGO-App der jeweilige FE-Typ (laut vorheriger Erhebung und Auswertung) hinterlegt und die Aufgaben der Studie (siehe Kapitel 6.3) vorgelesen wurden. Die Versuchsleiterin verweilte dabei in ihren Pausen still und hielt, wie in Abbildung 7.5 zu sehen, einen angemessenen Abstand zur Testperson.



Abbildung 7.5.: Typische Situation während dem praktischen User-Test

7.1.5. Versuchsaufbau

Um die App möglich realitätsnah zu testen, war es notwendig verschiedene Lebensmittel wie in einem Supermarkt aufzubauen. Die MANGO-App kann zwölf verschiedene Lebensmittel (siehe Abbildung 6.1) erkennen. Zum Zeitpunkt der Durchführung der Studie stand eines der zwölf Lebensmittel, Kohl, saisonal bedingt nicht zur Verfügung. Neben diesem Aspekt musste bei der Auswahl der Lebensmittel berücksichtigt

7. Empirische Studiendurchführung

werden, ob ein Lebensmittel zu einem oder mehreren der FE-Typen zugehörig ist, oder nicht. Die zur Verfügung stehenden Lebensmittel und deren Zugehörigkeit zu den einzelnen FE-Typen sind in Tabelle 7.1 zusammengefasst und die ausgewählten Lebensmittel gekennzeichnet.

	MANGO Lebensmittel	in Studie	Beauty Food	Brain Food	Power Food	Soul Food
1	Äpfel rot-gelb	⊗		×		×
2	Äpfel gelb			×		×
3	Fleischtomaten		×		×	×
4	Rispentomaten	⊗	×		×	×
5	Salat (Kopfsalat)	⊗	×		×	×
6	Karfiol	⊗		×		
7	Paprika gelb	⊗				×
8	Paprika rot		×			×
9	Paprika grün	⊗	×			×
10	Weintrauben grün		×	×		×
11	Bananen					
12	(Kohl)		(×)	(×)	(×)	(×)
	Anzahl/FE-Typ in Studie		3	2	2	5

Tabelle 7.1.: Auswahl der sechs Lebensmittel der Studie aus den elf zur Verfügung stehenden Lebensmitteln, sowie deren Zugehörigkeit zu den einzelnen FE-Typen. Kohl stand saisonal bedingt nicht zur Verfügung, wäre jedoch aufgrund seiner FE-Typenzugehörigkeiten bestens geeignet gewesen.

Zur Präsentation der Lebensmittel, ähnlich wie in einem Supermarkt, wurde ein kleines Regal mit schräger Oberfläche gebaut, das auf einem Tisch platziert wurde. Der fertige Versuchsaufbau ist in Abbildung 7.6 dargestellt.

7.1.6. Studienziel

Die durchgeführte Studie diente nicht primär der Identifikation und Interpretation von Usability-Problemen. Das Hauptziel der Studie war die quantitative Beurtei-



Abbildung 7.6.: Fertiger Versuchsaufbau zur Präsentation der Lebensmittel

lung der Usability der App MANGO in der Version 0.8.5, die eine Grundlage, einen Benchmark, für die Testung von folgenden Versionen darstellt. Die Ergebnisse der Aktivitäts-Analyse bezüglich des Kameramodus liefern die Basis, um die UX bei dieser Kernfunktionalität in späteren Versionen zu verbessern, da besser auf das Verhalten der Benutzerinnen und Benutzer eingegangen werden kann, wenn dieses bekannt ist.

7.2. Studienablauf

Die Testpersonen wurden direkt im Hauptraum des Human-Factors-Labors von JR durch die Versuchsleiterin empfangen und begrüßt und ihr Assistent wurde vorgestellt. Um eine lockere Atmosphäre zu schaffen, stellte man sich informell mit Vornamen vor und einleitend wurde ein kurzer Smalltalk geführt. Nach der Klärung eventueller Vorab-Fragen wurde die Testperson gebeten am Arbeitstisch Platz zu nehmen und darauf wurden die Studie bzw. der User-Test inhaltlich sowie vom Ablauf her kurz beschrieben. Im Anschluss wurde die Zustimmungserklärung (siehe Anhang D), mit der Aufforderung die Doppelseite durchzulesen und mit der Bitte

diese bei Einverständnis zu unterschreiben, ausgehändigt. Die Zustimmungserklärungen wurden gesondert und in zufälliger Reihenfolge archiviert, um die Anonymität der Testpersonen zu gewährleisten. Nach dem Unterschreiben wurde noch einmal für die Bereitschaft an der Studie teilzunehmen gedankt und der Fragebogen „Vorbefragung“ (siehe Kapitel 6.2) mit dem Hinweis, dass alles Geforderte genau beschrieben sei und man für Fragen jederzeit bereit stünde, der Testperson zum Ausfüllen gegeben.

Der Eye-Tracking-Teil startete mit dem Erklären der Hardware und dem Anlegen des Affectivia-Armbandes, dem Einschalten des Motion Trackers sowie dem Anpassen der ETG. Die ETG wurden mit dem integrierten Zugband von den Testpersonen selbst so fest angezogen, dass diese gut saßen und das Tragen angenehm war. Die ETG waren während dem gesamten Experiment durch ein Kabel mit dem Versuchs-Laptop verbunden. Das Bild der Augenkameras der ETG wurde überprüft und optimiert, danach wurde die Testperson ausdrücklich darauf hingewiesen, die ETG bzw. den Kopf nicht mehr zu berühren. Währenddessen wurde die FE-Typenzuordnung vom Assistenten vor Ort ausgewertet und das Ergebnis auf der Aufgabenbeschreibung vermerkt. Zur Kalibrierung nahmen die Testperson in einem Meter Abstand zur dafür vorgesehenen Ebene mit Markern an der Wand (siehe Abbildung 7.4) Platz, der Stuhl wurde zuvor entsprechend der Markierungen am Boden umgestellt. Nach dem Kalibrieren wurde die Aufnahme mit den ETG gestartet und zur Synchronisierung der Geräte wurden der Aufnahmestart der Screen-Recording-App und der App für den Motion Tracker durch die ETG gefilmt, und das Affectivia-Armband durch zweimaliges Klatschen aktiviert. Danach wurde den Testpersonen beschrieben welcher Bildausschnitt ihres Blickfeldes von den ETG aufgenommen und wie das Smartphone am besten in der Hand gehalten wird, damit dies später automatisch (durch die blaue Umrandung) getrackt werden kann. Bevor der Laptop, zu dem die ETG während der Untersuchung verbunden bleiben, in einem Rucksack angelegt wurde, erfolgte eine Validierung der Kalibrierung, in der die Testpersonen selbst sagten, auf welchen Marker sie blickten.

Fertig ausgerüstet erfolgte für die Testpersonen nun der praktische Teil der Untersuchung beim Lebensmittel-Versuchsaufbau (siehe Kapitel 7.1.5). Die Videoaufzeichnung wurde aktiviert, in der App der FE-Typ der Testperson hinterlegt und es wurden die Instruktionen von der Versuchsleiterin vorgelesen, die um den Hin-

weis, dass es keine Zeitvorgabe zu den Aufgaben gibt, ergänzt wurden. Während der Durchführung der Aufgaben kam es vor, dass die Testpersonen hingewiesen wurden das Smartphone im Aufnahmefeld der ETG zu halten, also den Kopf entsprechend mitzubewegen. Nach Beendigung der Aufgaben wurden die Testpersonen darüber informiert, dass der praktische Teil bereits fast vorbei ist, aber eine zweite Validierung anschließt, also die ETG noch nicht berührt werden dürfen. Zugleich wurde das Smartphone entgegengenommen und es wurden die Videokamera und die Aufnahmen durch das Smartphone gestoppt. Nach dem Ablegen des Laptops wurde neben dem Arbeitstisch wieder eine Validierung der ETG durchgeführt, die danach abgenommen werden konnten. Abschließend wurde von den Testpersonen der Fragebogen zur App-Bewertung ausgefüllt.

Vor der Verabschiedung erhielten die Testpersonen ihre Aufwandsentschädigung, deren Erhalt bestätigt werden musste, und den Testpersonen wurde angeboten in den Testpersonen-Pool von JR aufgenommen zu werden, was alle Testpersonen annahmen. Nach der Verabschiedung wurden alle Daten gesichert, alle Geräte wieder aufgeladen, alle aufgezeichneten Daten von Smartphone gelöscht und alle Papierunterlagen beschriftet.

All diese Schritte wurden in einer Checkliste (siehe Anhang E) festgehalten und, um die Qualität des Experiments zu gewährleisten, wurde diese Checkliste für jede Testperson ausgefüllt. Alle Unterlagen (bis auf die Zustimmungserklärung) und Dateien wurden mit „T##“ (## steht für die fortlaufende Nummer) beschriftet bzw. gespeichert. Für die Testpersonen wurde im Vorfeld eine Probandenversicherung durch JR abgeschlossen.

7.3. Pre-Test

Um Unstimmigkeiten bzw. Schwierigkeiten beim Ablauf der Studie zu identifizieren und die Abstimmung der Hardware zu überprüfen, wurde vorab ein sogenannter Pre-Test durchgeführt. Da für diesen bereits der gesamte Aufbau der verderblichen Lebensmittel (siehe Kapitel 7.1.5) wie in der eigentlichen Studie vorhanden sein musste, fand unter dem Aspekt der Ressourcenschonung der Pre-Test zeitnah vor der Durchführung der Studie statt.

Schon beim Bereitstellen der Hardware wurde festgestellt, dass der Befestigungsmechanismus des Motion-Trackers (siehe Kapitel 7.1.3) zwischenzeitlich verändert wurde, was aber leicht zu beheben war. Der Pre-Test wurde ohne Motion-Tracker durchgeführt.

Zum Pre-Test wurden zwei Testpersonen aus dem Bekanntenkreis der Versuchsleiterin eingeladen, welche, wie die übrigen Testpersonen, versichert wurden und ebenfalls eine Aufwandsentschädigung erhielten. Der gesamte Studienablauf (siehe Kapitel 7.2) wurde mit Hilfe einer Vorab-Version der eigentlichen Checkliste (siehe Anhang E) durchgeführt.

Aufgrund des Vortests wurde eine Excel-Vorlage zur Erfassung der FE-Typen eingeführt, um den Vorgang der Ermittlung zu optimieren und gleichzeitig zu dokumentieren. Weiters konnten kleine Fehler in der Aufgabenbeschreibung ermittelt und ausgebessert werden.

Nach dem Pre-Test wurde die Checkliste um den genauen Ablauf der Synchronisation der Geräte ergänzt, da dieser erst vor dem Pre-Test festgelegt und besprochen wurde, außerdem wurden die erhobenen Daten gesichtet. Eine Analyse mit dem SMET-Tool konnte nach dem Pre-Test noch nicht durchgeführt werden, da dieses erst für die neue Anwendung in dieser Studie angepasst werden musste.

8. Ergebnisse

In diesem Kapitel werden die gesamten Ergebnisse präsentiert. Zu Beginn werden die Testpersonen näher durch ihre Angaben bei der Vorbefragung beschrieben und die übrigen Ergebnisse der Vorbefragung präsentiert.

Die Testpersonen ($n = 16$) waren zwischen 20 und 30 Jahre alt ($M = 24,88$; $SD = 2,85$) und schätzten ihre Erfahrung mit Informationstechnologie bzw. mit Smartphones und Tablets auf einer 8-stufigen Skala (1 – sehr gering bis 8 – sehr groß) jeweils zwischen 3 und 8 ein ($M_{IT} = 5,38$; $SD_{IT} = 1,36$ bzw. $M_{ST} = 6,0$; $SD_{ST} = 1,75$). 15 der 16 Personen kommen aus einem universitären Umfeld.

Die Testpersonen leben in Haushaltsgrößen zwischen ein und vier Personen ($M = 2,06$; $SD = 0,85$) und bis auf zwei Personen erledigen alle selbst hauptsächlich die Einkäufe in ihrem Haushalt. Drei der Testpersonen gaben an, nicht auf eine abwechslungsreiche, ausgewogene Ernährung zu achten, wohingegen zwölf auf eine solche achten. Ein Drittel der Testpersonen ist nach der FE Typenzuordnung dem Typ Power Food zuzurechnen, die gesamte Verteilung auf die vier FE-Typen sowie zwei Mischtypen ist in Abbildung 8.1 dargestellt. Neun Personen (sechs Männer und drei Frauen) schätzten ihren Lebensstil ohne zusätzliche Informationen zu FE zu haben in der Vorbefragung (Kapitel 6.2) richtig ein.

Die Vorbefragung beinhaltete drei weitere Fragen, die Bezug auf die zu testende App nahmen. Der erste Punkt setzt sich aus Ergebnissen der Vor- und Nachbefragung zusammen. Vor und nach der Testung wurden die Testpersonen gebeten auf einer Skala von 0 bis 5 (gar nicht wahrscheinlich bis äußerst wahrscheinlich) einzuschätzen, wie wahrscheinlich sie die (zuvor beschriebene bzw. danach getestete) App herunterladen und verwenden würden. Vor der Testung konnten höhere Werte ($M = 2,88$; $SD = 1,5$) festgestellt werden als nach der Testung ($M = 2,63$; $SD = 1,7$). Das Gesamtergebnis wird in Abbildung 8.2a präsentiert. Wie in Tullis et al. (2013, S. 27f.)

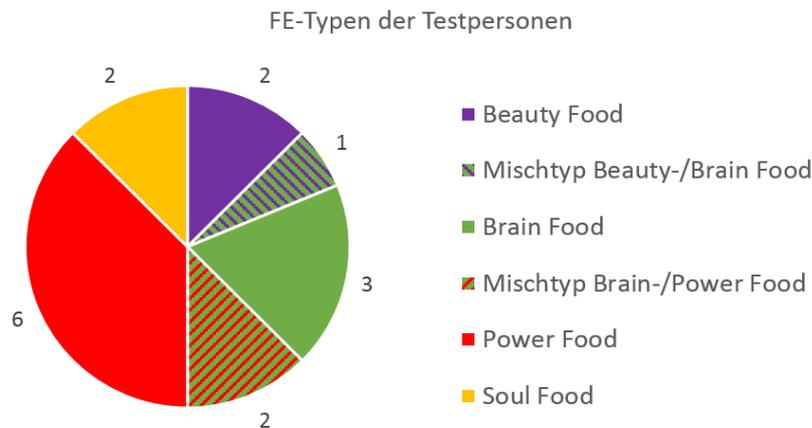
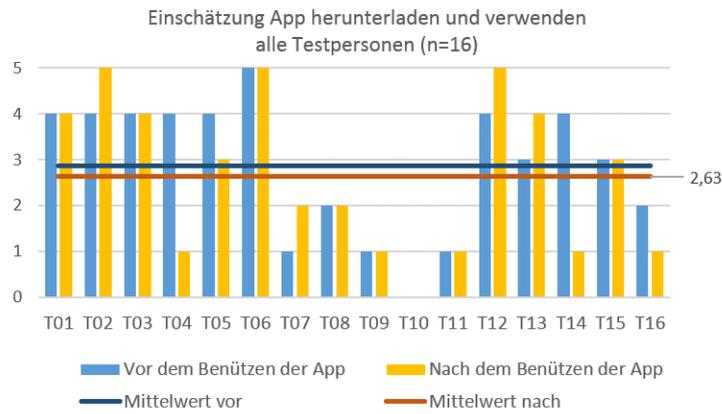


Abbildung 8.1.: Verteilung der Testpersonen auf die vier FE-Typen und zwei Mischtypen

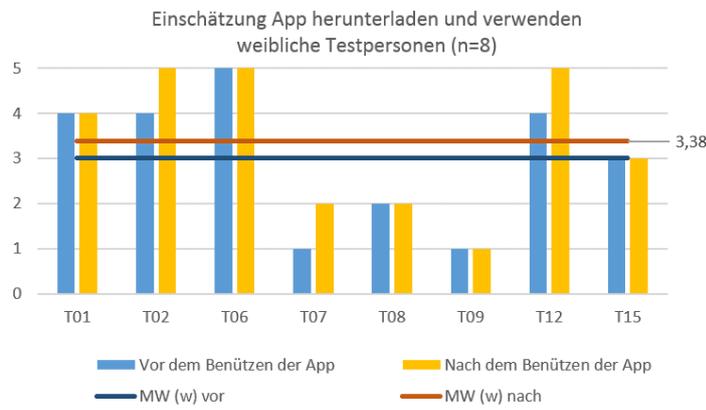
beschrieben, wurde ein T-Test für gepaarte Differenzen und einem 90 % Konfidenzintervall durchgeführt. Der Unterschied konnte nicht als signifikant nachgewiesen werden ($T(15) = 0,808$; $p = 0,432$; 90 % Konfidenzintervall $[-0,74; 0,24]$).

Die Bewertung der Fragen wurde daraufhin nach den Geschlechtern getrennt betrachtet. Bei der weiblichen Gruppe war die Bewertung vor der Testung niedriger ($M = 3$; $SD = 1,5$) als nach der Testung ($M = 3,38$; $SD = 1,6$), dieser Unterschied konnte hingegen als signifikant nachgewiesen werden ($T(7) = 2,049$; $p = 0,08$; 90 % Konfidenzintervall $[0,07; 0,68]$). Bei der Gruppe der Männer war es genau umgekehrt, hier war die Bewertung vor der Testung höher ($M = 2,75$; $SD = 1,58$) als nach der Testung ($M = 1,88$; $SD = 1,55$), dieser Unterschied konnte ebenfalls als signifikant nachgewiesen werden ($T(7) = -1,698$; $p = 0,133$; 90 % Konfidenzintervall $[-1,72; -0,03]$). Die Detailergebnisse nach Geschlechtern sind in Abbildung 8.2b und 8.2c dargestellt.

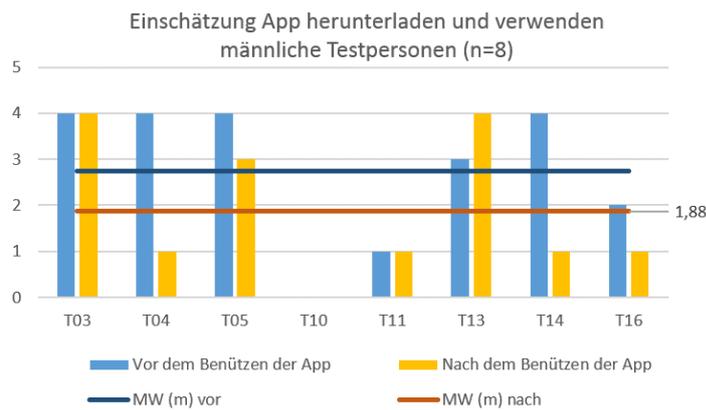
Bei der Frage, warum die Testpersonen die App herunterladen würden, gab es acht Antwortmöglichkeiten (Mehrfachantworten möglich) und ein Feld für freie Antworten. Elf Personen würden die App aus Neugierde herunterladen, sechs, weil sie sich gesünder ernähren wollen, sieben, weil sie an gesunder Ernährung interessiert sind, zwei, weil sie abnehmen wollen, drei, weil der Zugang zu einer individuellen Ernährungsweise neu für sie ist, sechs, um mehr zu Herkunft/Regionalität der Produkte



(a) Ergebnis aller Testpersonen



(b) Ergebnis weiblicher Testpersonen



(c) Ergebnis männlicher Testpersonen

Abbildung 8.2.: Auswertung zur vor und nach der Testung gestellten Frage, wie wahrscheinlich die TP die App heruntergeladen und verwenden würden.

zu erfahren, und eine Person, weil sie immer am neuesten Stand der Technik ist. Zwei Personen würden die App gar nicht herunterladen, dazu findet sich mit „kein Smartphone“ eine Erklärung bei den freien Antworten. Die beiden weiteren freien Antworten waren „Erkennung von Lebensmittel per App ist interessantes Feature“ und „weil ich mein Wissen in den Ernährungswissenschaften bestätigt haben möchte“.

Als nächstes werden in Kapitel 8.1 alle Ergebnisse des Usability Benchmarks präsentiert, darauf in Kapitel 8.2 die Ergebnisse der Post-Test-Fragebögen zusammengefasst und abschließend in Kapitel 8.3 die Ergebnisse der Aktionsanalyse nach den Sieben Stages of Action mit Eye Tracking dargestellt.

8.1. Usability Benchmark

Als wesentliche Usability-Metriken wurden für die Beurteilung der MANGO-App und für die Benchmark-Bildung die Time on Task, der Task Success, die subjektive Bewertung der Testpersonen und die Fehlerfreiheit des Systems in verschiedenen Variationen berechnet. Alle Daten wurden basierend auf den Beobachtungen aus den Videos der ETG erfasst, eines der 16 Ergebnisse musste ausgeschieden werden, da sich das Smartphone laufend zu mehr als 50 % außerhalb des Blickfeldes der ETG befand.

Für die Usability-Benchmark-Zielvorgaben ergeben sich die Ergebnisse wie in Tabelle 8.1 dargestellt. Die Kernaufgaben konnten bis auf KA 5 relativ gut von den Testpersonen bewältigt werden, die dazu definierten Ziele wurden jedoch zu 77 % nicht erfüllt. Auf die Tabelle 8.1 markierten Ziele, jene Ziele, die nicht erreicht werden konnten, wird in Kapitel 9 näher eingegangen.

Beschreibung	Ergebnis
KA 1: <i>Auffinden der eigenen FE-Typenbeschreibung.</i>	100 %
Ziel 1a: 90 % der User werden bei der KA 1 ohne Umwege direkt den Menüpunkt functional eating ansteuern.	73 %★
Ziel 1b: 75 % der User werden die KA 1 innerhalb von 15 Sekunden abschließen können.	53 %★
KA 2: <i>Öffnen des Kameramodus, um Lebensmittel zu scannen.</i>	100 %
Ziel 2a: 90 % der User werden bei der KA 2 ohne Umwege direkt den Kameramodus öffnen.	47 %★
Ziel 2b: 75 % der User werden die KA 2 innerhalb von 10 Sekunden abschließen können.	60 %★
KA 3: <i>Durchführen von 3 (oder mehr) Lebensmittel-Scans im Kameramodus.</i>	87 %
Ziel 3a: 75 % der User werden die ersten drei Lebensmittel-Erkennungen (egal ob richtig oder falsch) in der KA 3 innerhalb von 1 Minute abschließen können.	47 %★
Ziel 3b: Der Anteil der richtigen Scan-Ergebnisse pro User in der KA 3 wird im Mittel mindestens 70 % erreichen.	74 %
Ziel 3c: 55 % der User werden beim Scannen in der KA 3 ausschließlich richtige Ergebnisse erhalten.	33 %★
Ziel 3d: Der Kameramodus funktioniert in der KA 3 bei 90 % der User ohne abzustürzen.	40 %★
KA 4: <i>Ausfindig machen des am besten passenden Lebensmittels (einer Auswahl von Lebensmitteln vor Ort) im eigenen FE-Typ.</i>	67 %
Ziel 4a: 75 % der User werden die KA 4 innerhalb von 2 Minuten erfolgreich abschließen können.	53 %★

Fortsetzung auf der nächsten Seite

Tabelle 8.1.: Ergebnisse des Usability-Benchmark für die MANGO-App. Hervorgehobene Ergebnisse haben die Zielvorgabe nicht erreicht.

8. Ergebnisse

Beschreibung	Ergebnis
KA 5: <i>Ausfindig machen der drei wichtigsten Gemüsesorten im eigenen FE-Typ.</i>	13 %
Ziel 5a: 50 % der User werden die KA 5 innerhalb von 1 Minute erfolgreich abschließen können.	13 %★
Ziel 5b: 75 % der User werden die KA 5 innerhalb von 2 Minuten teilweise richtig abschließen können.	47 %★
<i>Gesamt-Ziele</i>	
Ziel 6: Die User werden zumindest 80 % der KA erfolgreich abschließen können.	73 %★
Ziel 7: Die User werden zumindest 55 % der Ziele 1a, 1b, 2a, 2b, 3a, 3b, 4a, 5b entsprechend den darin definierten Schwellenwerten für sich erfolgreich abschließen können.	55 %
Ziel 8: Die User werden die App mit einen durchschnittlichen SUS-Score von mindestens 80 % bewerten.	80 %

Tabelle 8.1.: Usability-Benchmark-Ergebnisse – Fortsetzung

Eine weitere Möglichkeit die Benchmark-Ziele zu betrachten ist es, den Erfüllungsgrad jedes Ziels wie in Abbildung 8.3 darzustellen. Für die User Performance lassen sich die Erfolgsquoten für die Erfüllung der Kernaufgaben sowie für das Erreichen der Ziele 1a, 1b, 2a, 2b, 3a, 3b, 4a, 5b entsprechend den darin definierten Schwellenwerten (vgl. Ziel 7) wie in Abbildung 8.4 zusammenfassen.

Die durchschnittliche Time on Task der Kernaufgaben ist in Abbildung 8.5 dargestellt. Eine andere Sicht bietet wieder die Performance der Testpersonen, in Abbildung 8.6 dargestellt als Anteil der Testpersonen, die für die jeweilige Kernaufgabe weniger als eine Minute benötigt haben.

Als Fehler wurden die Abstürze des Kameramodus in allen KA und falsche Lebensmittel-Scan-Ergebnisse in KA 3 dokumentiert. Zweiteres nur in KA 3, da nur in dieser Aufgabe die Lebensmittel-Scan-Funktion von allen Personen verwendet wurde. Die Abstürze des Kameramodus werden gestaffelt und nach der Anzahl pro Person auf die Kernaufgaben verteilt in Abbildung 8.7 dargestellt. In Abbildung 8.8 werden pro Person die Anzahl der durchgeführten Lebensmittel-Scans und der Anteil der

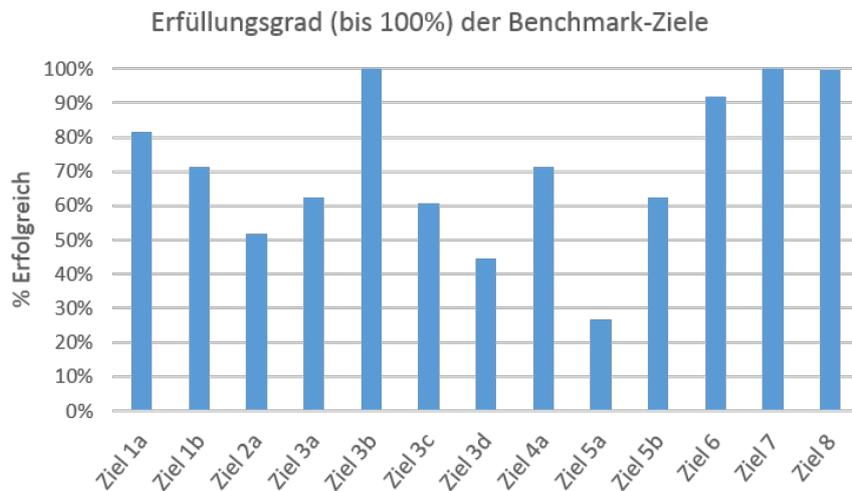


Abbildung 8.3.: Erfüllungsgrade der Usability-Benchmark-Zielvorgaben, dargestellt bis 100 %, berechnet aus dem Verhältnis der erreichten Ergebnisse zu den Zielvorgaben.

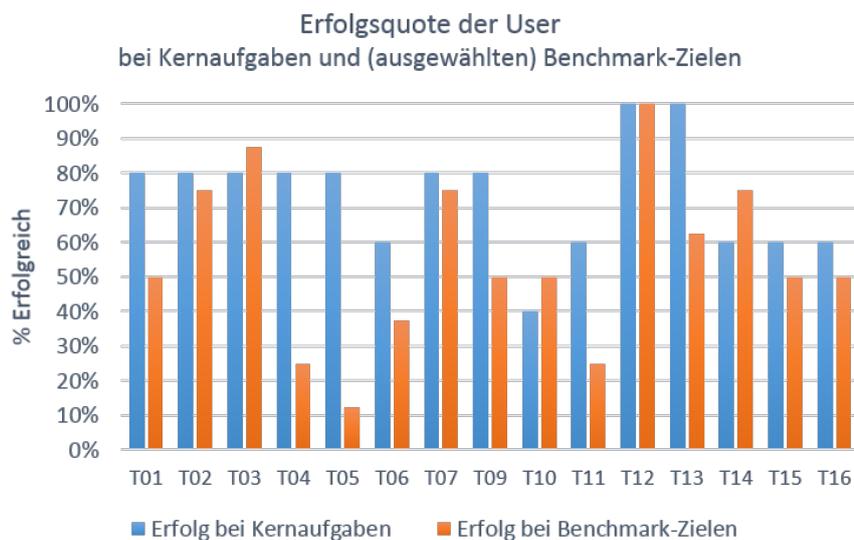


Abbildung 8.4.: User Performance bei den Kernaufgaben und den Benchmark-Zielen 1a, 1b, 2a, 2b, 3a, 3b, 4a, 5b bezüglich dem individuellen Erreichen der Ziel-Schwellenwerte, im Besonderen wird das Ziel 3b hier im Gegensatz zur Bewertung im Benchmark für jede Testperson als erreicht oder nicht erreicht bewertet, wie beim Benchmark-Ziel 7.

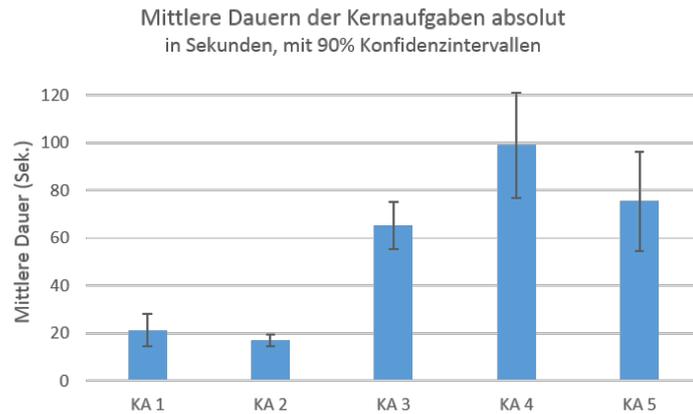


Abbildung 8.5.: Mittlere Dauern jeder Kernaufgabe, dargestellt mit 90 % Konfidenzintervallen.

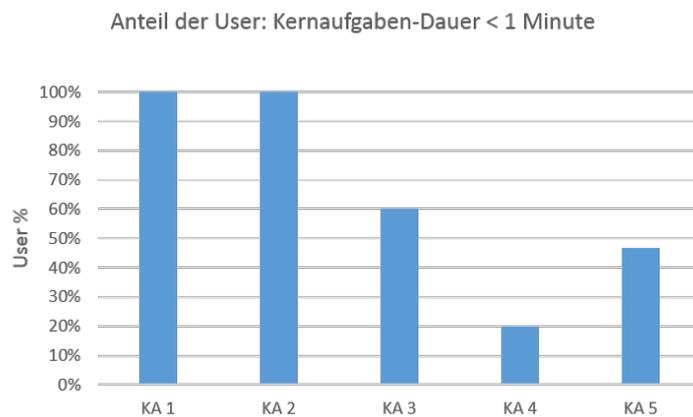


Abbildung 8.6.: Anteil der Testpersonen, die für die jeweilige Kernaufgabe unter einer Minute benötigt haben.

nicht richtig erkannten Lebensmittel aus KA 3 einander gegenüber gestellt. Insgesamt enthielten 31 % aller Scan-Ergebnisse nicht das richtige Lebensmittel.

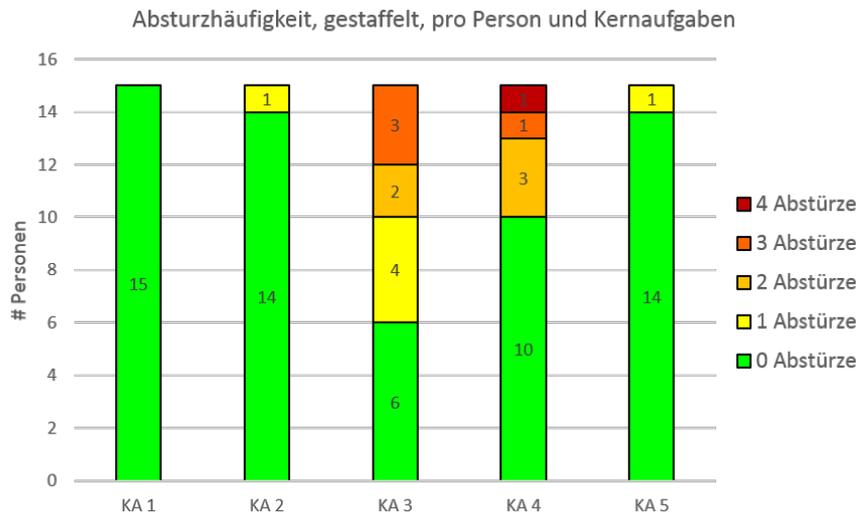


Abbildung 8.7.: Gestaffelte Abstürzhäufigkeit der App im Kameramodus zu den Kernaufgaben.

Eine Vergleich der Ergebnisse der subjektiven Bewertungen der Testpersonen durch den SUS und UEQ ist in Tabelle 8.2 dargestellt, bei dem alle Ergebnisse in Prozente transformiert wurden. Die UEQ-Ergebnisse (von -3 bis $+3$) wurden hierfür um $+3$ skaliert, um Bewertungen von 0 bis 6 zu erhalten, die in Prozentwerte durch Division durch 6 umgewandelt wurden. Mit den skalierten Bewertungen (von 0 bis 6) wurden keine Berechnungen durchgeführt.

Zusammenfassend ergibt sich für die Gesamt-Performance der Testpersonen in der Studie (siehe Tabelle 8.3) ein durchschnittlicher Wert von 66 % (mit 90 % Konfidenzintervall 59 bis 73 %). Das Einbeziehen des transformierten UEQ-Ergebnisses, wie in Tabelle 8.2, hat nur einen geringen Einfluss auf die durchschnittliche Performance (69 %, mit 90 % Konfidenzintervall 64 bis 74 %), verändert jedoch die Performance von einzelnen Testpersonen merklich (siehe Markierungen in Tabelle 8.3).

# Test- person	SUS	UEQ						Mittel- wert
		Attrak- tivi- tät	Durch- führ- bar- keit	Effek- tivi- tät	Steu- er- bar- keit	Sti- mula- tion	Ori- gina- lität	
T01	90%	83%	92%	83%	71%	71%	75%	81%
T02	83%	81%	58%	75%	42%	100%	100%	77%
T03	90%	75%	92%	67%	50%	58%	88%	74%
T04	68%	19%	83%	29%	33%	75%	92%	57%
T05	85%	69%	79%	75%	79%	83%	83%	79%
T06	90%	89%	63%	67%	79%	75%	67%	76%
T07	78%	72%	58%	63%	58%	58%	54%	63%
T09	50%	58%	42%	58%	50%	71%	54%	55%
T10	90%	69%	100%	83%	75%	71%	58%	78%
T11	90%	92%	100%	67%	63%	63%	96%	81%
T12	90%	86%	92%	83%	83%	92%	92%	88%
T13	88%	83%	88%	67%	71%	83%	83%	80%
T14	70%	44%	50%	71%	58%	29%	25%	50%
T15	68%	83%	63%	88%	71%	75%	83%	76%
T16	70%	64%	88%	67%	67%	71%	67%	70%
Mittel	80%	71%	76%	69%	63%	72%	74%	72%

Tabelle 8.2.: Gesamtbewertung der MANGO-App durch die Testpersonen. Die Testpersonen bewerteten die App mit 72 % (± 5 %, 90 % Konfidenzintervall). Anmerkung: Für das Transformieren der Werte in Prozente kann der SUS-Score direkt als Prozentwert angesehen werden. Die originalen UEQ-Ergebnisse (von -3 bis $+3$) mussten zuvor um $+3$ in Bewertungen von 0 bis 6 skaliert werden, um diese dann in Prozentwerte durch Division durch 6 umzuwandeln. Mit den skalierten Bewertungen (von 0 bis 6) wurden keine Berechnungen durchgeführt.

Testperson #	Dauer pro KA (in Sek.)	Erfolgreiche KA (von 5)	Erfüllte Ziele (von 8)	Anzahl Abstürze	sus-Score	Dauer	KA	Ziele	Fehlerfreiheit	sus	Mittelwert	Mittelwert inkl. sechs UEQ-Skalen
T01	58	4	4	0	90	50%	80%	50%	100%	90%	74%	77%
T02	40	4	6	2	82,5	81%	80%	75%	71%	83%	78%	77%
T03	54	4	7	2	90	56%	80%	88%	71%	90%	77%	74%
T04	80	4	2	7	67,5	11%	80%	25%	0%	68%	37%	47%*
T05	87	4	1	6	85	0%	80%	13%	14%	85%	38%	60%*
T06	80	3	3	3	90	12%	60%	38%	57%	90%	51%	63%*
T07	49	4	6	1	77,5	65%	80%	75%	86%	78%	77%	68%
T09	68	4	4	1	50	32%	80%	50%	86%	50%	60%	57%
T10	29	2	4	0	90	100%	40%	50%	100%	90%	76%	76%
T11	85	3	2	4	90	3%	60%	25%	43%	90%	44%	64%*
T12	32	5	8	2	90	94%	100%	100%	71%	90%	91%	89%
T13	35	5	5	1	87,5	89%	100%	63%	86%	88%	85%	82%
T14	39	3	6	0	70	81%	60%	75%	100%	70%	77%	60%*
T15	40	3	4	0	67,5	80%	60%	50%	100%	68%	72%	75%
T16	58	3	4	3	70	49%	60%	50%	57%	70%	57%	64%
Gesamt-Score											66%	69%

Anmerkung: Die Dauer pro KA ist die durchschnittliche Dauer, die für jede KA gebraucht wurde, in Sekunden. Erfolgreiche KA ist die Anzahl von KA (von 5), die von der Testperson erfolgreich abgeschlossen werden konnte. Erfüllte Ziele ist die Anzahl der Benchmark-Ziele 1a, 1b, 2a, 2b, 3a, 3b, 4a, 5b, die von der Testperson bezüglich dem individuellen Erreichen der Ziel-Schwellenwerte erfolgreich abgeschlossen werden konnte. Anzahl Abstürze ist die Anzahl, wie oft die App im Kameramodus abgestürzt ist (Blackscreen, von dem man zurück zur Startseite der App kommt). sus-Score (von 0 bis 100) ist die Bewertung der App durch die Testperson mit dem sus, höher ist besser. Transformieren in Prozente: Für die Dauern wurde die Differenz der längsten Dauer (87) und der gebrauchten Zeit durch die Differenz der längsten (87) und kürzesten (29) Dauer dividiert, um von 0 bis 100% skalierte Werte zu erhalten. Die KA bzw. Ziele wurden durch 5 bzw. 8 dividiert. Die Fehlerfreiheit ist die Anzahl der Abstürze dividiert durch die höchste Anzahl der Abstürze (7), subtrahiert von 1. Der sus-Score kann direkt als Prozent-Wert angesehen werden.

Tabelle 8.3.: Gesamt-User-Performance im Test der MANGO-App, Version 0.8.5. Im Mittel ergibt sich eine User Performance von 66 % ($\pm 7\%$, 90 % Konfidenzintervall), unter Einbezug der transformierten UEQ-Ergebnisse aus Tabelle 8.2 ergibt sich eine User Performance von 69 % ($\pm 5\%$, 90 % Konfidenzintervall). Die Gesamt-Performance verändert sich nur gering, die deutlichsten Veränderungen bei einzelnen Testpersonen wurden markiert.

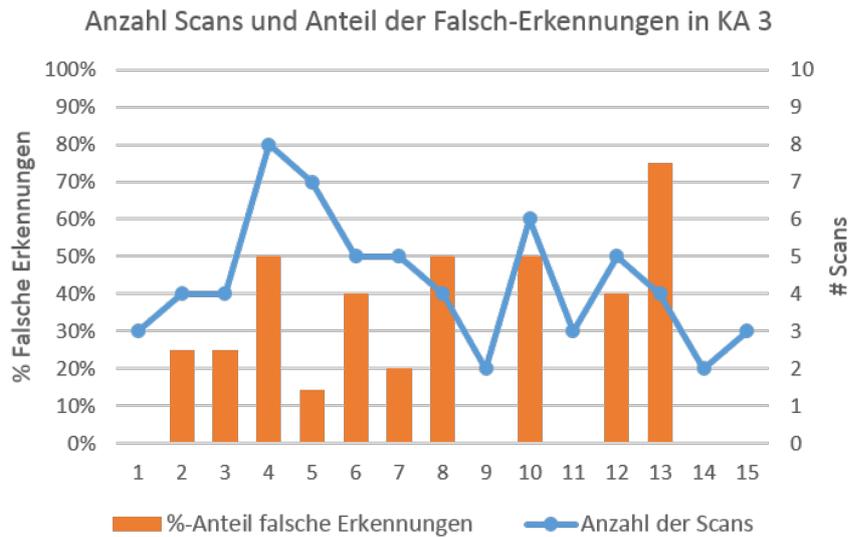


Abbildung 8.8.: Die Anzahl der Lebensmittel-Scans pro Person (blaue Line) und der Anteil der falschen Lebensmittel-Erkennungen in % (orange Balken) in der Kernaufgabe 3.

8.2. Post-Test-Fragebögen

Die Beurteilung der App durch die Testpersonen wurde mit zwei vorgefertigten Post-Test-Fragebögen, dem SUS und UEQ, abgefragt. In die Auswertung flossen die Ergebnisse aller sechzehn Testpersonen ein. Das Ergebnis des SUS zur Bewertung der Usability kann einen Wert zwischen 0 und 100 ergeben. Die App wurde von den Testpersonen mit einem durchschnittlichen SUS-Score von 78 bewertet, die Einzelergebnisse sind in Abbildung 8.9 dargestellt. Der UEQ liefert je ein Ergebnis zwischen -3 und $+3$ auf den sechs Dimensionen Attraktivität, Durchschaubarkeit, Effizienz, Steuerbarkeit, Stimulation und Originalität, zur Bewertung der UX. Ein Ergebnis $> 0,8$ kann als positiv angesehen werden, was sich in der Ergebnis-Darstellung in Abbildung 8.10 als grüne Farbe widerspiegelt. Relativ zu anderen Produkten, die mit dem UEQ bewertet wurden, lässt sich die MANGO-App wie in Abbildung 8.11 einstufen.

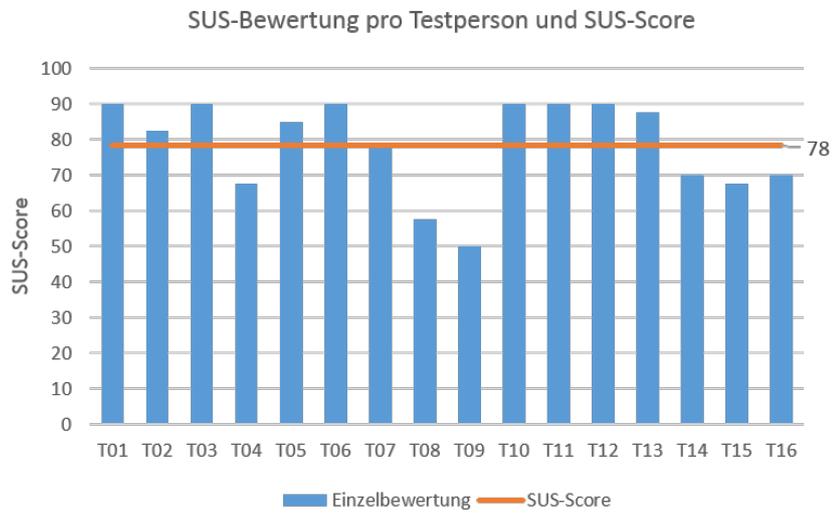


Abbildung 8.9.: SUS-Score und Einzelbewertungen der MANGO-App, Version 0.8.5.

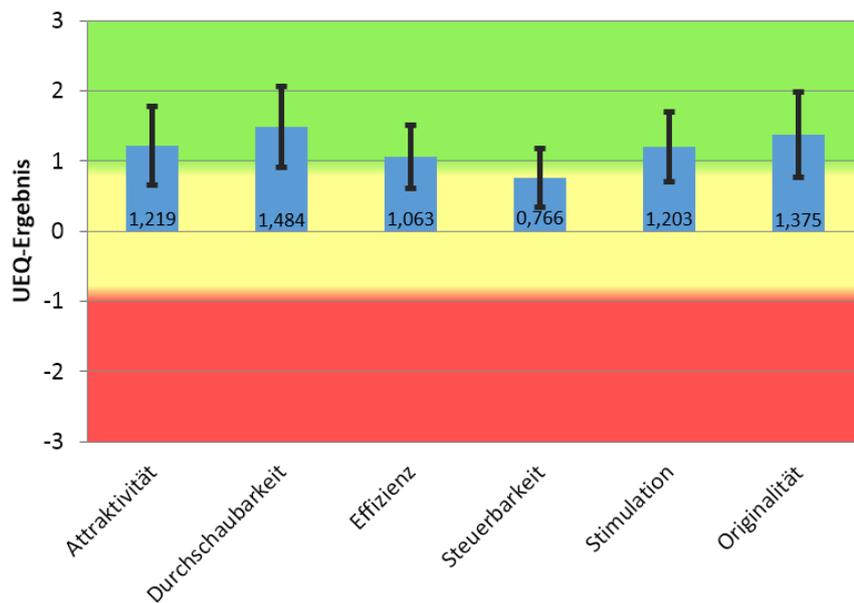


Abbildung 8.10.: UEQ-Ergebnis der MANGO-App, Version 0.8.5.

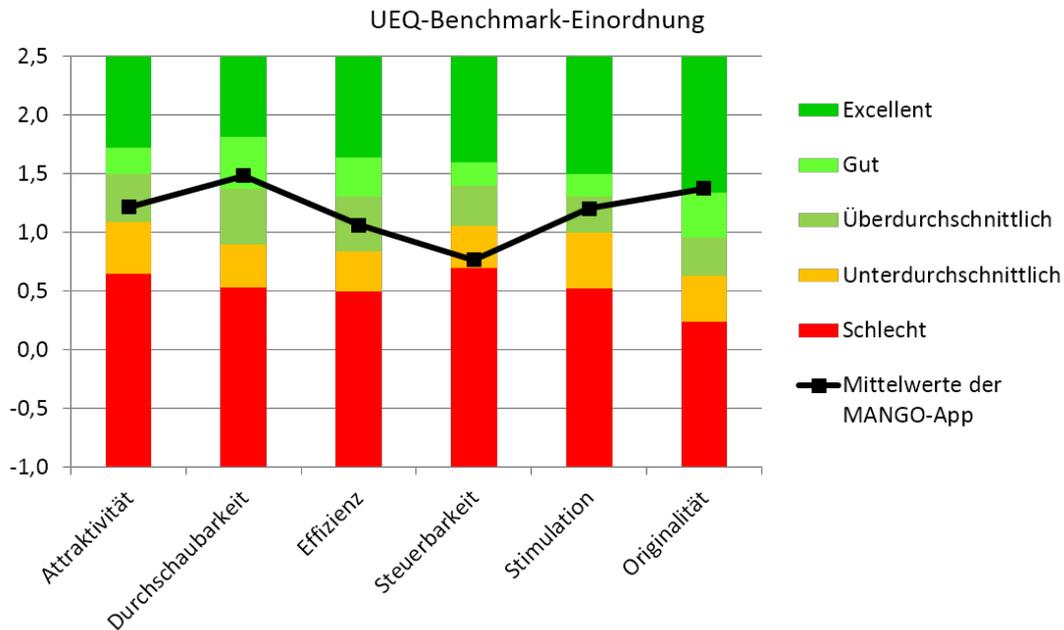


Abbildung 8.11.: UEQ-Benchmark-Einordnung der MANGO-App, Version 0.8.5.

8.3. Eye Tracking und die Seven Stages of Action

Die Aufgabe 2 eignet sich am besten zur weiteren Analyse von Blickbewegungsdaten, da für diese Aufgabe von allen Testpersonen einheitlich der Kameramodus verwendet wurde. Da für JR der Kameramodus am interessantesten ist, kann die Auswertung auf die Kernaufgabe 3 (siehe Tabelle 6.3) reduziert werden. Von den Ergebnissen der 16 Testpersonen konnten 15 für diese Auswertung herangezogen werden, ein Ergebnis musste ausgeschieden werden, da sich das Smartphone ständig zu mehr als 50 % außerhalb des Blickfeldes der ETG befand.

8.3.1. Aktionsbestimmung mit den SSA

Bei der Aufgabe 2 hatten sieben Personen (47% von $n = 15$) exakt dieselbe – zu erwartende – Abfolge von unterschiedlichen Aktionen: sie öffneten den Kameramodus, scannten ein Lebensmittel, riefen Detailinformationen auf, scannten das

zweite Lebensmittel, riefen dazu Detailinformationen auf und scannten das dritte Lebensmittel und riefen schließlich auch dazu Detailinformationen auf. Es können also drei unterschiedliche Aktionen, nämlich Kamera öffnen (K), Scannen (S) und Detailinformationen aufrufen (D), definiert werden, in der zuvor genannten Abfolge sind das sieben Schritte, als Kurzform dargestellt: [KSDSDSD]. Weitere fünf Personen (33 %) führten zwischen einem und sechs zusätzlichen Schritten aus, die sich aus wiederholtem Scannen und/oder wiederholtem Aufrufen von Detailinformationen zusammensetzten [z. B. KSDDSSDSDSD]. Drei Personen (20 %) führten nur sechs Schritte durch, zwei Personen davon scannten nach dem Kamera öffnen vier Mal am Stück, gefolgt von einem Aufruf von Detailinformationen [KSSSSD], eine Person rief nach dem zweiten Scan Details von zwei verschiedenen Lebensmitteln auf [KSDSDD] und hatte somit nur zwei Scan-Vorgänge.

Der Schritt Kamera öffnen ist in der KA 3 nicht mehr enthalten und wird nicht näher beschrieben. Die beiden übrigen Schritte wurden aufgrund des Videomaterials der ETG in die sieben Stages of Action (SSA) unterteilt. Diese beinhalten, neben dem Ziel, jeweils drei Phasen, die der Durchführung (Planen, Spezifizieren, Ausführen) bzw. Evaluation (Wahrnehmen, Interpretieren, Vergleichen) zuzuordnen sind (siehe Kapitel 5). Durch wiederholtes Durchlaufen von Phasen, wie dem Ausführen mehrerer Aktionen zur Erreichung des selben Zieles, können Phasen in einem Schritt mehrfach vorkommen. In die Evaluation kommen die Testpersonen, sobald eine Systemresponse (z. B. eine Displayänderung) sichtbar ist. Ein SSA-Lebensmittel-Scan kann frühestens nach einer Ergebniseinblendung beendet sein, kann aber, im Gegensatz zu den Scans wie sie in Ziel 3a des Usability Benchmarks (siehe Tabelle 6.4) formuliert sind, auch mehrere Scan-Ergebniseinblendungen beinhalten, wenn beispielsweise ein oder mehrere falsche Ergebnisse zuvor angezeigt wurden, es aber weiter versucht wird das selbe Lebensmittel zu scannen. Wird nach der Einblendung eines falschen Ergebnisses das Lebensmittel gewechselt beginnt ein neuer SSA-Scan, wird das Lebensmittel gewechselt, ohne dass vorher eine Ergebniseinblendung stattgefunden hat, bleibt man im selben SSA-Scan. Die meisten SSA-Scans wurden durch das Aufrufen von Detailinformationen, also durch das Formen eines neuen Zieles, beendet.

Nicht alle Handlungen in den Phasen werden von den Testpersonen bewusst ausgeführt, also können Phasen oft nicht scharf voneinander getrennt werden. Für die

Unterteilung wurden daraufhin Handlungen zu den einzelnen Phasen spezifiziert und die Unterteilung erfolgte konsequent nach diesen Vorgaben für alle Testpersonen, sodass vergleichbare Ergebnisse entstehen konnten. In dieser Diplomarbeit wurden so alle Phasen zu allen Schritten mit ihrer Zeit erfasst und zu den entstandenen Zeit-Intervallen wurden die Eye-Tracking-Rohdaten mit einem MATLAB-Skript von JR wie in Perko et al. (2014) nach Holmqvist et al. (2011), zugeordnet und zusammengefasst, sodass darauf hin für alle Testpersonen die Anzahl der Fixationen, die Fixationen pro Sekunde, die durchschnittliche Blickdauer in Sekunden, die Anzahl der Sakkaden, die Sakkaden/Sekunde und die durchschnittlichen Pupillendurchmesser für das linke und rechte Auge pro Zeit-Intervall ermittelt waren. Für jede Testperson wurden die Eye-Tracking-Daten zuerst in jedem Schritt (1. scannen, 1. lesen, etc.) exakt auf die SSA reduziert (Wiederholungen wurden zusammengefasst) und daraufhin wurden über alle Testpersonen die Mittelwerte jeweils zu den ersten drei Schritten (1. scannen, 1. lesen, . . . , 3. scannen, 3. lesen) gebildet, die die Datengrundlage für die nun präsentierten ET-Daten darstellen.

8.3.2. Eye-Tracking-Ergebnisse

Es wurden die Dauer der SSA, die Fixationen pro Sekunde und die durchschnittlichen Blickdauern in den SSA für die ersten drei SSA-Scans analysiert. Weiters werden einige erhobene Daten von konkretem User-Verhalten präsentiert. Nach der Unterteilung aller Schritte in die Phasen der SSA wurde der Schritt Detailinformationen aufrufen bzw. nachlesen nicht weiter analysiert, da keine Testpersonen Probleme bei diesem Schritt hatten und sich beliebig lange über die Lebensmittel informiert haben. Die Testpersonen verweilten zwischen vier Sekunden und einer Minute auf den aufgerufenen Detailseiten.

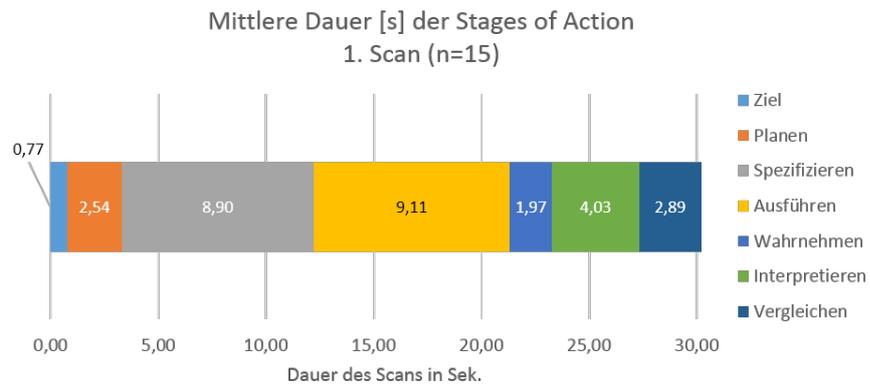
Die Dauer der jeweiligen Stages of Actions für die ersten drei SSA-Scans ist in Abbildung 8.12 dargestellt. Das Spezifizieren (grau) und Ausführen (gelb) kann als zusammengehörig aufgefasst werden, da diese Zeiten die immer wiederholten Versuche der Testpersonen, eine Auslösung der Erkennung (z. B. durch Berühren des Displays) zu bewirken inkludieren, die App auf diese Berührungen des Kamerabildes aber nicht reagiert. Die Formulierung des Ziels dauert immer nur einen kurzen

Moment, da sofort mit der Planung, dem neu Positionieren, begonnen wird. Die Planung wird sofort durch das Spezifizieren und Ausführen von Handlungen abgelöst. In das Wahrnehmen kommen die Testpersonen, sobald eine Displayänderung erkennbar wird. Mit diesem Wissen können die mittleren Blickdauern in den SSA-Scans in Abbildung 8.13 besser interpretiert werden, denn die Phasen Ziel und Planen sind somit kaum noch relevant. Abbildung 8.14 zeigt die gemittelten Anzahlen der Fixationen pro Sekunde in den SSA der Scans, welche nun ohne die Dauer der Stages visualisiert sind, da diese bereits pro Sekunde ausgedrückt sind.

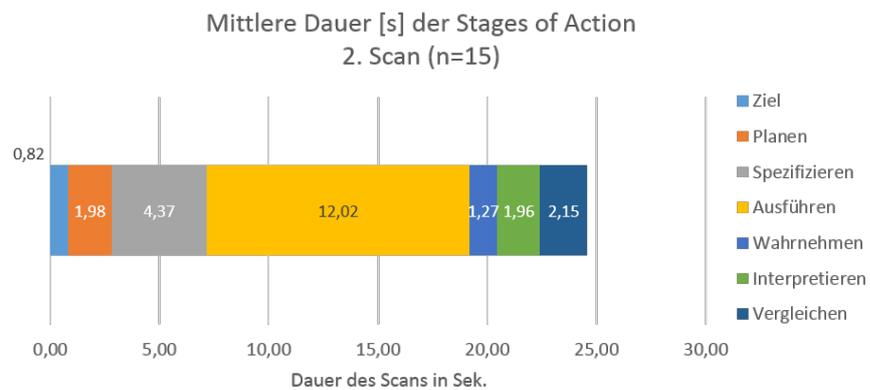
In Heat Maps werden die Anzahl und Dauer der Fixationen als farbcodierte Visualisierungen dargestellt und mit Hilfe der *Smartphone Eye Tracking Toolbox* (SMET) (Paletta et al., 2014) (siehe Kapitel 6.5.2), kann direkt der Smartphone-Bildschirm mit diesen farbcodierten Visualisierungen überlagert werden. Da die Testpersonen das Scannen von Lebensmitteln mit der App sehr unterschiedlich gelöst haben, werden zunächst typische Benutzerverhalten identifiziert, um Heat Maps nach diesen Gruppierungen zum ersten, zweiten und dritten Scan zu erstellen.

Die Testpersonen haben während der drei SSA-Scans mit der App zwischen zwei und sechs (teilweise falsche) Lebensmittel-Erkennungen durchgeführt. Zusätzlich zu den durchgeführten Erkennungen ist in Abbildung 8.15 die Anzahl der Abstürze des Kameramodus der einzelnen Testpersonen dargestellt. Nach einem Absturz kommen die Testpersonen bezüglich der SSA in die Evaluation (erste Systemresponse – auch wenn nicht gewollt), in der sie wieder in den Kameramodus zurückkehren. Die Erkennung von Lebensmitteln geschieht mit der App im Kameramodus kontinuierlich durch eine Videoaufnahme, es ist kein Eingreifen der Testpersonen erforderlich (siehe Kapitel 6.1). 14 von 15 Testpersonen haben in ihrem ersten SSA-Scan durch bis zu 8-maligem Berühren des Apfel-Symbols, das im Kamerasucher dargestellt wird und keine Funktionalität besitzt, versucht eine Erkennung durchzuführen. Acht Personen (53 %) haben über alle drei SSA-Scans versucht Erkennungen durch Berührung des Apfel-Symbols aktiv auszulösen. Die Anzahl der Berührungen des Apfel-Symbols in den drei SSA-Scans ist in Abbildung 8.16a pro Testperson dargestellt, Berührungen an anderen Displaystellen sind nicht mitgezählt. Beim wiederholten Berühren des Apfel-Symbols, das der Phase der Ausführung in den SSA zuzuordnen ist, bleiben die Testperson in der Durchführung und kommen gar nicht in die Evaluation, da es keine Systemresponse gibt.

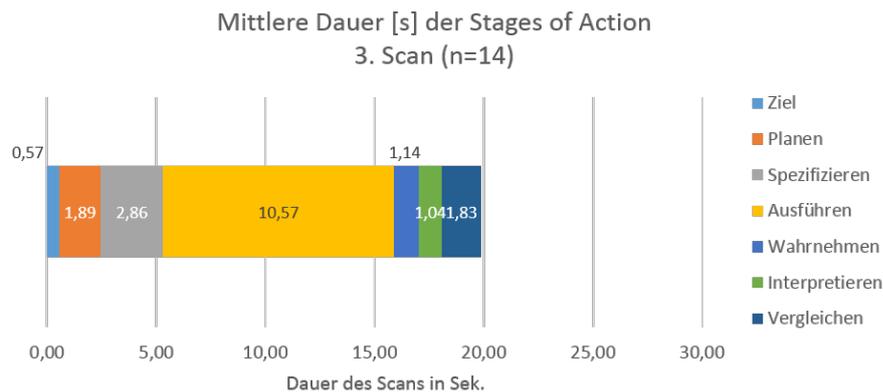
8. Ergebnisse



(a) Durchschnittliche Dauer der SSA (in Sek.) 1. Scan



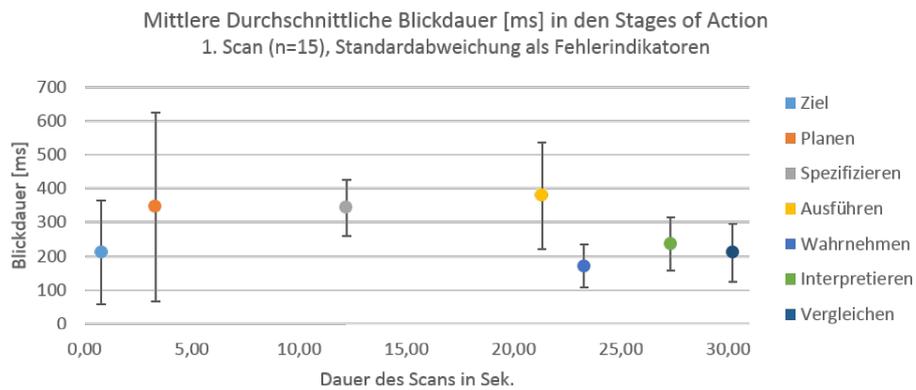
(b) Durchschnittliche Dauer der SSA (in Sek.) 2. Scan



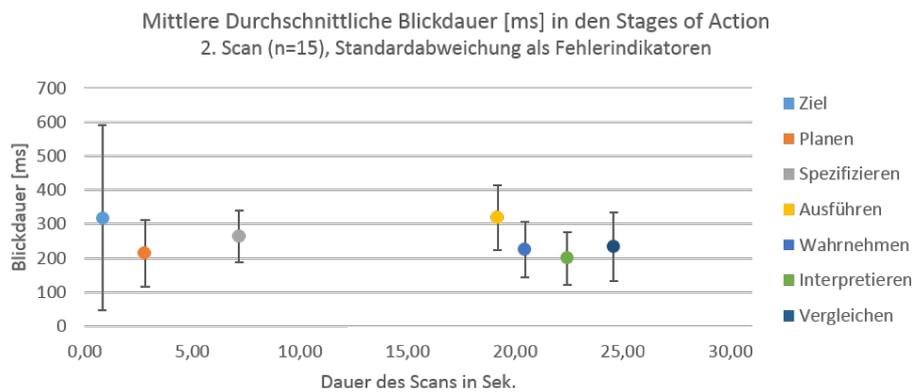
(c) Durchschnittliche Dauer der SSA (in Sek.) 3. Scan

Abbildung 8.12.: Die durchschnittliche Dauer der SSA in Sekunden, dargestellt bezüglich der ersten drei Lebensmittel-Scans.

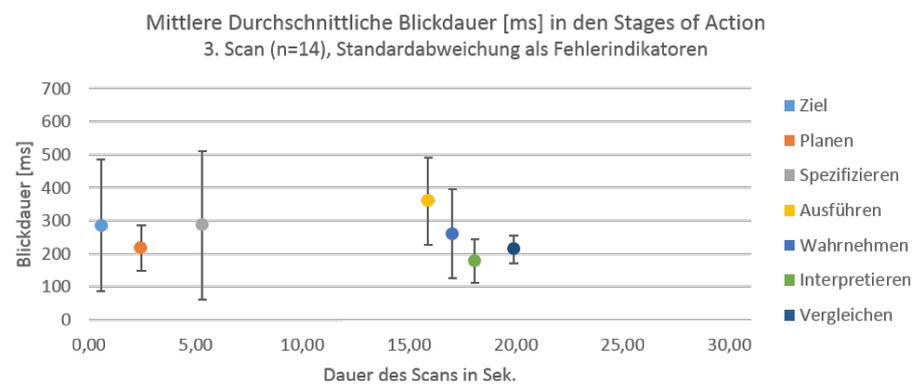
8.3. Eye Tracking und die Seven Stages of Action



(a) Durchschnittliche Blickdauer (in ms) 1. Scan



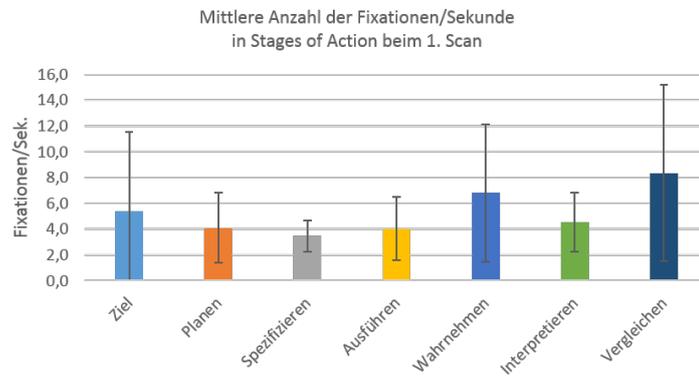
(b) Durchschnittliche Blickdauer (in ms) 2. Scan



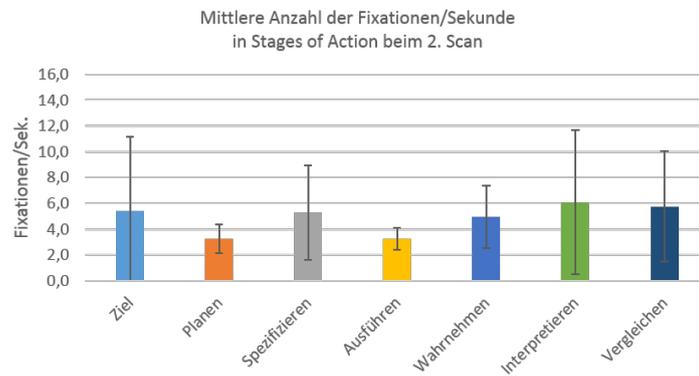
(c) Durchschnittliche Blickdauer (in ms) 3. Scan

Abbildung 8.13.: Durchschnittliche Blickdauern in den ersten drei SSA-Scans in Millisekunden mit Standardabweichung, dargestellt bezüglich der Dauern der SSA.

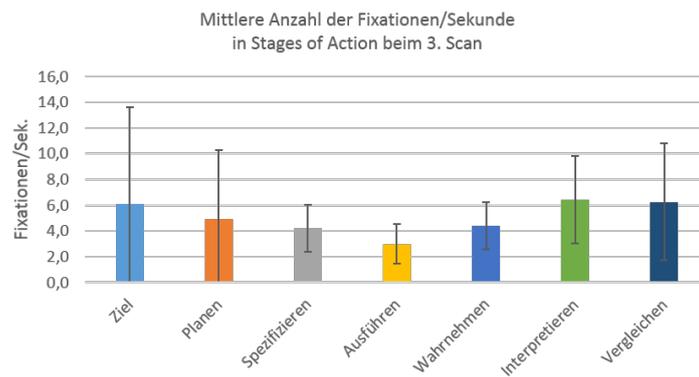
8. Ergebnisse



(a) Durchschnittliche Fixationen/Sek. in den SSA 1. Scan



(b) Durchschnittliche Fixationen/Sek. in den SSA 2. Scan



(c) Durchschnittliche Fixationen/Sek. in den SSA 3. Scan

Abbildung 8.14.: Durchschnittliche Fixationen pro Sekunde in den ersten drei SSA-Scans.

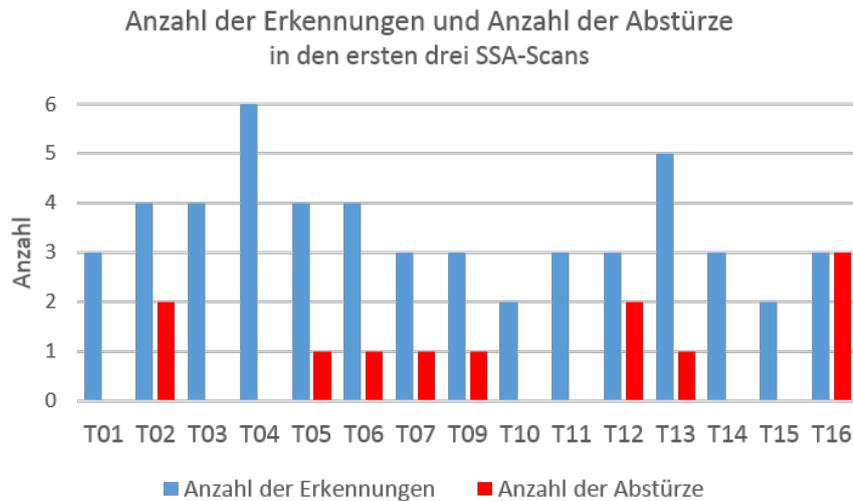
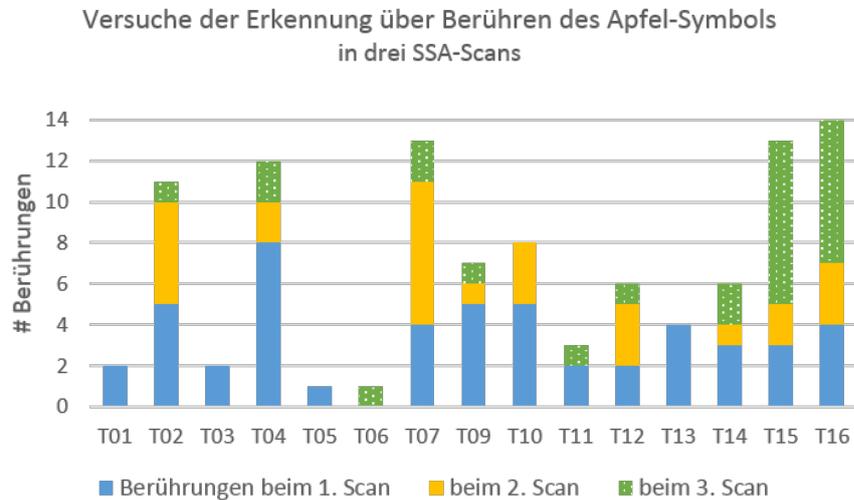


Abbildung 8.15.: Anzahl der Lebensmittel-Erkennungen der App und Anzahl der Abstürze des Kameramodus bei den drei SSA-Scans.

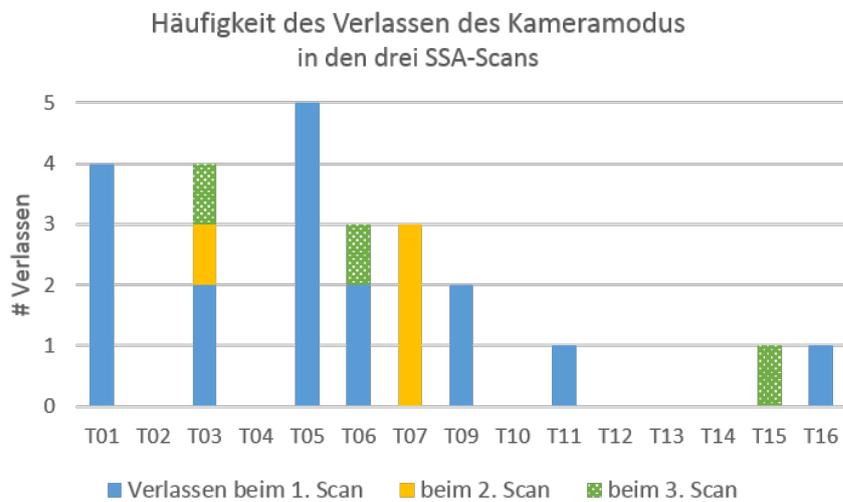
Beim Versuch aktiv eine Erkennung auszulösen haben 7 von 15 Testpersonen im ersten SSA-Scan bis zu fünf Mal die App geschlossen oder anderweitig den Kameramodus verlassen (Abstürze wurden nicht mitgezählt). Eine Person hat in allen drei SSA-Scans den Kameramodus verlassen, sechs Personen jedoch in keinem SSA-Scan. Die Aufteilung, welche Testpersonen wie oft die App in den drei SSA-Scans verlassen haben, ist in Abbildung 8.16b dargestellt. Beim Verlassen des Kameramodus kommen die Testpersonen bezüglich der SSA, wie bei einem Absturz, in die Evaluation, da es eine Systemresponse gegeben hat. In den Phasen der Evaluation kehren die Testpersonen wieder in den Kameramodus zurück.

Zusammenfassend haben die Abstürze des Kameramodus (siehe Abbildung 8.15), die Versuche aktiv eine Erkennung auszulösen (vgl. Abbildung 8.16a) und die Häufigkeit des Verlassen des Kameramodus (siehe Abbildung 8.16b) eine Auswirkung auf das Durchlaufen der SSA, was als Iterationen von Durchläufen der SSA, wie in Abbildung 8.17, pro Testperson und Scan dargestellt werden kann.

Bezüglich der unterschiedlichen Herangehensweisen der Testpersonen wurden Heat Maps mit SMET erstellt. Die Testpersonen wurden aufgrund der Unterschiede in Abbildung 8.16 in die folgenden Gruppen unterteilt: Personen, die in jedem SSA-Scan



- (a) Häufigkeit der Verwendung des Apfel-Symbols, um eine Erkennung aktiv auszulösen, über alle drei SSA-Scans. Berührungen an anderen Displaystellen sind nicht enthalten.



- (b) Häufigkeitsverteilung des Verlassen des Kameramodus in den drei SSA-Scans.

Abbildung 8.16.: Einflussgrößen für die Gruppeneinteilung der Testpersonen für die Heat-Map-Erstellung

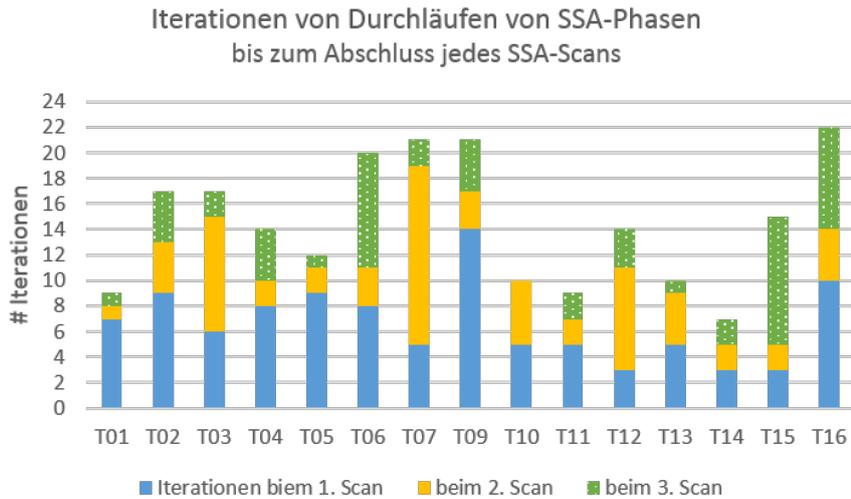


Abbildung 8.17.: Häufigkeit der Iterationen der SSA pro Testperson in den drei SSA-Scans.

das Apfel-Symbol als Auslöser verwendet haben, aber nicht den Kameramodus verlassen haben (T02, T04, T12, T14), und Personen, die den Kameramodus oft verlassen haben, wenn möglich aber bezüglich des Apfel-Symbols nicht sehr auffällig waren (T03, T01, T05, T06), und Personen, die gemischte Vorgehensweisen hatten.

Für die Auswertungen mit SMET wurden die Videos der ETG, die Screen Recordings des Smartphones und die ET-Rohdaten benötigt. Aufgrund zahlreicher Spiegelungen des Lichts der Leuchtstoffröhren der Deckenbeleuchtung am Display des Smartphones und weil die meisten Testpersonen, trotz des Hinweises auf die Korrekte Verwendung, das Smartphone wie gewohnt in der Hand gehalten und dabei oft die Ecken des Smartphones verdeckt haben, war das Display-Tracking durch SMET nicht in einem Ausmaß möglich, als das aussagekräftige Heat Maps hätten erstellt werden können. Abbildung 8.18 zeigt einen Ausschnitt einer dieser teilweisen erfolgreichen Erkennungen. Die Ergebnisse der unterschiedlichen User-Wahrnehmung des Displays konnte also nicht ausgewertet werden.

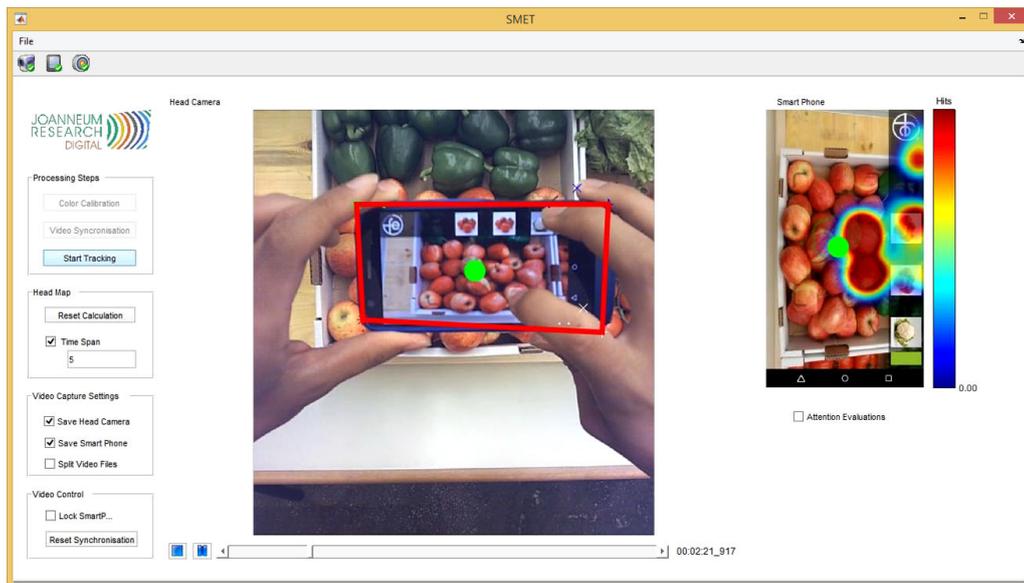


Abbildung 8.18.: Teilweise Erkennung von Daten von T04 durch SMET. Erkennt SMET das Smartphone im Video, wird dies durch einen roten Rahmen gekennzeichnet und die errechneten Blickdaten als farbcodierte Überlagerung des Screenrecordings dargestellt. Durch zahlreiche Störungen konnten die Daten nur sehr lückenhaft in Heat-Maps umgewandelt werden, was eine objektive Beurteilung nicht möglich macht.

9. Diskussion und Erkenntnisse

Dieses Kapitel dient der Diskussion und Interpretation der wesentlichen Ergebnisse der ersten Usability-Testung der App MANGO (Version 0.8.5). Der erstellte Usability Benchmark (siehe Tabelle 8.1) stellt die Basis für folgende Usability-Tests dar.

9.1. Gesamtbewertung der App

Auf den ersten Blick wirft die Tatsache, dass 77% der Ziele, die für den Usability Benchmark definiert wurden, von den Testpersonen bzw. der App nicht erfüllt wurden ein schlechtes Licht auf die Usability der App. Ein wesentlicher Grund hierfür war, dass die definierten Zeitvorgaben nicht eingehalten werden konnten. Ein Hinweis dafür ist auch, dass die in Tabelle 6.4 definierten Gesamt-Ziele fast im vollen Maße erfüllt wurden, in denen die Ziele zu Performance und Stabilität der App ausgeklammert sind.

Zieht man die Bewertungen durch die Testpersonen mit dem SUS und UEQ (siehe Kapitel 8.2) heran, so schneidet die App bezüglich Usability deutlich positiver ab, als es die Gesamt-Performance der Testpersonen (zwischen 59 und 73%, siehe Tabelle 8.3) aussagen würde. Betrachtet man nur die Ergebnisse der User-Bewertungen ist man geneigt darauf zu schließen, dass es sich bei der App bereits um eine runde Sache handelt, die von den Testpersonen gut angenommen wird, jedoch die Steuerbarkeit verbessert werden muss, denn lediglich die Steuerbarkeit liegt laut dem UEQ Benchmark in einem unterdurchschnittlichen Bereich, knapp über den schlechtesten 25% der Vergleichswerte. Es sei zu den Fragebögen noch erwähnt, dass bei solchen subjektiven Bewertungen durch Testpersonen extreme Antworttendenzen meist vermieden werden, und die Ergebnisse einen Bias aufweisen können.

Um ein größeres Gesamtbild zu erhalten und um diesen Bias auszugleichen ist es aber auf jeden Fall besser, noch andere Maße zu berücksichtigen. Die schlechte Bewertung der Steuerbarkeit kann auf die zahlreichen Abstürze des Kameramodus (vgl. Ziel 3d Usability Benchmark – zweit schlechtesten Erfüllungsgrad) zurückgeführt werden. Diese Abstürze trugen wiederum dazu bei, dass die Testpersonen an der Erfüllung der zu den Kernaufgaben 3 und 4 definierten Ziele gehindert wurden. Bei allen Personen die zum Lösen der KA 4 den Kameramodus verwendeten, kam es zu Abstürzen – auch bei Personen, die in KA 3 keine Abstürze erlebt hatten. Es konnte aus den Beobachtungen nicht festgestellt werden, ob diese Abstürze aufgrund einer gewissen Verwendungsweise der App aufgetreten sind.

Ein unerwartetes Ergebnis war, dass sich die Akzeptanz der App vor und nach der Benutzung bei den männlichen und weiblichen Teilnehmern an der Studie signifikant unterschied. Während die Bewertung durch die männlichen Teilnehmer zurückgegangen war, gaben umgekehrt mehr weibliche Teilnehmerinnen nach dem Test an die App wahrscheinlicher verwenden zu wollen. Eine mögliche Erklärung wäre, dass Frauen sich generell stärker als Männer für eine individuelle Ernährungsform, wie es FE bietet, interessieren könnten. Eine Betrachtung der Einzelergebnisse zeigt, dass speziell bei den Männern einige Studienteilnehmer die vorher großes Interesse an der App bekundet hatten, nach der Studie kaum noch Interesse hatten die App zu verwenden. Dieser Interessensverlust könnte inhaltlich begründet sein oder auf die Abstürze bzw. die schlechte Steuerbarkeit zurückzuführen sein. Tatsächlich lässt sich aus der Gesamt-User-Performance (Tabelle 8.3) feststellen, dass eine dieser männlichen Testpersonen die meisten Abstürze im Kameramodus und mit der damit verbundenen länger gebrauchten Zeit insgesamt die schlechteste User Performance mit 37 % (bzw. 47 % unter Berücksichtigung der Bewertung durch den UEQ) hatte. Für diese Person wäre es also plausibel, dass sie die App aufgrund der Fehleranfälligkeit nicht verwenden wollen würde. Eine andere betroffene männliche Testperson wies eine durchschnittliche Performance auf, auffällig war jedoch, dass diese Person im Gegensatz zur anderen Testperson im Laufe der Untersuchung keinen einzigen Absturz des Kameramodus erlebt hat. Für diese Person könnte der Ablehnungsgrund auf einer inhaltlichen Ebene gefunden worden sein. Für eine genaue Analyse dieser Beobachtungen war jedoch das Feld der Testpersonen zu klein.

9.2. Identifizierte Probleme

Die Aktivitäts-Analyse hat verschiedene Vorgehensweisen der Testpersonen bei der Verwendung des Kameramodus aufgezeigt. Die App bietet ihren Benutzerinnen und Benutzern während des Lebensmittel-Erkennungsprozesses kein Feedback zur richtigen Handhabung oder darüber, ob Daten bereits verarbeitet worden sind. Dieses fehlende Design-Prinzip „Feedback“ (siehe Tabelle 5.2) wirkte sich auch negativ auf die Treffergenauigkeit der Bilderkennung aus. Nur ein Drittel der Testpersonen erhielt ausschließlich richtige Scan-Ergebnisse (vgl. Ziel 3c Usability Benchmark – viert schlechtesten Erfüllungsgrad). Mitunter könnte das fehlende Feedback ebenfalls Einfluss auf die schlechte Bewertung der Steuerbarkeit gehabt haben.

Auch die Erkennungsraten hingen dadurch sehr stark von der individuellen Benutzung durch die Testpersonen ab. Vergeht beispielsweise bei laufendem Kameramodus längere Zeit bis sich eine Person für das nächste zu scannende Lebensmittel entscheidet, werden durch die kontinuierliche Erkennung nicht relevante Bilder in die Auswertung miteinbezogen, welche das Ergebnis negativ beeinflussen können. Außerdem kam es beim Test vor, dass Personen, die sehr schnell zwischen den Lebensmitteln wechselten, das Ergebnis des vorherigen Scans eingeblendet bekamen, was in Bezug auf das aktuell ausgewählte Lebensmittel als Fehlererkennung gewertet werden musste.

Die Dauer der Scans wurde ebenfalls oft von den Testpersonen beeinflusst. Die meisten Personen haben zu einem Kameramodus im Allgemeinen das konzeptuelle Modell entwickelt, dass ein Ergebnis einen Auslöser – eine Aktion der Person – benötigt. Dieses konzeptuelle Modell wird durch die App nicht aufgegriffen, was die Testpersonen jedoch nicht aufhält, sich auf gewohnten Pfaden zu bewegen. Alles was aussieht, als ob es eine Funktion hätte, wurde auf diese hin getestet. Dies beinhaltete die Android-Navigationsleiste, über die die App verlassen werden kann, ebenso, wie das Apfel-Symbol im Kamerasucher, das hinweisen sollte, dass der Modus nur mit Lebensmitteln verwendet werden kann.

Ein kurzes Intro mit einem ebenso kurzen Tutorial, die Integration von Feedback während des Scannens, oder die Rücksichtnahme auf das allgemeine konzeptuelle

Modell einen Kameramodus betreffend, könnten die Schwierigkeiten der Testpersonen im Kameramodus deutlich reduzieren.

Die Implementierung der Darstellung der wichtigsten Lebensmittel einer Kategorie (Obst, Milchprodukte, etc.) in der Lebensmitteldatenbank wurde von den Testpersonen nicht im vollen Umfang verstanden. Die Kernaufgabe 5 diente zur Überprüfung, ob die Kernaufgabe 4, die zu 67 % erfolgreich abgeschlossen wurde, tatsächlich durch Informationen aus der App gelöst wurde. Indirekt wurde damit abgefragt, ob die Darstellung der wichtigsten Lebensmittel wahrgenommen wird. Diese Darstellung findet sich nicht auf der Übersichtsseite der Lebensmitteldatenbank bei der geöffneten Kategorie (z. B. Gemüse), sondern auf jeder Einzelseite jedes Lebensmittels (für die Kategorie des gewählten Lebensmittels). Erstere Ansicht ist alphabetisch sortiert (vgl. Abbildung 6.2b unten links) und enthält die Gewichtsangaben der Portionsgrößen. Diese beiden Informationen wurden von den Testpersonen oft fälschlicherweise zur Lösung der Kernaufgabe 5 verwendet.

9.3. Eye Tracking

Die gewonnenen Erkenntnisse aus der Aktionsanalyse sind in den bisherigen Schlüssen bereits eingeflossen. Bei den Eye-Tracking-Daten zu den SSA-Scans sind vor allem die Phasen Spezifizieren und Ausführen interessant, denn alle anderen Phasen werden so schnell durchlaufen, dass sie kaum auf den Einfluss eines gewissen Bestandteils zurückgeführt werden können. Bei den durchschnittlichen Blickdauern fallen im ersten SSA-Scan die erhöhten durchschnittlichen Blickdauern (250–300 ms wäre eine typische Fixation (Goldberg et al., 1999)) der Phasen Spezifizieren und Ausführen auf, die über die Fortdauer der Scans aber nur leicht abnimmt. Bei der mittleren Anzahl der Fixationen pro Minute (Abbildung 8.14) kann ebenfalls keine positive Entwicklung gesehen werden.

Es konnten aufgrund der in Kapitel 8.3.2 beschriebenen Probleme keine vollständigen Heat Maps erstellt werden. Aus der Beobachtung kann jedoch berichtet werden, dass Personen der Gruppe, die oft den Kameramodus verlassen haben, eher auf die Funktions-Icons am Display fokussiert waren. Personen, die oft das Apfel-Symbol als Auslöser verwendet haben, waren hingegen mit dem Blick beim Apfel-Symbol

selbst. Nach der ersten Ergebniseinblendung bleibt die Ergebnisleiste bis zur nächsten Erkennung eingeblendet und die Testpersonen erwarteten ab dem zweiten Scan auch eine Änderung des Ergebnisses, und fokussierten immer wieder auf die Ergebnisleiste.

Für das Unterteilen in die SSA ist ein sehr hoher Aufwand notwendig. Es muss auf den Frame genau bestimmt werden, wann Phasen beginnen und aufhören. Die Eye-Tracking-Daten bzgl. dieser Unterteilungen zu berechnen und zu betrachten, brachte keine Aufschlüsse über ein mögliches Erlernen der Scan-Funktion. Eine Fortführung dieser Methode ist also zu hinterfragen.

9.4. Limitation der Arbeit

Die Auswertung der Eye-Tracking-Daten bei den Scan-Vorgängen gestaltete sich schwieriger als angenommen. Diese wurden anhand der Zeitintervalle der einzelnen Stages of Action berechnet. Wie bereits dargelegt, sind für die Interpretation die Phasen Spezifizieren und Ausführen am interessantesten, da die beiden Phasen davor, Ziel formulieren und Planen der Handlung, sehr schnell durchlaufen werden, und sich die Durchführung des Scannens der Benutzerinnen und Benutzer auf die Phasen Spezifizieren und Ausführen beschränkt, da die drei nachfolgenden Phasen der Evaluation, Wahrnehmen, Interpretieren und Vergleichen, Reaktionen auf eine Systemresponse darstellen. Im Fall des Scannens ist diese Systemresponse meist eine Ergebnisdarstellung, die, wenn sie falsch ist, von den Benutzerinnen und Benutzern erkannt wird und das gewollte Lebensmittel wird abermals gescannt und somit werden wieder die Phasen Spezifizieren und Ausführen durchlaufen, und wenn die Erkennung richtig ist, wird sofort durch die Auswahl des gewünschten Lebensmittels durch die Benutzerinnen und Benutzer in die Detailansicht gewechselt, die Phasen der Evaluation werden sehr schnell durchlaufen, und der Scan somit beendet. In diesen beiden Phasen konnte keine positive Entwicklung der durchschnittlichen Blickdauer und der mittleren Anzahl der Fixationen über das wiederholte Scannen festgestellt werden, was zuvor erwartet wurde.

Da es aufgrund von Spiegelungen am Display und dem Verdecken des Smartphone-Randes durch die Testpersonen nicht möglich war vollständige Heat Maps mit dem

Auswertungstool SMET zu erstellen, konnten die User-Wahrnehmung des Displays und Areas of Interest nicht ausgewertet werden. Es wurde zu Beginn dieser Arbeit angenommen, dass mit SMET, das für eine Anwendung in einem anderen Kontext konzipiert wurde, auch im Human-Factors-Labor vollständige Auswertungen erzielt werden können. Vielleicht hätten durch das vermehrte Hinweisen der Testpersonen, die Ränder des Smartphones nicht zu verdecken, vollständigere Ergebnisse erzielt werden können. SMET ist ein stabiles und validiertes Programm zur automatisierten Auswertung von Eye-Tracking-Daten, wurde aber, im Gegensatz zur Anwendung in dieser Arbeit, für eine Auswertung also auch Verwendung des Smartphones im Hochformat konzipiert. Für die Testpersonen ist es im Hochformat leichter die Ränder des Smartphones nicht zu verdecken, als wenn das Smartphone im Querformat verwendet werden muss und auch die Kamera nicht verdeckt werden darf, wie in der vorliegenden Arbeit. Auf diese ständigen Hinweise zur richtigen Handhabung wurde zugunsten eines natürlichen bzw. intuitiven Benutzens des Kameramodus des Smartphones verzichtet, was wohl die unvollständigen Heat Maps nach sich zog.

9.5. Beantwortung der Forschungsfrage

Können mit Usability Benchmark und Aktivitäts-Analyse durch Eye-Tracking-Unterstützung vertiefende Kenntnisse für die App-Entwicklung im Bereich Ernährungsempfehlung gewonnen werden?

In der vorliegenden Arbeit konnten Erkenntnisse für die App-Entwicklung (siehe Kapitel 9.2) aufgrund des erstellten Usability Benchmarks und der weiterführenden Aktivitäts-Analyse gewonnen werden, obwohl, wie in der Limitation der Arbeit dargelegt, die Eye-Tracking-Daten an sich nicht sehr aufschlussreich waren. Basis für die Erstellung des Usability Benchmarks und für die Unterteilung in die sieben Stages of Action bei der Aktivitäts-Analyse waren die durch die Eye Tracking Glasses aufgenommen Szenen-Videos, also jeweils die Sicht der Testpersonen auf das Smartphone, mit Hilfe derer alle Daten erfasst und ausgewertet werden konnten. In der Aktivitäts-Analyse der Scan-Funktion konnte alleine durch das Erfassen der sieben Stages of Action, also dem Transkribieren der Handlungssequenzen der Testpersonen, viele Handlungsabläufe soweit nachvollzogen werden, dass Mängel in der

Bedienbarkeit der App abgeleitet werden konnten. Die gewonnen Erkenntnisse legen der App-Entwicklung die Erstellung eines allgemeinen Tutorials, die Integration von Feedback während dem Benutzen der Lebensmittelerkennung und eine Überarbeitung der Darstellung der wichtigen Lebensmittel der einzelnen FE-Typen nahe.

10. Zusammenfassung und Ausblick

Es wurde ein erster Usability-Test zur Benchmark-Erstellung für die von JR entwickelte App MANGO durchgeführt. Diese App ermöglicht es dem Benutzer/der Benutzerin Informationen zur individuell angepassten Ernährungsform im Ernährungskonzept FE abzurufen. Zentrales Element der App ist die automatische Lebensmittelerkennung samt Ernährungsempfehlungen mit mobiler Augmented Reality. In Ergänzung des Usability-Tests wurde eine Analyse der Aktivitäten der Testpersonen bei der Verwendung dieser Lebensmittel-Scan-Funktion durchgeführt, wobei Eye-Tracking-Daten mit ETG erhoben und ausgewertet wurden.

Für den Usability-Test wurden zu Anwendungs-Szenarien der App entsprechende Aufgaben erstellt, die von den sechzehn Testpersonen während der Untersuchung durchzuführen waren. Für die Benchmark-Erstellung wurden diese Aufgaben in messbare Kernaufgaben unterteilt, für die Zielvorgaben festgelegt wurden. Darüber hinaus wurden drei Gesamt-Ziele in den Benchmark aufgenommen. Die Zielerreichung wurde nach der Testung durch die Analyse von quantitativen Daten, die aus den Szenen-Videos der ETG gewonnen wurden, überprüft. Der Usability-Test bestand aus einer Vorbefragung, dem praktischen Test der App mit Testpersonen und wurde mit einem Fragebogen zur App-Bewertung abgeschlossen. Für den praktischen Test wurde ein kleiner Obststand, ähnlich dem aus der Frische-Abteilung eines Supermarks aufgebaut und mit Obst und Gemüse befüllt. Der abschließende Fragebogen zur Erhebung der Usability bestand aus zwei Hauptteilen: dem System-Usability-Scale-Fragebogen (SUS) und dem User Experience Questionnaire (UEQ). Zusätzlich enthielt der Fragebogen eine weitere Frage, die auch schon in der Vorbefragung gestellt wurde, nämlich, wie wahrscheinlich die Testpersonen die (zuvor beschriebene bzw. danach getestete) App herunterladen und verwenden würden.

Die Akzeptanz der Testpersonen für die App ist positiv, es gab jedoch erhebliche Probleme in der Steuerbarkeit, vornehmlich durch Abstürze der getesteten Beta-Version. Diese spiegeln sich vor allem deutlich darin nieder, dass ein Großteil der Usability-Benchmark-Ziele nicht erfüllt werden konnten. Die Gesamtziele im Usability Benchmark wurden jedoch fast vollständig erfüllt.

Im Laufe des Tests konnten mehrere Punkte der Verbesserungsmöglichkeit identifiziert werden, von denen ausgehend die Weiterentwicklung der App in Bezug auf Usability aufbauen kann. Zum einen wurde festgestellt, dass den Benutzerinnen und Benutzern ein Einführungstutorial den Umgang mit der App erheblich erleichtern würde. Auch fehlt Feedback im Kameramodus, was sich negativ auf die Erkennungsrate auswirkt, da die Benutzerinnen und Benutzer über den laufenden Scan-Vorgang nicht informiert sind und aktiv auf das Auslösen der Erkennung einwirken wollen.

Eine interessante Beobachtung war die unterschiedliche Akzeptanz der App bei männlichen und weiblichen Testpersonen nach dem Test der App. Die genauen Gründe dafür konnten im Rahmen der Studie nicht gefunden werden.

10.1. Ausblick

Der erstellte Benchmark ist eine Ausgangsbasis für die Weiterentwicklung der App MANGO. Die Integration des empfohlenen Einführungstutorials ist bereits in Planung.

Weitere Usability-Tests der App im Vergleich zu dem im Rahmen dieser Arbeit erstellten Benchmark mit den vorgestellten Methoden werden angestrebt. Zuvor wird jedoch versucht, die Probleme, die während des Tests mit der Steuerung aufgetreten sind, genau zu identifizieren und zu lösen.

Literatur

- Andrews, Keith (2008). „Evaluation Comes in Many Guises“. Revised Position Paper. In: *CHI Workshop: BEyond time and errors: novel evaLuation methods for Information Visualization (BELIV'08)*. CHI'08. (Florence, Italy). URL: <http://www.dis.uniroma1.it/beliv08/pospap/andrews.pdf> (besucht am 21.08.2015) (siehe S. 13, 14).
- Andrews, Keith (2015). „Human–Computer Interaction“ *Lecture Notes*. URL: <http://courses.iicm.tugraz.at/hci/hci.pdf> (besucht am 21.08.2015) (siehe S. 5, 7–13, 15–19, 21).
- Bangor, Aaron, Philip T. Kortum und James T. Miller (2008). „An Empirical Evaluation of the System Usability Scale“. In: *International Journal of Human–Computer Interaction* Bd. 24, Nr. 6, S. 574–594. DOI: 10.1080/10447310802205776 (siehe S. 29, 64).
- Barnum, Carol M. (2010). *Usability Testing Essentials: Ready, Set... Test!* 1. Aufl. Morgan Kaufmann. ISBN: 9780123750921 (siehe S. 27, 61).
- Berkun, Scott (2003). *#27 – The art of usability benchmarking*. URL: <http://scottberkun.com/essays/27-the-art-of-usability-benchmarking/> (besucht am 21.08.2015) (siehe S. 12, 65).
- Bojko, Agnieszka (2006). „Using Eye Tracking to Compare Web Page Designs: A Case Study“. In: *Journal of Usability Studies* Bd. 1, Nr. 3, S. 112–120. URL: http://uxpajournal.org/wp-content/uploads/pdf/bojko_eye_tracking.pdf (besucht am 21.08.2015) (siehe S. 39, 70).
- Brooke, John (1996). „SUS: A 'Quick and Dirty' Usability Scale“. In: *Usability Evaluation in Industry*. Hrsg. von Patrick W. Jordan, B. Thomas, Bernard Weerdmeester und Ian Lyall McClelland. London, UK: Taylor & Francis, S. 189–194. ISBN: 0748403140, 0748404600 (siehe S. 27, 28, 64).

- Chynał, Piotr, Jerzy M. Szymański und Janusz Sobiecki (2012). „Using Eyetracking in a Mobile Applications Usability Testing“. In: *Intelligent Information and Database Systems: Proceedings of the 4th Asian Conference on Intelligent Information and Database Systems, Part III*. ACIIDS'12. (Kaohsiung, Taiwan). Hrsg. von Jeng-Shyang Pan, Shyi-Ming Chen und Ngoc Thanh Nguyen. Bd. 7198. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer-Verlag, S. 178–186. ISBN: 9783642284922. DOI: 10.1007/978-3-642-28493-9_20 (siehe S. 43).
- Çöltekin, Arzu, Benedikt Heil, Simone Garlandini und Sara Irina Fabrikant (2009). „Evaluating the Effectiveness of Interactive Map Interface Designs: A Case Study Integrating Usability Metrics with Eye-Movement Analysis“. In: *Cartography and Geographic Information Science* Bd. 36, Nr. 1, S. 5–17. DOI: 10.1559/152304009787340197 (siehe S. 42).
- DIN EN ISO 9241-11 (1999). *Ergonomische Anforderungen für Bürotätigkeiten mit Bildschirmgeräten - Teil 11: Anforderungen an die Gebrauchstauglichkeit; Leitsätze (ISO 9241-11:1998); Deutsche Fassung EN ISO 9241-11:1998; Ausgabedatum: 1999-01*. Berlin: Beuth-Verlag (siehe S. 6–8).
- DIN EN ISO 9241-210 (2011). *Ergonomie der Mensch-System-Interaktion - Teil 210: Prozess zur Gestaltung gebrauchstauglicher interaktiver Systeme (ISO 9241-210:2010); Deutsche Fassung EN ISO 9241-210:2010; Ausgabedatum: 2011-01*. Berlin: Beuth-Verlag (siehe S. 9).
- Duchowski, Andrew (2007). *Eye Tracking Methodology: Theory and Practice*. 2. Aufl. London: Springer-Verlag. ISBN: 9781846286087. DOI: 10.1007/978-1-84628-609-4 (siehe S. 36, 37).
- Gibson, James J. (1979). *The Ecological Approach to Visual Perception*. Boston, MA, USA: Houghton Mifflin (siehe S. 46).
- Goldberg, Joseph H. und Xerxes P. Kotval (1999). „Computer Interface Evaluation using Eye Movements: Methods and Constructs“. In: *International Journal of Industrial Ergonomics* Bd. 24, Nr. 6, S. 631–645. ISSN: 0169-8141. DOI: 10.1016/S0169-8141(98)00068-7 (siehe S. 36, 39, 70, 112).
- Hassenzahl, Marc (2001). „The Effect of Perceived Hedonic Quality on Product Appealingness“. In: *International Journal of Human-Computer Interaction* Bd. 13, Nr. 4, S. 481–499. DOI: 10.1207/S15327590IJHC1304_07 (siehe S. 30).
- Hassenzahl, Marc, Michael Burmester und Franz Koller (2003). „AttrakDiff: Ein Fragebogen zur Messung wahrgenommener hedonischer und pragmatischer Qua-

- lität“. In: *Mensch & Computer 2003: Interaktion in Bewegung*. (Stuttgart, DE). Hrsg. von Gerd Szwillus und Jürgen Ziegler. Bd. 57. Berichte des German Chapter of the ACM. Stuttgart: B. G. Teubner, S. 187–196. ISBN: 9783519004417. DOI: 10.1007/978-3-322-80058-9_19 (siehe S. 29).
- He, Jibo, Christina Siu, Barbara Chaparro und Jonathan Strohl (2014). „Mobile“. In: *Eye Tracking in User Experience Design*. Hrsg. von Jennifer Romano Bergstrom und Andrew Jonathan Schall. Boston: Morgan Kaufmann, S. 255–290. ISBN: 978-0-12-408138-3. DOI: 10.1016/B978-0-12-408138-3.00010-8 (siehe S. 41, 42, 44).
- Holmqvist, Kenneth, Marcus Nyström, Richard Andersson, Richard Dewhurst, Halszka Jarodzka und Joost Van de Weijer (2011). *Eye Tracking: A Comprehensive Guide to Methods and Measures*. New York, NY, USA: Oxford University Press. ISBN: 9780199697083 (siehe S. 35–38, 100).
- Laugwitz, Bettina, Theo Held und Martin Schrepp (2008). „Construction and Evaluation of a User Experience Questionnaire“. In: *HCI and Usability for Education and Work: Proceedings of the 4th Symposium of the Workgroup Human-Computer Interaction and Usability Engineering of the Austrian Computer Society*. USAB’08. (Graz, Austria). Hrsg. von Andreas Holzinger. Bd. 5298. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer-Verlag, S. 63–76. ISBN: 9783540893493. DOI: 10.1007/978-3-540-89350-9_6 (siehe S. 31).
- Laugwitz, Bettina, Martin Schrepp und Theo Held (2006). „Konstruktion eines Fragebogens zur Messung der User Experience von Softwareprodukten“. In: *Mensch und Computer 2006: Mensch und Computer im Strukturwandel*. (Gelsenkirchen, DE). Hrsg. von Andreas M. Heinecke und Hansjürgen Paul. München: Oldenbourg, S. 125–134. ISBN: 3486581295 (siehe S. 29–31, 63).
- Lewis, James R. und Jeff Sauro (2009). „The Factor Structure of the System Usability Scale“. In: *Human Centered Design: Proceedings of the 1st International Conference on Human Centered Design: Held As Part of HCI International 2009*. (San Diego, CA, USA). Hrsg. von Masaaki Kurosu. Bd. 5619. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer-Verlag, S. 94–103. ISBN: 9783642028052. DOI: 10.1007/978-3-642-02806-9_12 (siehe S. 27).
- Mayhew, Deborah J. (1999). *The Usability Engineering Lifecycle: A Practitioner’s Handbook for User Interface Design*. San Francisco, CA, USA: Morgan Kaufmann. ISBN: 9781558605619 (siehe S. 5, 6).

- Nielsen, Jakob (1993). *Usability Engineering*. San Diego, CA, USA: Academic Press. ISBN: 0125184069 (siehe S. 7–9, 21, 22, 61).
- Norman, Donald A. (1986). „Cognitive Engineering“. In: *User Centered System Design; New Perspectives on Human-Computer Interaction*. Hrsg. von Donald A. Norman und Stephen W. Draper. Hillsdale, NJ, USA: L. Erlbaum Associates Inc., S. 31–65. ISBN: 0898598729 (siehe S. 50).
- Norman, Donald A. (1988). *The Psychology of Everyday Things*. United States: Basic Books. ISBN: 9780385267748 (siehe S. 45, 46, 49).
- Norman, Donald A. (1999). „Affordance, Conventions, and Design“. In: *interactions* Bd. 6, Nr. 3, S. 38–43. ISSN: 10725520. DOI: 10.1145/301153.301168 (siehe S. 45, 46, 49).
- Norman, Donald A. (2008). „THE WAY I SEE IT: Signifiers, not affordances“. In: *interactions* Bd. 15, Nr. 6, S. 18–19. ISSN: 10725520. DOI: 10.1145/1409040.1409044. URL: http://jnd.org/dn.mss/signifiers_not_affordances.html (besucht am 21.08.2015) (siehe S. 46, 47).
- Norman, Donald A. (2013). *The Design of Everyday Things*. Revised and Expanded Edition. United States: Basic Books. ISBN: 9780465050659 (siehe S. 45–53).
- Olmsted-Hawala, Erica, Temika Holland und Victor Quach (2014). „Usability Testing“. In: *Eye Tracking in User Experience Design*. Hrsg. von Jennifer Romano Bergstrom und Andrew Jonathan Schall. Boston: Morgan Kaufmann, S. 49–80. ISBN: 978-0-12-408138-3. DOI: 10.1016/B978-0-12-408138-3.00003-0 (siehe S. 38–40, 70).
- Paletta, Lucas, Helmut Neuschmied, Michael Schwarz, Gerald Lodron, Martin Pszeida, Stefan Ladstätter und Patrick Luley (2014). „Smartphone Eye Tracking Toolbox: Accurate Gaze Recovery on Mobile Displays“. In: *Proceedings of the Symposium on Eye Tracking Research and Applications*. ETRA’14. (Safety Harbor, Florida, USA). New York, NY, USA: ACM, S. 367–368. ISBN: 9781450327510. DOI: 10.1145/2578153.2628813 (siehe S. 70, 101).
- Perko, Roland, Michael Schwarz und Lucas Paletta (2014). „Aggregated Mapping of Driver Attention from Matched Optical Flow“. In: *2014 IEEE International Conference on Image Processing (ICIP)*. ICIP 2014. (Paris, France), S. 214–218. DOI: 10.1109/ICIP.2014.7025042 (siehe S. 100).

- Pernice, Kara und Jakob Nielsen (2015). *How to Conduct Eyetracking Studies*. report. Nielsen Norman Group. URL: <http://www.nngroup.com/reports/how-to-conduct-eyetracking-studies/> (besucht am 21.08.2015) (siehe S. 38).
- Poole, Alex und Linden J. Ball (2006). „Eye Tracking in Human-Computer Interaction and Usability Research: Current Status and Future Prospects“. In: *Encyclopedia of Human-Computer Interaction*. Pennsylvania: Idea Group, Inc. Hrsg. von Claude Ghaoui. IGI Global, S. 211–219. DOI: 10.4018/978-1-59140-562-7.ch034. URL: <http://alexpoole.info/blog/eyetracking-metrics/> (besucht am 21.08.2015) (siehe S. 39, 70).
- Rampl, Hansjörg (2007). *Handbuch Usability: Definition Usability*. URL: <http://www.handbuch-usability.de/begriffsdefinition.html> (besucht am 21.08.2015) (siehe S. 5).
- Rauer, Matthias (2011). *Quantitative Usability-Analysen mit der System Usability Scale (SUS)*. URL: <https://blog.seibert-media.net/blog/2011/04/11/usability-analysen-system-usability-scale-sus/> (besucht am 21.08.2015) (siehe S. 28, 64).
- Richter, Michael und Markus D. Flückiger (2013). *Usability Engineering kompakt: Benutzbare Produkte gezielt entwickeln*. 3. Aufl. IT kompakt. Berlin, Heidelberg: Springer-Verlag. ISBN: 9783642348310. DOI: 10.1007/978-3-642-34832-7 (siehe S. 5).
- Romano Bergstrom, Jennifer C., Erica L. Olmsted-Hawala und Matt E. Jans (2013). „Age-Related Differences in Eye Tracking and Usability Performance: Website Usability for Older Adults“. In: *International Journal of Human-Computer Interaction* Bd. 29, Nr. 8, S. 541–548. DOI: 10.1080/10447318.2012.728493 (siehe S. 43).
- Santner, Katrin, Gerald Fritz, Lucas Paletta und Heinz Mayer (2013). „Visual recovery of saliency maps from human attention in 3D environments“. In: *2013 IEEE International Conference on Robotics and Automation*. ICRA 2013. (Karlsruhe), S. 4297–4303. DOI: 10.1109/ICRA.2013.6631185 (siehe S. 75, 76).
- Sarodnick, Florian und Henning Brau (2011). *Methoden der Usability Evaluation: Wissenschaftliche Grundlagen und praktische Anwendung*. 2. Aufl. Huber: Wirtschaftspsychologie in Anwendung. Bern, Schweiz: Verlag Hans Huber. ISBN: 9783456848839 (siehe S. 6, 9, 10, 30, 64).

- Schrepp, Martin, Siegfried Olschner und Ulf Schubert (2013). „User Experience Questionnaire Benchmark: Praxiserfahrungen zum Einsatz im Business-Umfeld“. In: *Usability Professionals 2013*. Hrsg. von Henning Brau, Andreas Lehmann, Kostanja Petrovic und Matthias C. Schroeder. Stuttgart: German UPA e.V., S. 348–353 (siehe S. 29–31, 64).
- Tullis, Thomas und William Albert (2013). *Measuring the User Experience: Collecting, Analyzing, and Presenting Usability Metrics*. 2. Aufl. Amsterdam: Morgan Kaufmann. ISBN: 9780124157811 (siehe S. 9, 23–27, 29, 32–34, 85).
- Waltner, Georg, Michael Schwarz, Stefan Ladstätter, Anna Weber, Patrick Luley, Horst Bischof, Meinrad Lindschinger, Irene Schmid und Lucas Paletta (2015). „MANGO — Mobile Augmented Reality with Functional Eating Guidance and Food Awareness“. In: *First International Workshop on Multimedia Assisted Dietary Management (MADiMa2015)*. ICIAP 2015. (Genova, Italy). Im Erscheinen (siehe S. 1, 2, 56).

Anhang

Nachfolgend sind die verwendeten Unterlagen zur Studie in originaler Form zu finden.

A. Unterlage „Vorbefragung“

Projekt: MANGO (DIG.01-12.GF.006-02)

Probandencode: _____

Vielen Dank, dass Sie sich Zeit nehmen an der Eye-Tracking-Studie zur Usability einer Ernährungs-App teilzunehmen. Die Studie ist Bestandteil der Diplomarbeit von Verena Rexeis, die die Studie für Joanneum Research durchführt.

Bevor es los geht, bitten wir Sie den nachfolgenden Fragebogen auszufüllen und die Fragen spontan und ehrlich zu beantworten. Es gibt keine „richtigen“ oder „falschen“ Antworten. Für die Beantwortung werden maximal 10 Minuten benötigt.

Ihre Angaben werden vertraulich behandelt und fließen anonymisiert in die Studiauswertung ein.

Fragebogen

Einleitung Ihr Lebensstil beeinflusst Ihre Entscheidungen, Handlungen und Gewohnheiten, auch Essgewohnheiten. Haben Sie sich schon einmal gefragt, ob Ihre Ernährung auch eine positive Auswirkung auf genau Ihren Lebensstil haben kann? Jeder Mensch hat unterschiedliche Bedürfnisse, die auch individuell angepasste Ernährungsformen verlangen. Die Menschen unterscheiden sich in ihrem Charakter und auch in ihrem Lebensstil. Manche haben eine vorwiegend sitzende Tätigkeit, andere den Terminkalender mit „außer Haus“ Terminen voll, andere haben körperliche Anstrengungen oder müssen geistige Höchstleistungen erbringen. Dieses Anforderungsprofil zeigt aber auch, dass jeder von uns unterschiedliche Nährstoffe in unseren natürlichen Lebensmitteln benötigt.

Die Inhalte und Informationen zu Ihrer individuell angepassten Ernährungsform wären als Applikation (App) für Ihr Smartphone verfügbar. Neben den Informationen zu Ihrer Ernährungsform wären auch Informationen zu Lebensmitteln, wie Obst und Gemüse, verfügbar. Außerdem gibt es die Möglichkeit einer automatische Erkennung der Lebensmittel im Kameramodus, um sofort zu erfahren wie gut das Lebensmittel zu Ihrem individuellen Ernährungstyp passt.

1. Wie wahrscheinlich würden Sie diese App herunterladen und verwenden?

äußerst wahrscheinlich gar nicht wahrscheinlich

5 4 3 2 1 0

2. Warum würden Sie sich diese App herunterladen? (*Mehrfachantworten möglich*)

- aus Neugierde
- weil Sie sich gesünder ernähren wollen
- weil Sie an gesunder Ernährung interessiert sind
- weil Sie abnehmen wollen
- weil der Zugang zu einer individuellen Ernährungsweise neu für Sie ist
- weil Sie immer am neuesten Stand der Technik sind
- um mehr zu Herkunft/Regionalität der Produkte zu erfahren
- gar nicht
- sonstige Gründe:

3. Welche dieser Kategorien passt am ehesten zu Ihrem Lebensstil? (Bitte eine Antwort wählen)

- Beauty Food (innere Schönheit und Hautalterung)
- Brain Food (Konzentration, geistige Leistungsanforderungen)
- Power Food (Sport, körperliche Anstrengungen)
- Soul Food (Stress, hektischer Alltag)

4. Wären Sie bereit persönliche Daten (wie Alter, Gewicht, Geschlecht...) bekanntzugeben, um Informationen zu Ihrer individuell angepassten Ernährungsform zu erhalten?

- ja
- nein
- weiß nicht

5. Achten Sie auf eine abwechslungsreiche, ausgewogene Ernährung?

- ja
- nein
- weiß nicht

6. Erledigen Sie selbst hauptsächlich die Einkäufe in Ihrem Haushalt?

- ja
- nein

7. Wie viele Personen leben in Ihrem Haushalt?

- 1
- 2
- 3
- 4
- 5 oder mehr

8. Wie würden Sie Ihre Erfahrung mit Informationstechnologie einstufen?

- sehr gering sehr groß
- 1
 - 2
 - 3
 - 4
 - 5
 - 6
 - 7
 - 8

9. Wie würden Sie Ihre Erfahrung mit Smartphones und Tablets einstufen?

- sehr gering sehr groß
- 1
 - 2
 - 3
 - 4
 - 5
 - 6
 - 7
 - 8

10. Geschlecht

- männlich
- weiblich

11. Alter

Jahre

12. Familienstand

- ledig
- verheiratet/in Partnerschaft lebend
- verwitwet
- geschieden

13. Höchste abgeschlossene Ausbildung

- Pflichtschule
- Lehre
- Fachschule
- Matura (AHS, BHS)
- Kolleg
- Universität/Fachhochschule

14. Beruf (*Mehrfachantworten möglich*)

- arbeitslos
- Pensionist/-in
- selbständig / freiberuflich
- unselbstständig Erwerbstätig als _____
bei _____
- Student/-in: Hochschule _____,
Studienrichtung _____

Typenzuordnung nach functional eating

Bitte geben Sie max. eine Antwortmöglichkeit an und füllen Sie jede Frage aus!

15. Müssen Sie im Alltag oder Berufsleben vorwiegend geistig fit sein?

- | | | | |
|--------------------------|--------------------------|--------------------------|--------------------------|
| <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| trifft sehr zu | trifft zu | trifft kaum zu | trifft gar nicht zu |

16. Leidet Ihr äußeres Erscheinungsbild (Haut, Haare und Nägel) durch „falsche“ Ernährung oder durch Stress?

- | | | | |
|--------------------------|--------------------------|--------------------------|--------------------------|
| <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| trifft sehr zu | trifft zu | trifft kaum zu | trifft gar nicht zu |

17. Fühlen Sie sich oft gestresst?

- | | | | |
|--------------------------|--------------------------|--------------------------|--------------------------|
| <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| trifft sehr zu | trifft zu | trifft kaum zu | trifft gar nicht zu |

18. Wie viele Tage in der Woche betreiben Sie Ausdauer- und/oder Kraftsport (mind. 30 Minuten)?

- | | | | |
|--------------------------|--------------------------|--------------------------|--------------------------|
| <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| >4 Tage | 3-4 Tage | 1-2 Tage | 0 Tage |

19. Ist Ihnen das Wochenende zu kurz, um sich von der Arbeitswoche zu erholen bzw. zu entspannen?

trifft sehr zu trifft zu trifft kaum zu trifft gar nicht zu

20. Ist Ihr Arbeitsalltag mit hohen geistigen Anforderungen (Konzentration) verbunden?

trifft sehr zu trifft zu trifft kaum zu trifft gar nicht zu

21. Ernährung und Flüssigkeit haben Einfluss auf den Alterungsprozess?

trifft sehr zu trifft zu trifft kaum zu trifft gar nicht zu

22. Regelmäßig wöchentlich sportliche Aktivitäten haben für Sie Priorität?

trifft sehr zu trifft zu trifft kaum zu trifft gar nicht zu

23. Ablenkungen durch Telefon, Gespräche, Emails usw. erfordern eine höhere Konzentration bei Ihrer Arbeit?

trifft sehr zu trifft zu trifft kaum zu trifft gar nicht zu

24. „Vital – leistungsfähig – schön“ ist für Sie wichtig?

trifft sehr zu trifft zu trifft kaum zu trifft gar nicht zu

25. Ist der Leistungsdruck an Ihrem Arbeitsplatz für Sie kaum zu bewältigen?

trifft sehr zu trifft zu trifft kaum zu trifft gar nicht zu

26. Wird von Ihnen Genauigkeit/Pünktlichkeit und/oder Kreativität im Beruf verlangt?

trifft sehr zu trifft zu trifft kaum zu trifft gar nicht zu

27. Zu welcher Kategorie zählen Sie Ihre täglichen körperlichen Anstrengungen?

- körperlich anstrengende berufliche Tätigkeit, zB Bauarbeiter, Landwirt, Waldarbeiter, Bergarbeiter, Leistungssportler
- überwiegend gehende und stehende Tätigkeit zB Hausfrauen, Verkäufer, Kellner, Mechaniker, Handwerker
- sitzende Tätigkeit, zeitweilig auch zusätzlicher Energieaufwand für gehende und stehende Tätigkeiten, zB Kraftfahrer, Studierende, Fließbandarbeiter
- ausschließlich sitzende Tätigkeit mit wenig oder keiner anstrengenden Freizeitaktivität, zB Büroangestellte, Feinmechaniker

28. Investieren Sie viel Zeit in Ihr äußeres Erscheinungsbild und sind regelmäßige Beauty-Maßnahmen (Kosmetikprodukte, Maniküre, Pediküre, Frisör. . .) ein wichtiger Teil in Ihrem Leben?

trifft sehr zu trifft zu trifft kaum zu trifft gar nicht zu

29. Wenn Sie nicht zu einer sportlichen Aktivität kommen werden Sie dann „unruhig“ bzw. fehlt Ihnen etwas?

trifft sehr zu trifft zu trifft kaum zu trifft gar nicht zu

30. Haben Sie Probleme mehrere Dinge gleichzeitig zu erledigen (Beruf, Familie, Haushalt, Hobbies. . .)?

trifft sehr zu trifft zu trifft kaum zu trifft gar nicht zu

31. Altersflecken, brüchige Nägeln oder übermäßiger Haarausfall sind für Sie ein großes Problem?

trifft sehr zu trifft zu trifft kaum zu trifft gar nicht zu

32. Fällt es Ihnen schwer sich Dinge zu merken ohne diese aufzuschreiben?

trifft sehr zu trifft zu trifft kaum zu trifft gar nicht zu

33. Ist für Sie körperliche Fitness im täglichen Leben sehr wichtig?

trifft sehr zu trifft zu trifft kaum zu trifft gar nicht zu

34. Verursachen Stresssituationen bei Ihnen oftmals Schlafstörungen, Kopfschmerzen, Magenbeschwerden oder auch Gemütsschwankungen?

trifft sehr zu trifft zu trifft kaum zu trifft gar nicht zu

Vielen Dank!

Weiter geht es mit dem Eye-Tracking-Teil, bei dem Sie auch Ihren functional eating-Typ erfahren.

B. Unterlage „Aufgaben der Eye Tracking Studie“

Projekt: MANGO (DIG.01-12.GF.006-02)

Probandencode: _____

Aufgaben der Eye Tracking Studie

Sie bekommen einzelne Szenarien von mir beschrieben und jeweils Aufgaben dazu genannt. Bei keiner Aufgabe gibt es ein Zeitlimit.

Szenario 1

Im Supermarkt wird aktuell eine App für Smartphones beworben, die helfen soll, dass Sie sich passend zu Ihrem Lebensumfeld ernähren und so ihr volles Potential, ihre volle Leistung ausschöpfen können. Sie haben sich die App heruntergeladen und installiert.

Ihr Einstufungstest hat ergeben, dass Sie nach functional eating am ehesten dem Typ Beauty Food/Brain Food/Power Food/Soul Food entsprechen. In der App wurde Ihr Ernährungstyp bereits eingestellt und gespeichert.

Aufgabe 1 Verwenden Sie die App und lesen sich Ihre Typenbeschreibung, also Beauty Food/Brain Food/Power Food/Soul Food, durch!

Szenario 2

In der Werbung zur App haben Sie weiters gelesen, dass man mit der App Obst oder Gemüse scannen kann, um dann weitere Informationen zu erhalten. Sie wollen nun diese Funktion ausprobieren.

Aufgabe 2 Probieren Sie die Funktion „Lebensmittel scannen“ bei drei Lebensmitteln aus und lernen Sie die App kennen!

Szenario 3

Sie wollen nun erfahren, welche Lebensmittel am besten Ihren Ernährungstyp und somit Lebensstil unterstützen. Wie bereits gesagt wurde in der App Ihr Ernährungstyp bereits eingestellt und gespeichert.

Aufgabe 3 Finden Sie heraus welches der aufgebauten Lebensmittel am besten zu Ihrem Lebensstil passt. Nennen Sie Ihr Ergebnis wenn Sie fertig sind!

1.

Aufgabe 4 Finden sie die drei wichtigsten Gemüsesorten ihres Ernährungstypes Beauty Food/Brain Food/Power Food/Soul Food. Nennen Sie Ihr Ergebnis wenn Sie fertig sind!

1.

2.

3.

C. Unterlage „App-Bewertung“ (Bewertungsskalen)

Projekt: MANGO (DIG.01-12.GF.006-02)

Probandencode: _____

App-Beurteilung

Um die App zu bewerten, füllen Sie bitte den Fragebogen auf der nächsten Seite aus. Er besteht aus Gegensatzpaaren von Eigenschaften, die das Produkt haben kann. Abstufungen zwischen den Gegensätzen sind durch Kreise dargestellt. Durch Ankreuzen eines dieser Kreise können Sie Ihre Zustimmung zu einem Begriff äußern.

Beispiel:

attraktiv ○ ⊗ ○ ○ ○ ○ ○ unattraktiv

Mit dieser Beurteilung sagen Sie aus, dass Sie das Produkt eher attraktiv als unattraktiv einschätzen.

Entscheiden Sie möglichst spontan. Es ist wichtig, dass Sie nicht lange über die Begriffe nachdenken, damit Ihre unmittelbare Einschätzung zum Tragen kommt.

Bitte kreuzen Sie immer eine Antwort an, auch wenn Sie bei der Einschätzung zu einem Begriffspaar unsicher sind oder finden, dass es nicht so gut zum Produkt passt.

Es gibt keine „richtige“ oder „falsche“ Antwort. Ihre persönliche Meinung zählt!

C. Unterlage „App-Bewertung“ (Bewertungsskalen)

1. Bitte geben Sie nun Ihre Einschätzung der App ab. Kreuzen Sie bitte nur einen Kreis pro Zeile an.

	1	2	3	4	5	6	7	
unerfreulich	<input type="radio"/>	erfreulich						
unverständlich	<input type="radio"/>	verständlich						
kreativ	<input type="radio"/>	phantasielos						
leicht zu lernen	<input type="radio"/>	schwer zu lernen						
wertvoll	<input type="radio"/>	minderwertig						
langweilig	<input type="radio"/>	spannend						
uninteressant	<input type="radio"/>	interessant						
unberechenbar	<input type="radio"/>	voraussagbar						
schnell	<input type="radio"/>	langsam						
originell	<input type="radio"/>	konventionell						
behindernd	<input type="radio"/>	unterstützend						
gut	<input type="radio"/>	schlecht						
kompliziert	<input type="radio"/>	einfach						
abstoßend	<input type="radio"/>	anziehend						
herkömmlich	<input type="radio"/>	neuartig						
unangenehm	<input type="radio"/>	angenehm						
sicher	<input type="radio"/>	unsicher						
aktivierend	<input type="radio"/>	einschläfernd						
erwartungskonform	<input type="radio"/>	nicht erwartungskonform						
ineffizient	<input type="radio"/>	effizient						
übersichtlich	<input type="radio"/>	verwirrend						
unpragmatisch	<input type="radio"/>	pragmatisch						
aufgeräumt	<input type="radio"/>	überladen						
attraktiv	<input type="radio"/>	unattraktiv						
sympathisch	<input type="radio"/>	unsympathisch						
konservativ	<input type="radio"/>	innovativ						

2. Wie sehr stimmen Sie diesen Aussagen zu?

	stimme gar nicht zu				stimme voll zu
Ich kann mir sehr gut vorstellen, die App regelmäßig zu nutzen	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Ich empfinde die App als unnötig komplex	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Ich empfinde die App als einfach zu nutzen	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Ich denke, dass ich technischen Support brauchen würde, um die App zu nutzen	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Ich finde, dass die verschiedenen Funktionen der App gut integriert sind	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Ich finde, dass es in der App zu viele Inkonsistenzen gibt	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Ich kann mir vorstellen, dass die meisten Leute die App schnell zu beherrschen lernen	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Ich empfinde die Bedienung als sehr umständlich	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Ich habe mich bei der Nutzung der App sehr sicher gefühlt	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Ich musste eine Menge Dinge lernen, bevor ich mit der App arbeiten konnte	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

3. Wie wahrscheinlich würden Sie diese App herunterladen und verwenden?

äußerst wahrscheinlich												gar nicht wahrscheinlich
	<input type="checkbox"/>											
	5	4	3	2	1	0						

Big Five Persönlichkeitstest

Auf den folgenden Seiten finden Sie Listen von Aussagen, die mehr oder weniger auf Sie zutreffen können. Es gibt keine „richtigen“ oder „falschen“ Antworten. Bitte geben Sie an, wie sehr die einzelnen Aussagen auf Sie zutreffen. Bitte beschreiben Sie sich im Hier und Jetzt und nicht so, wie Sie sich gerne künftig sehen würden.

Dafür stehen Ihnen immer die folgenden Antwortmöglichkeiten zur Verfügung:

- 1: Sehr unzutreffend
- 2: Eher unzutreffend
- 3: Weder zutreffend noch unzutreffend
- 4: Eher zutreffend
- 5: Sehr zutreffend

Bitte antworten Sie auf jede Aussage, auch wenn Sie sich Ihrer Antwort nicht ganz sicher sind.

4. Wie sehr treffen diese Aussagen auf Sie zu?

	Sehr un- zutreffend		neutral		Sehr zu- treffend
Ich bin eine Stimmungskanone.	<input type="checkbox"/>				
Ich kümmere mich kaum um andere.	<input type="checkbox"/>				
Ich bin stets vorbereitet.	<input type="checkbox"/>				
Ich lasse mich leicht stressen.	<input type="checkbox"/>				
Ich habe ein großes Vokabular.	<input type="checkbox"/>				
Ich spreche nicht viel.	<input type="checkbox"/>				
Ich bin an Leuten interessiert.	<input type="checkbox"/>				
Ich lasse meine Sachen herumliegen.	<input type="checkbox"/>				
Die meiste Zeit bin ich entspannt.	<input type="checkbox"/>				
Mir fällt es schwer, abstrakte Ideen zu verstehen.	<input type="checkbox"/>				

	Sehr un- zutreffend		neutral		Sehr zu- treffend
Ich fühle mich in Gesellschaft anderer wohl.	<input type="checkbox"/>				
Ich beleidige Leute.	<input type="checkbox"/>				
Ich achte auf Kleinigkeiten.	<input type="checkbox"/>				
Ich Sorge mich über Dinge.	<input type="checkbox"/>				
Ich verfüge über eine lebhaft Phantasie.	<input type="checkbox"/>				
Ich halte mich eher im Hintergrund.	<input type="checkbox"/>				
Ich kann die Gefühle anderer nachempfinden.	<input type="checkbox"/>				
Ich hinterlasse ein Durcheinander.	<input type="checkbox"/>				
Nur selten fühle ich mich traurig.	<input type="checkbox"/>				
Ich bin an abstrakten Ideen nicht interessiert.	<input type="checkbox"/>				
Ich beginne Unterhaltungen.	<input type="checkbox"/>				
Ich bin nicht an den Problemen anderer Leute interessiert.	<input type="checkbox"/>				
Ich erledige Routineaufgaben sofort.	<input type="checkbox"/>				
Ich bin leicht zu beunruhigen.	<input type="checkbox"/>				
Ich habe großartige Ideen.	<input type="checkbox"/>				
Ich habe meist nur wenig zu sagen.	<input type="checkbox"/>				
Ich habe ein weiches Herz.	<input type="checkbox"/>				
Ich vergesse es oft, Dinge an ihren Platz zurückzustellen.	<input type="checkbox"/>				
Mich kann man leicht aufregen.	<input type="checkbox"/>				
Ich habe keine lebhaft Phantasie.	<input type="checkbox"/>				

C. Unterlage „App-Bewertung“ (Bewertungsskalen)

	Sehr un- zutreffend		neutral		Sehr zu- treffend
Auf Parties spreche ich mit vielen verschiedenen Leuten.	<input type="checkbox"/>				
Ich bin an anderen nicht wirklich interessiert.	<input type="checkbox"/>				
Ich mag Ordnung.	<input type="checkbox"/>				
Meine Laune ändert sich häufig.	<input type="checkbox"/>				
Ich verstehe Dinge schnell.	<input type="checkbox"/>				
Ich mag es nicht, Aufmerksamkeit auf mich zu ziehen.	<input type="checkbox"/>				
Ich nehme mir Zeit für andere.	<input type="checkbox"/>				
Ich drücke mich um die Erfüllung meiner Pflichten.	<input type="checkbox"/>				
Ich habe häufig Stimmungsschwankungen.	<input type="checkbox"/>				
Ich verwende schierige Wörter.	<input type="checkbox"/>				
Es macht mir nichts aus im Mittelpunkt zu stehen.	<input type="checkbox"/>				
Ich kann die Gefühle anderer nachempfinden.	<input type="checkbox"/>				
Ich folge einem Terminplan.	<input type="checkbox"/>				
Ich werde schnell ärgerlich.	<input type="checkbox"/>				
Ich verbringe Zeit damit Dinge zu reflektieren.	<input type="checkbox"/>				
Ich bin still unter Fremden.	<input type="checkbox"/>				
Leute fühlen sich mit mir wohl.	<input type="checkbox"/>				
Ich bin sehr genau bei meiner Arbeit.	<input type="checkbox"/>				
Oft fühle ich mich traurig.	<input type="checkbox"/>				
Ich bin ideenreich.	<input type="checkbox"/>				

Sie haben nun alle Teile der Studie abgeschlossen! Wir bedanken uns herzlich für Ihre Teilnahme und hoffen, dass die Zeit auch für Sie gewinnbringend war!

D. Zustimmungserklärung



ZUSTIMMUNGSERKLÄRUNG

Projekt:

MANGO (DIG.01-12.GF.006-02)

**ZUSTIMMUNGSERKLÄRUNG für die Teilnahme an der
Eye-Tracking Studie zur Usability einer Ernährungsapp**

Durch Unterzeichnung des vorliegenden Formulars erkläre ich mich damit einverstanden, dass meine persönlichen Daten (Name, Alter, Geschlecht), die biometrischen Daten des *Blickverhaltens*, sowie *Bildmaterial (Foto oder Video)*, vom Datenhalter **JOANNEUM RESEARCH Forschungsgesellschaft mbH, Leonhardstraße 59, 8010 Graz, Österreich**, gemäß den geltenden Gesetzen und den genannten Bestimmungen erhoben, verarbeitet und verwendet werden dürfen.

Dieses Projekt (MANGO, DIG.01-12.GF.006-02) ist eine Eye-Tracking Studie zur Usability einer Ernährungsapp, die zur Klassifizierung von Lebensmitteln verwendet werden kann.

Auf jeden Fall werden bei der Durchführung dieser Studie vom österreichischen Datenschutzbeauftragten die entsprechenden organisatorischen Sicherheits- und Datenschutzmaßnahmen eingehalten¹.

Joanneum Research garantiert, dass *die Aufnahmen* ausschließlich zu Zwecken der „Wissenschaftlichen Forschung und Statistik“ im Sinne des aktuellen Datenschutzgesetzes (§46 DSGVO 2000) und damit eindeutig *nicht personenbezogen verwendet* werden.

Die Daten werden vollständig im Sinne des Paragraphen §46 DSGVO 2000 des österreichischen Datenschutzgesetzes gehandhabt: „Die Verwendbarkeit im Rahmen Wissenschaftlicher Forschung und Statistik“ ist erlaubt. Für Zwecke wissenschaftlicher oder statistischer Untersuchungen, die keine personenbezogenen Ergebnisse zum Ziel haben, darf der Auftraggeber der Untersuchung alle Daten verwenden, die für den Auftraggeber nur indirekt personenbezogen sind.“

Im Besonderen wird folglich hiermit informiert, dass die Daten anonymisiert für statistische Zwecke ausgewertet werden. Im Falle einer Verwendung biometrischer Daten, sowie Bildmaterial (Foto oder Video) in wissenschaftlichen Zeitschriften oder auf Konferenzen erfolgen keine Angaben zur Identität der Person.

Ich bin berechtigt, Zugang zu meinen Daten zu verlangen, und meine persönlichen Daten in Übereinstimmung mit den geltenden Gesetzen zu korrigieren und, falls erforderlich, zu löschen. Zu diesem Zweck kann ich den Datenhalter, das Projektkonsortium, kontaktieren (derzeitige Kontaktperson: **Michael Schwarz, Tel +43 316 876-1727**, (bitte beachten Sie, dass der Datenhalter andere Kontaktpersonen bestimmen kann)). Ich kann den Datenhalter

¹ siehe auch europäische Richtlinien unter http://ec.europa.eu/justice_home/fsi/privacy/nationalcomm/index_de.htm

auffordern, meine persönlichen Daten, die ungenau, unvollständig, zweifelhaft oder veraltet sind oder deren Erhebung, Nutzung, Veröffentlichung oder Speicherung untersagt ist, je nach Sachlage richtig zu stellen, zu vervollständigen, zu aktualisieren, zu sperren oder zu löschen. Auf meine Aufforderung hin hat der Datenhalter für mich kostenfrei zu begründen, dass er die erforderlichen Schritte gesetzt hat.

Wird dieses Recht ausgeübt, so muss jede weitergeleitete Kopie der biometrischen und persönlichen Daten korrigiert, richtig gestellt oder gelöscht werden.

Ich habe die obigen Bedingungen gelesen und nehme zur Kenntnis, dass es mir freisteht, der Verarbeitung meiner persönlichen Daten wie oben beschrieben zuzustimmen. Ich wurde über den Inhalt der Studie informiert und offene Fragen wurden beantwortet. Die Teilnahme an der Studie erfolgt freiwillig, auf eigene Gefahr und Verantwortung. Bei einer Nichtteilnahme wären mir keinerlei Nachteile erwachsen. Ich kann die Teilnahme an der Studie jederzeit ohne Angabe von Gründen beenden. Ich wurde über mögliche Risiken, die mit der Ableitung von physiologischen Maßen und dem Versuchsablauf in Verbindung stehen könnten, aufgeklärt. Ich versichere, die mir aufgetragenen Anweisungen und Verhaltensrichtlinien einzuhalten. Durch Unterzeichnung des vorliegenden Formulars erkläre ich mich mit den oben angeführten Bedingungen einverstanden.

Datum:

Name:

Unterschrift:

E. Checkliste zum Studienablauf

Projekt: MANGO (DIG.01-12.GF.006-02)

Datum: _____

Versuchsleiterin: _____

Probandencode: _____

Checkliste MANGO Usability Studie

- Checkliste, Einverständniserklärung, Pre-Fragebogen 1, Aufgabenstellung, Post-Fragebogen 2, Aufwandsentschädigung für nächste Person richten
- Lebensmittelaufbau kontrollieren
- Computer hochfahren, Handy und Affektiva Armband bereitlegen, Motion Tracker ->ET Glasses
- (iView starten) Eye-Tracking Glasses starten
- Verbindung Eye-Tracking Glasses mit Programm

Vorgehen:

- Begrüßung
- Inhalt Studie
- Einverständniserklärung
- Schriftlicher Fragebogen 1 ausfüllen lassen
 - functional eating Auswertung machen (Assistent)
- Start Eye-Tracking Teil:
 - Affektiva anlegen
 - Experiment auswählen – MANGO Usability_Rexeis
 - Experiment „lock“
 - Motion Tracker einschalten
 - Anlegen/Anpassen der Eye-Tracking Glasses (ev. höhenverstellbaren Teil fixieren)
 - Hinweis: Nicht mehr berühren
 - Stuhl umstellen
 - Eingabe Probandencode etc.
 - Kalibrierung
 - Smartphone Verbindung zu Head-Tracker herstellen (in FACTS-App)
 - Recording in iView starten
- SYNC mit Aufnahme via ET-Glasses:
 - Smartphone Screen Capturing starten
 - In FACTS-App: starten
 - 2x Klatschen für Affektiva (vorzeigen: Ruhe – 2x Klatschen – Ruhe)
 - Handy geben und Umgang beschreiben (blauer Rand sichtbar, Display sichtbar)

E. Checkliste zum Studienablauf

Projekt: MANGO (DIG.01-12.GF.006-02)

Datum: _____

Versuchsleiterin: _____

Probandencode: _____

- Validierung ETG durch User (sagt wo er/sie hinschaut)
- Laptop anlegen
- Videocam starten
- In App fe-Typ einstellen
- Mündliche Instruktionen durch Untersuchungsleiterin
- Durchführung Task
- Nach Ende: Screen Capturing gleich stoppen, FACTS stoppen (ET-Glasses unberührt lassen)
- Videocam stoppen
- VALIDIERUNG. Wieder an Schachbrettwand (User sagt wohin er/sie schaut)
- ET-Glasses + Affektiva abnehmen (Head Tracker ausschalten)
- Postbefragung: Fragebogen 2 ausfüllen lassen
- Handy und Laptop aufladen, Head tracker, Kameraakku, Affektiva aufladen
- Aufwandsentschädigung
- Frage nach Datenaufnahme in ETG-Studien Liste
- Verabschiedung
- Datensicherung
- Daten von Handy löschen
- Alle Papier-Unterlagen beschriften

Unterschrift Versuchsleiterin: _____