

Messung und Bewertung der Usability von Kommunikationsendeinrichtungen

Inhaltsverzeichnis

0	Vorbemerkungen	3
1	Einführung und Definitionen	3
1.1	Usability	4
1.1.1	Ease of Use	4
1.1.2	Joy of Use	5
1.2	Taxonomie von Qualitätsaspekten	5
2	Klassifikation von Anwendungen	7
2.1	Gültigkeitsbereich	7
2.2	Klassifikation nach Funktionsbereich	8
2.2.1	Kommunikationsanwendungen	8
2.2.2	Medienaufzeichnungs- und -abspielanwendungen	8
2.2.3	Informationsabrufanwendungen	9
2.3	Klassifikation nach Nutzungsszenarien	9
2.4	Klassifikation nach Ein- und Ausgabemodalitäten	9
3	Usability Engineering Lifecycle	11
4	Evaluationsverfahren	11
4.1	Übersicht und Einführung	11
4.2	Expertenzentrierte Verfahren	13
4.2.1	Cognitive Walkthrough (CW)	13
4.2.2	Prüflisten	14
4.2.3	Heuristische Evaluation (HE)	14
4.2.4	Modellbasierte Evaluation (ME)	15
4.3	Nutzerzentrierte Verfahren	15
4.3.1	Experiment	15
4.3.2	Feldstudien	16
4.3.3	Interviews/Fragebögen	16
4.3.4	Protokollanalysen und Lautes Denken (<i>Thinking Aloud</i>)	16
5	Praktischer Leitfaden	17
5.1	Tabelle zu Verfahren und deren Eignung	17
5.2	Ermittlung der Testaufgaben mittels Fragekatalog	18
5.2.1	Ermittlung der Hauptfunktionen	18
5.2.2	Ermittlung der Neben- und Zusatzfunktionen	18
5.3	Praktische Hinweise	19
5.3.1	Generelle Hinweise	19
5.3.2	Methodische Hinweise	19
5.3.3	Leitfragen zur Planung einer Evaluationsstudie	20
5.3.4	Der Umgang mit Versuchspersonen und ihren Daten	21

Fortsetzung Seite 2

Anhang A – Beispiel für eine Versuchsbeschreibung + Einverständniserklärung	23
Anhang B – Beschreibung von Beispielfragebögen	24
Anhang C – Aufgaben- und Prüfliste nach VDE-ITG-Richtlinie 9.4. 1.01	27
Anhang D – Instruktion und Protokoll - Heuristische Evaluation	29
Anhang E – Fallbeispiele	31
Anhang F – Liste der englischen Termini und deren Übersetzung	37
Literatur	39

Messung und Bewertung der Usability von Kommunikationsendeinrichtungen

Diese Richtlinie ist eine Aktualisierung der Richtlinie ITG 9.4.1.01 vom 31.10.1995 „Prüfung und Bewertung der Benutzerfreundlichkeit von Kommunikationsendgeräten und -systemen“.

0 Vorbemerkungen

Die im Folgenden vorgestellte Richtlinie beschreibt Verfahren zur Messung und Bewertung der *Usability* (häufig nicht ganz zutreffend mit „Gebrauchstauglichkeit“ übersetzt) von Kommunikationsendeinrichtungen. Sie richtet sich insbesondere an Praktiker (z. B. Ingenieure und Informatiker, die mit der Entwicklung und Evaluierung von Anwendungen auf solchen Kommunikationsendeinrichtungen befasst sind) und soll ihnen eine konkrete Handreichung zur Messung und Bewertung verschiedener Aspekte der Usability geben. Entsprechend liegt der Fokus dieser Richtlinie, im Gegensatz zu anderen Richtlinien, Empfehlungen oder Normen (bspw. des DIN, ISO, ETSI oder ITU-T), nicht auf der Erarbeitung von Definitionen oder der rein funktionalen Prüfung von Hardware- oder Softwarefunktionen, sondern auf der Beschreibung verfügbarer und der Auswahl passender Mess- und Bewertungsverfahren. Dabei beschränken sich die Hinweise nicht auf grafische Nutzeroberfläche (wie bspw. in DIN/ISO 9241), sondern auch Verfahren zur Bewertung multimodaler Schnittstellen werden berücksichtigt.

Hinsichtlich der Verwendung der entsprechenden englischsprachigen Fachbegriffe sei an dieser Stelle angemerkt, dass, sofern keine eindeutige und gebräuchliche deutsche Übersetzung vorlag, die Begriffe im englischen Original belassen wurden. Eine Liste der englischen Termini und deren Übersetzung findet sich im Anhang.

1 Einführung und Definitionen

Das vorliegende Verfahren soll die Bewertung der Gebrauchstauglichkeit (*Usability*) von Kommunikationsendeinrichtungen ermöglichen. Einführend wird ein Einblick in Grundlagen der *Usability*-Forschung geben. Den zweiten Teil stellt ein praktischer Leitfaden zur Prüfung von Kommunikationsendeinrichtungen dar. Konkrete Anwendungsbeispiele dazu werden in den Anhängen gegeben.

Im Folgenden werden zunächst grundlegende Begriffe zur Bewertung der Gebrauchstauglichkeit von Kommunikationsendeinrichtungen eingeführt und definiert. Anschließend werden Zusammenhänge zwischen diesen Konzepten in einer Taxonomie von Qualitätsaspekten dargestellt. Kapitel 2 dient der Klassifikation von Anwendungen, die in einer Kommunikationsendeinrichtung integriert sein können. Dabei wird eine Einteilung nach Funktionsbereichen sowie nach Ein- und Ausgabemodalitäten beschrieben. Anschließend werden in Kapitel 3 verschiedene Verfahren zur Evaluation von Kommunikationsendeinrichtungen vorgestellt. Es wird eine Unterscheidung zwischen experten- und nutzerzentrierten Verfahren getroffen. Im 4. Kapitel wird zunächst eine zusammenfassende Übersicht aller in Kapitel 3 vorgestellten Verfahren und deren Eignung für den jeweiligen Einsatz gegeben. Darauf folgt ein Fragenkatalog zur Ermittlung von Testaufgaben zur Durchführung einer Evaluation. Den Abschluss des Kapitels bilden generelle und methodische Hinweise, die den praktischen Einsatz dieser Richtlinie unterstützen sollen.

1.1 Usability

Gemäß der ISO-Norm 9241-11:1998 beschreibt Gebrauchstauglichkeit (*Usability*) „das Ausmaß, in dem ein Produkt durch bestimmte Benutzer in einem bestimmten Nutzungskontext genutzt werden kann, um bestimmte Ziele effektiv, effizient und zufriedenstellend zu erreichen.“ Obwohl die ISO-Norm wohl die wahrscheinlich am weitesten verbreitete Definition von Usability darstellt, sind in der Literatur vielfältige weitere Definitionen zu finden.

Darüber hinaus sei zu erwähnen, dass in jüngster Zeit der Begriff *User Experience (UX)*, das Nutzungserlebnis, an Popularität gewinnt. In der ISO-Norm 9241-210:2010 ist *User Experience* wie folgt definiert: „*User Experience* umfasst alle Wahrnehmungen und Reaktionen einer Person auf eine tatsächliche oder eine antizipierte Nutzung eines Produktes, eines Systems oder einer Dienstleistung.“

Nach dieser Definition lässt sich *User Experience* gemäß Bevan auf unterschiedliche Art und Weise konzeptualisieren: *User Experience* kann einerseits als Sammelbegriff für alle Wahrnehmungen und Reaktionen der Nutzer verstanden werden. Andererseits kann *User Experience* als Gegenkonzept zu Usability aufgefasst werden, da Usability historisch vorrangig auf Performanzparameter abzielte. Eine andere mögliche Betrachtungsweise versteht *User Experience* als ausführlichere Darstellung der Zufriedenheitskomponente der *Usability*. In der vorliegenden Richtlinie wird dieser Betrachtungsweise gefolgt.

Es werden im Folgenden zwei Aspekte der Gebrauchstauglichkeit betrachtet: Die interaktionsbezogenen Aspekte (*Ease of Use*) sowie affekt-bezogene, oft mit dem Begriff *User Experience* assoziierte, Aspekte (*Joy of Use*). Beide wirken auf die Akzeptanz eines Dienstes. Jeder dieser Aspekte lässt sich durch weitere Kriterien beschreiben, die im Weiteren dargestellt werden.

Folglich wird Usability hier nicht ausschließlich auf Performanzparameter bezogen und entspricht damit trotz Überschneidungen nicht dem Konzept der „*Fitness for Use*“ nach DIN 55350-11.

1.1.1 Ease of Use

Die (wahrgenommene) *Ease of Use* beschreibt inwiefern Nutzer annehmen, dass das Benutzen eines Systems frei von Anstrengung sein wird (Davis, 1989). Relevant sind dabei insbesondere die auch in der ISO-Norm 9241 zu findenden Aspekte Effektivität und Effizienz sowie Intuitivität und Erlernbarkeit.

Effektivität bezieht sich auf die „Genauigkeit und Vollständigkeit, mit der Benutzer ein bestimmtes Ziel erreichen“. Effizienz beschreibt den im Verhältnis zur Genauigkeit und Vollständigkeit der Zielerreichung eingesetzten Aufwand, mit dem Benutzer ein bestimmtes Ziel erreichen. Intuitivität ist das Ausmaß, in dem ein Benutzer mit der unbewussten Anwendung von Vorwissen effektiv mit einem System interagieren kann (Mohs et al., 2006). Erlernbarkeit meint die Leichtigkeit, mit der neue Nutzer effektive Interaktionen beginnen können und maximale Performanz erreichen (Dix et al., 2004).

Der Nutzen (*Utility*) eines Systems beschreibt, ob das System die Funktionalitäten zur Verfügung stellt, die der Nutzer zur Erreichung seiner Ziele benötigt. Die Nützlichkeit (*Usefulness*) schließlich ergibt sich aus dem Nutzen und der *Ease of Use*.

Anhand der oben vorgestellten Konzepte wird deutlich, dass *Ease of Use*, die Einfachheit der Nutzung, eine stark aufgabenorientierte, pragmatische Sichtweise darstellt. Emotionale, nicht-instrumentelle Gründe der Nutzung bzw. Nichtnutzung eines Systems werden nicht abgedeckt. Doch ist Techniknutzung (bspw. bei Diensten oder Geräten aus dem Unterhaltungsbereich) nicht immer instrumentell durch die Ausführung bestimmter Aufgaben motiviert (vgl. Hassenzahl 2006). Das nachstehend beschriebene Konzept *Joy of Use* (Spaß bei der Benutzung) umfasst genau diese affektbezogenen Kriterien.

1.1.2 Joy of Use

Joy of Use ist nach Schleicher & Trösterer (2009) das bewusst positive Erleben der Qualität der Interaktion. Ausschlaggebend ist dabei unter anderen die Ästhetik des Produktes. Dabei wird das Erleben nicht nur auf die rein visuelle Ästhetik bezogen, sondern wird als die durch sensorische Wahrnehmung vermittelte Freude verstanden (Hekkert, 2006). Dementsprechend kann Ästhetik durch alle Sinne bzw. alle Modalitäten wahrgenommen werden und ist nicht allein auf den visuellen Kanal beschränkt.

Weiterhin von Bedeutung ist die *Persönlichkeit* des Systems. Diese bezieht sich auf die Faktoren, die das Verhalten des Systems als Interaktionspartner betreffen (Möller et al., 2006). Die Persönlichkeit ist vor allem bei Systemen wichtig, die menschliches Verhalten imitieren, bspw. natürlichsprachliche Systeme oder solche, die einen Avatar als Systemausgabe verwenden. Der *Appeal* eines Systems setzt sich demnach aus der Ästhetik und der Persönlichkeit zusammen und beschreibt entsprechend das Ausmaß, zu dem das System überraschende, interessante und neuartige Eigenschaften besitzt.

Anzumerken bleibt, dass sich im aktuellen Stand der Forschung widersprüchliche Positionen zum Verhältnis instrumenteller und nicht-instrumenteller Aspekte finden. Auf der einen Seite stehen dabei die Arbeiten von Tractinsky und Kollegen (1997, 2000): „What is beautiful is usable“ – Ästhetische Systeme sind dementsprechend gebrauchtauglichere Systeme. Aufgegriffen wurde diese Sichtweise u.a. von Norman mit dem Slogan „Attractive things work better“ (Norman, 2002). Andererseits konnten einige Studien hier keinen Zusammenhang feststellen (Lindgaard, 2007; vgl. auch Hassenzahl, 2008; Hassenzahl und Monk, 2010). Als Ursache für die uneindeutige Befundlage zur Beziehung zwischen *Ease of Use* und (visueller) Ästhetik vermuten Hassenzahl und Monk (2010) uneinheitliche Auffassungen und Operationalisierungen der relevanten Konzepte wie Schönheit, Ästhetik, Güte und Attraktivität.

1.2 Taxonomie von Qualitätsaspekten

Das Zusammenspiel sämtlicher beschriebener Faktoren ist in Abb. 1 dargestellt (Möller et al., 2009). Gezeigt wird eine Taxonomie der oben genannten Qualitätsaspekte auf drei Ebenen: Der Ebene der Einflussfaktoren auf die Qualität (Qualitätsfaktoren, *Influencing Factors*), der Ebene der Beschreibung der Interaktion zwischen Nutzer und System (*Interaction Performance*), sowie der Ebene der wahrgenommenen Qualitätsaspekte. Letztere wird auch als *Quality of Experience (QoE)* bezeichnet, während die ersten beiden Ebenen verschiedene Aspekte der Leistung des Systems, zusammengefasst als *Quality of Service (QoS)*, umfassen.

In der obersten Ebene sind die Qualitätsfaktoren dargestellt. Hier wird grob zwischen den Faktoren, die der Benutzer einbringt, denen des Systems, sowie denen des Benutzungskontextes unterschieden. Bei den Benutzerfaktoren wird grob unterschieden zwischen denen, die

quasi-statisch und konstant sind (Alter, Geschlecht, Einstellung, etc.) sowie solchen, die sich von Interaktion zu Interaktion oder auch während einer Interaktion ändern können (Emotionen, Motivation, etc.). Auf der Systemseite wird neben den (technischen) *Agent Factors* auch das äußerliche Erscheinungsbild des Systems (*Appearance*) berücksichtigt. Aufgabenbezogene Aspekte (*Task Factors*) sind hier mit anderen funktionalen Aspekten (*Functional Factors*) zusammengefasst.

Auf der mittleren Ebene findet die Interaktion zwischen Benutzer und System statt. Diese ist durch zwei Kreisläufe angedeutet, bei denen jeweils alle Schritte von der Informationseingabe über die Informationsverarbeitung bis hin zur Informationsausgabe dargestellt sind. Für jeden der beteiligten Schritte können Leistungsindikatoren angegeben werden. Beim System decken sich diese häufig mit Leistungen der Systemkomponenten.

Auf der unteren Ebene werden die zuvor besprochenen Qualitätsaspekte zusammengefasst. Ausgangspunkt ist hier zunächst die Interaktionsqualität, d.h. die Beurteilung der direkten Interaktion zwischen Nutzer und System. Diese wird unterschieden in die Eingabequalität, die Ausgabequalität und die Kooperativität. Darunter sind die beiden Hauptkomponenten der *Usability* dargestellt, nämlich der *Ease of Use* und der *Joy of Use*. Rechts und Links hiervon sind wiederum die eher hedonischen bzw. eher pragmatischen Qualitätsaspekte angedeutet, wie eingangs erläutert. Auf der pragmatischen Seite leitet sich hieraus die Nützlichkeit (*Usefulness*) ab, welche zusammen mit *Usability* und den hedonischen Qualitätsaspekten die Akzeptanz (*Acceptability*) bestimmt. Letztere ist eine rein ökonomische Größe, welche die Anzahl der tatsächlichen Nutzer zur Größe der Zielgruppe in Beziehung setzt.

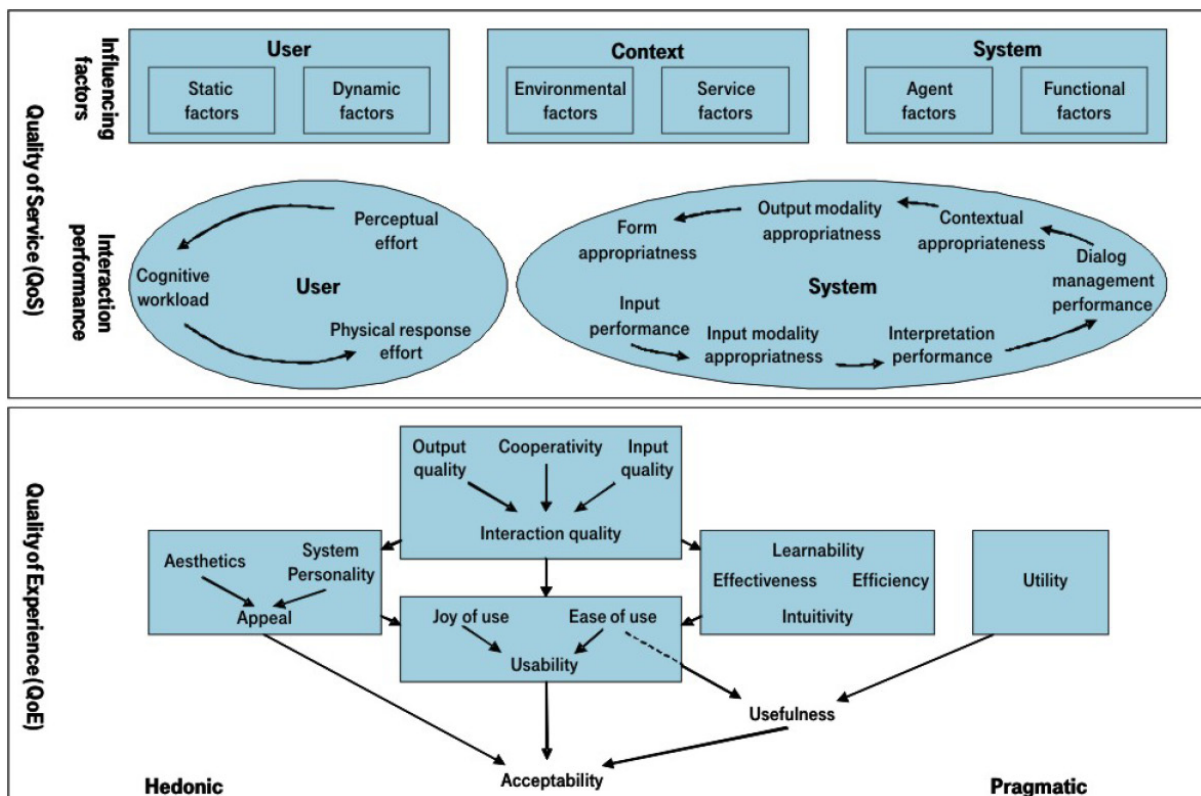


Abbildung 1. Taxonomie von Qualitätsaspekten (Möller et al., 2009). Deutsche Bezeichnungen einiger der dargestellten Aspekte finden sich im Text

2 Klassifikation von Anwendungen

Eine Vielzahl moderner Kommunikationsendeinrichtungen zeichnet sich dadurch aus, dass verschiedenste Anwendungen in einem Gerät integriert sind. Dies erschwert die standardisierte Bewertung der Qualität eines Gerätes, da bei Geräten, die im Grunde demselben Zweck dienen, häufig erhebliche Unterschiede im Funktionsumfang feststellbar sind. Mobiltelefone besitzen beispielsweise eine große Spannbreite an integrierten Anwendungen: einfache Varianten sind auf grundlegende Funktionen wie Telefonie und SMS beschränkt, umfangreichere Geräte (z. B. Smartphones) hingegen integrieren neben den Grundfunktionen ein variables Angebot an Anwendungen wie Internet-Browser, MP3-Player, E-Mail-Programme usw. Zur Bewertung der Qualität eines Gerätes sollte daher zuerst festgelegt werden, welche Gerätefunktionen überprüft werden sollen. Vor allem, wenn mehrere Geräte miteinander verglichen werden sollen, ist es notwendig vorher festzuhalten, welche Anwendungen für einen Vergleich betrachtet werden sollen. Eine Gesamtbewertung für ein Gerät kann aus der Summe der Bewertungen aller integrierten Anwendungen abgeleitet werden.

2.1 Gültigkeitsbereich

Die hier aufgeführten Methoden zur Bewertung von Kommunikationsendeinrichtungen beziehen sich nur auf die Schnittstelle zwischen Mensch und Gerät. Bei Systemen, die der zwischenmenschlichen Kommunikation über ein technisches System dienen (z. B. Telefonie), kann mit den dargestellten Verfahren nur die Bedienung des Gerätes durch einen menschlichen Benutzer bewertet werden. Die Qualität der Kommunikation selbst hängt bei derartigen Systemen in hohem Maße vom Übertragungskanal ab. Für dessen Bewertung stehen gesonderte Verfahren zur Verfügung (vgl. z. B. die ITU-T-Empfehlungen der P-Serie), die in Rahmen der vorliegenden Richtlinie nicht erörtert werden. Ähnliches gilt für Anwendungen, die dem Abruf von Informationen dienen. Bei der Übertragung von Daten über Netzwerke können unerwünschte Effekte (z. B. Verzögerungen und Paketverluste) auftreten, die durch die dargestellten Methoden nicht berücksichtigt werden.

Weiterhin stehen adaptive und kontextsensitive Systeme sowie intelligente Umgebungen nicht im Mittelpunkt dieser Richtlinie. Zwar mögen viele der Empfehlungen auch für die genannten Systeme Gültigkeit besitzen, doch wird eine erschöpfende Evaluierung mittels der hier vorgestellten Methoden nicht möglich sein.

Eine weitere Einschränkung bezieht sich auf die vorgestellten Methoden: Der Großteil der gegenwärtig verfügbaren Methoden ist in erster Linie für aufgabenorientierte Bewertungen geeignet. Erlebnisorientierte Bewertungsverfahren befinden sich in der Entwicklung, standardisierte Verfahren gibt es bisher jedoch kaum. Der Schwerpunkt der Richtlinie liegt daher auf den klassisch-aufgabenorientierten Ansätzen.

In vielen Kommunikationsendeinrichtungen sind neben den gerätespezifischen Haupt- und Nebenfunktionen auch Zusatzfunktionen wie Spiele integriert. Solche nicht gerätespezifischen Funktionen, die in Komplexität und Umfang beliebig variabel sein können, werden durch das Bewertungsverfahren ebenfalls nicht abgedeckt.

2.2 Klassifikation nach Funktionsbereich

Da moderne Kommunikationsendeinrichtungen zumeist unterschiedliche Anwendungen integrieren und die gleiche Anwendung auch auf verschiedenen Endgeräten verfügbar sein kann, ist es notwendig im Vorfeld die zu bewertenden Anwendungen festzulegen. Die hier beschriebene Klassifikation von Anwendungen nach Funktionsbereich bildet einen Ausgangspunkt zum späteren Festlegen von Testaufgaben. Für alle Anwendungen gibt es typische Haupt- und Nebenfunktionen, die im Folgenden kurz beschrieben werden. Aufgrund der stets wachsenden Vielfalt an Funktionen kann diese Beschreibung jedoch nicht vollständig sein. Generell ist anzunehmen, dass das Pareto-Prinzip auch für das Nutzungsverhalten von Telekommunikationsendeinrichtungen gültig ist. Demnach würden 80 Prozent der Nutzer 20 Prozent der Funktionalitäten ihres Systems tatsächlich nutzen. Die Verteilung ist dabei von den Nutzungsszenarien abhängig (vgl. Abschnitt 2.3.). Die Beschreibung soll lediglich als beispielhafte Anleitung dienen, wie gerätespezifische Haupt- und Nebenfunktionen festgelegt werden können. Mischformen der einzelnen Kategorien sind möglich, werden hier aber nicht explizit aufgeführt.

Hauptfunktionen sollten schnell und einfach durch den Dialog zwischen Benutzer und Gerät erreichbar sein. Nebenfunktionen sind wichtige aber weniger häufig genutzte Funktionen, die oft einen direkten Bezug zur Hauptfunktion haben. Weiterhin gibt es Zusatzfunktionen, die in der Regel nicht gerätespezifisch sind und deren Wegfall die Bedienung des Gerätes nur marginal beeinträchtigt. Die Einteilung in Haupt-, Neben- und Zusatzfunktionen ist nicht immer eindeutig. Sie kann für spezielle Anwendungsfälle variieren und verändert sich mit der Weiterentwicklung von Geräten. Im Mobilfunkbereich setzen sich beispielsweise *Smartphones* mehr und mehr durch, bei denen es möglich ist, mit relativ geringem Aufwand Anwendungen für unterschiedliche Bereiche zu integrieren. Bei der Klassifizierung von Haupt-, Neben- und Zusatzfunktionen muss demnach der jeweilige Kontext berücksichtigt werden. Weiter sollten wegen des rasanten technischen Fortschritts aktuelle Trends und neueste Entwicklungen mitbetrachtet werden, bevor eine vergleichende Qualitätsbewertung mehrerer Geräte durchgeführt wird.

2.2.1 Kommunikationsanwendungen

Kommunikationsanwendungen dienen der technikvermittelten menschlichen Kommunikation. Dabei ist die Übertragung von Informationen zwischen einem Sender und einem Empfänger üblicherweise die Hauptfunktion. Kommunikationsanwendungen sind zumeist durch eine Reihe typischer Nebenfunktionen wie z. B. Klingeltöne, Abschaltung aller akustischen Rückmeldungen, Termineintrag in einen Kalender, etc. gekennzeichnet.

Beispiele: (Internet-) Telefonie, Videotelefonie, Telefonkonferenzen, Videokonferenzen, SMS, MMS

2.2.2 Medienaufzeichnungs- und -abspielanwendungen

Mit Hilfe von Medienaufzeichnungsanwendungen werden Audio- oder Videodaten gespeichert. Diese können zu einem späteren Zeitpunkt durch Medienabspielanwendungen wiedergegeben werden. Diese Aufzeichnungs- und Abspielfunktionen sind dabei meist Hauptfunktionen (siehe Beispiele unten). Einige Nebenfunktionen von Medienaufzeichnungsanwendungen werden durch das Aufzeichnungsmedium beeinflusst. So können die Informationsträger

digital oder analog sein, wobei vor allem bei digitalen Speichermedien häufig unterschiedliche Datenformate und Qualitätsstufen wählbar sind. Häufig muss dabei beachtet werden, wie die gespeicherten Daten weiterverarbeitet werden sollen. Auch bei Medienabspielanwendungen können vielfältige Nebenfunktionen vorhanden sein, was ebenfalls darin begründet ist, dass analoge oder digitale Datenquellen verwendet werden. Beispielsweise bieten DVD-Abspielanwendungen meist einen größeren Funktionsumfang als Abspielanwendungen für Videokassetten.

Beispiele: Anrufbeantworter, Aufnahme- und Abspielgeräte für Audio und Video (z. B. CD, VHS, DVD, BluRay), Online-Videoportale

2.2.3 Informationsabrufanwendungen

Informationsabrufanwendungen umfassen ein sehr breites Spektrum und werden von einer breiten Masse an Benutzern alltäglich verwendet. Viele dieser Anwendungen sind z. B. im öffentlichen Raum zu finden und sollen von Menschen benutzbar sein, die keine Vorerfahrung mit der Bedienung der Anwendung haben. Daher ist intuitive Benutzbarkeit zumeist eine wichtige Systemeigenschaft. Dennoch ist häufig Vorwissen über die Anwendung von Vorteil, und oft werden zumindest grundlegende Computerkenntnisse zum Verständnis der Interaktionslogik vorausgesetzt. Hauptfunktion ist zumeist das Finden einzelner oder komplexer Informationen, wozu bei GUI¹-basierten Anwendungen beispielsweise Auswahllisten und Textfelder verwendet werden, um bestimmte Suchkriterien einzugrenzen. Die Nebenfunktionen dienen häufig dazu, die Suche nach einer Information durch weitere Details einzugrenzen oder z. B. durch Auswahl der passenden Ausgabesprache zu vereinfachen.

Beispiele: Informationsterminals, Fahrkartenautomat, Restaurantinformationssystem, Online-datenbanken wie Wikipedia

2.3 Klassifikation nach Nutzungsszenarien

Nicht jede Evaluierungsmethode ist für jeden Nutzungskontext tauglich. Entsprechend sollten vor der Planung der Evaluation die Nutzungskontexte festgelegt und analysiert werden. Eine grobe Einteilung lässt sich anhand der Mobilität (fest installiert vs. mobil) der Systeme treffen. Weiterhin ist die Freiwilligkeit bzw. Motivation eine relevante Einflussgröße. So kann der Nutzer bspw. durch beruflichen Kontext oder mangelnde Verfügbarkeit von Alternativen zum Gebrauch gezwungen sein. Außerdem von Einfluss ist die Privatheit der Nutzung. Verhalten in Gegenwart anderer unterscheidet sich deutlich von Verhalten ohne die Anwesenheit weiterer Personen.

2.4 Klassifikation nach Ein- und Ausgabemodalitäten

Häufig werden nur drei Modalitäten unterschieden, welche mit drei menschlichen Sinnen korrespondieren. Gegenwärtig sind diese drei Sinne zwar diejenigen mit zentralster Bedeutung hinsichtlich multimodaler Mensch-Maschine-Interaktion, doch sind in der Physiologie weitere Sinne definiert (Schomaker et al., 1995). Tabelle 1 stellt die Sinneswahrnehmungen, die Sinnesorgane und die jeweilige Modalität dar.

¹ GUI: Graphical User Interface: Grafische Nutzerschnittstelle.

Sinneswahrnehmung	Sinnesorgan	Modalität
Sehsinn	Augen	Visuell
Hörsinn	Ohren	Auditiv
Tastsinn	Haut	Taktil
Geruchssinn	Nase	Olfaktorisch
Geschmackssinn	Zunge	Gustatorisch
Gleichgewichtssinn	Gleichgewichtsorgan	Vestibular

Tabelle 1. Sinne und korrespondierende Modalitäten (übersetzt nach Schomaker et al., 1995)

Eine andere Definition schlägt Bernsen (1997) vor: Modalität ist die Art und Weise der Informationsrepräsentation zwischen Menschen oder Maschinen. Damit weicht Bernsen von der traditionellen kognitionspsychologischen Sichtweise ab: Modalitäten im Sinne von Bernsen beziehen sich auf die Darstellung von Informationen sowie Interaktionstechniken und nicht auf die menschlichen Sinneskanäle (Bernsen, 1997). Beispiele für Bernsens Modalitätsbegriff sind geschriebene und gesprochene Sprache oder auch die akustische Rückmeldung eines Rechners (Bernsen, 1997).

Diese verschiedenen Definitionen demonstrieren zwei verschiedene Ansätze: den nutzerzentrierten und den systemorientierten Ansatz (Baber, 2001). Der nutzerzentrierte Ansatz versteht Modalität im Sinne menschlicher Wahrnehmung. Der systemorientierte Ansatz ist auf Interaktionstechniken und Ein- und Ausgabegeräte fokussiert (Baber, 2001). Für multimodale Mensch-Maschine-Interaktion sollten beide Ansätze Berücksichtigung finden: Ein multimodales System agiert üblicherweise mit verschiedenen Interaktionstechniken und Ein- und Ausgabegeräten, gleichzeitig werden vom Nutzer solch eines Systems multiple sensorischen Ressourcen benötigt (Ilmonen, 2006).

Gegenwärtig sind die drei menschlichen Sinne Sehen, Hören und Fühlen (s. oben) hinsichtlich multimodaler Mensch-Maschine-Interaktion diejenigen mit zentraler Bedeutung. Der Sehsinn der Benutzer wird von klassischen Kommunikationsschnittstellen häufig mit Bildern, Videos und geschriebenen Text angesprochen. Dabei werden häufig *Icons* als vereinfachte, bildliche Darstellung gewählt, da mit ihnen auf schnelle Art und Weise Information dargeboten werden kann. Im Bereich der auditiven Wahrnehmung gibt es sowohl die Möglichkeit mit gesprochener Sprache als auch mit geeigneten Klängen und Geräuschen Informationen zu übermitteln. Sprachausgabe wird häufig in telefonbasierten Anwendungen genutzt und in Umgebungen, in denen der visuelle Kanal stark beansprucht ist (z. B. Navigationsgerät im Fahrzeug). Der Tastsinn wird von verschiedenen Systemen bspw. durch Vibration, Material und Form von Oberflächen und Tasten oder auch durch das Gewicht eines Gerätes angesprochen.

3 Usability Engineering Lifecycle

Um eine optimale *Usability* zu gewährleisten, müssen bereits vor dem Systementwurf Vorkehrungen getroffen werden, vgl. Möller (2010). So müssen die augenblickliche Situation, die Anforderungen der Nutzer, ihr Verhalten und das Benutzungsumfeld analysiert werden, und daraus Anforderungen und Ziele bezüglich verschiedener *Usability*-Metriken festzulegen. Nach erfolgter Implementierung dienen diese Ziele einer quantitativen Überprüfung des Erfolgs. Auch der finanzielle Rahmen für den Einsatz von *Usability-Engineering*-Praktiken kann abgesteckt werden; insbesondere vor dem Design sind diese Praktiken sehr preiswert einzusetzen, da noch keine Änderungen am System notwendig sind. Nach erfolgter Analyse geht es dann in das Systemdesign. Dieses ist von zentraler Bedeutung, werden doch hier die Grundlagen für eine gute Mensch-Maschine-Interaktion gelegt. Ein oder mehrere Design-Vorschläge werden anschließend in Prototypen umgesetzt, welche bezüglich der *Usability* evaluiert werden können. Hierzu bestehen zwei Möglichkeiten: Zum einen verschiedene Methoden der expertenbasierten Evaluierung, zum anderen Tests mit Nutzern. Beide Verfahren werden typischerweise parallel bzw. alternierend eingesetzt und liefern Hinweise auf *Usability*-Probleme sowie Verbesserungsmöglichkeiten.

Diese werden im iterativen Design umgesetzt, welches u. U. mehrere Zyklen von Re-Design und Evaluierung umfasst. Anschließend kann das neue System oder der neue Dienst im Feld (d.h. in der Praxis) benutzt werden. Rückmeldungen vom Feld gestatten eine weitere iterative Optimierung und liefern Ideen für zukünftige Generationen dieses und anderer Systeme und Dienste.

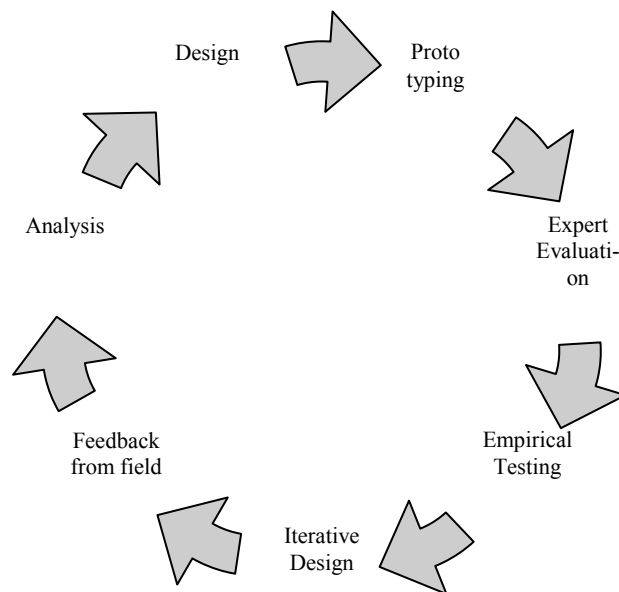


Abbildung 2: *Usability Engineering Lifecycle*, nach Nielsen (1993)

4 Evaluationsverfahren

4.1 Übersicht und Einführung

Mit einer Evaluation sind üblicherweise drei Hauptziele verbunden (Dix et al., 1993): Die Bewertung der Funktionalität eines Systems, die Bewertung der Wirkung der Schnittstelle auf den Nutzer und das Identifizieren von spezifischen Problemen mit dem System. Generell kann

man Evaluationsmethoden nach fünf verschiedenen Gesichtspunkten unterscheiden (Preece, et al., 1994). Eine erste Unterscheidung erfolgt nach der Fragestellung der Evaluation:

- Die Evaluation ist auf ein bestimmtes Ziel ausgerichtet, die Fragestellung lautet: „Ist das System gut genug?“
- Die Evaluation vergleicht zwei alternative Designs, die Fragestellung lautet: „Welches System ist besser?“
- Mit der Evaluation soll der realen Welt näher gekommen werden: „Wie gut arbeitet das System in der realen Welt?“
- Mit der Evaluation wird die Einhaltung von Standards überprüft: „Wie gut erfüllt das System den/die Standards?“

Je nach Fragestellung muss entschieden werden, wann, d.h. in welchem Stadium der Systementwicklung, die Evaluation stattfinden soll. Dabei wird unterschieden in:

- **Formative Evaluation:** Die formative Evaluation ist eine prozessorientierte Evaluation, die während des Entwicklungsprozesses stattfindet und schon vor dem Vorhandensein eines Prototyps, also schon im Designstadium, erfolgen kann. Ziel ist die Verbesserung des Produktes als Teil des iterativen Designprozesses.
- **Summative Evaluation:** Die summative Evaluation ist ergebnisorientiert und eine zusammenfassende Beurteilung des fertigen Systems. Diese Form der Evaluation erfolgt, wenn das System schon implementiert ist.

Zusätzliche Distinktionskriterien sind nach Dix et al. (1993):

- **Das Niveau und die Art der erhobenen Informationen:** Die Informationen, die durch eine Evaluation gewonnen werden, können Informationen von niedrigem Level (z. B.: „Welche Schriftart ist am Besten zu lesen?“) oder Informationen von hohem Level (z. B.: „Ist das System leicht zu benutzen?“) umfassen. Dabei ist zwischen quantitativen, d.h. numerischen Daten sowie qualitativen, d. h. verbalen bzw. nicht-numerischen Daten zu unterscheiden. Beide Datenarten haben besondere Vor- und Nachteile. Bspw. sind quantitative Daten im Gegensatz zu qualitativen Daten statistisch auswertbar. Allerdings können numerische Messwerte unter Umständen weniger detailliert und inhaltsreich sein als qualitative Daten (vgl. Bortz & Döring, 2006).
- **„Aufdringlichkeit“ des Verfahrens:** Die „Aufdringlichkeit“ des Verfahrens bezeichnet das Ausmaß, in dem der Nutzer während der Evaluation in seinem natürlichen Verhalten „gestört“ wird. Hier sei bspw. auf Evaluatoreffekte oder Beeinträchtigung durch Verkabelung der Versuchspersonen verwiesen.
- **Benötigte Ressourcen:** Evaluationsverfahren beanspruchen Ressourcen in unterschiedlicher Art und in unterschiedlichem Ausmaß. Bedacht werden müssen Faktoren wie Geld, Zeit, technische Ausrüstung, benötigte Expertise und benötigte Versuchspersonen.

Ein weiteres Kriterium zur Abgrenzung ist das Ausmaß der Einbindung der Benutzer. Die Einteilung erfolgt:

- Nach dem *Grad der Kontrolle*, den die Nutzer über ihre Aufgaben/Aktivitäten während der Evaluation und über die Evaluationsprozedur im Allgemeinen haben. Je formalisierter und wissenschaftlicher die Evaluation ist, umso weniger Kontrolle liegt bei den Nutzern (Preece et al., 1994).

- Nach dem *Grad der Nutzerbeteiligung*: Einige Verfahren, insbesondere formative Evaluationen im Rahmen iterativer Designprozesse, können auch ohne Beteiligung von Nutzern durchgeführt werden. Nutzerzentrierte/empirische Verfahren stehen dabei analytischen/formalen Verfahren gegenüber:
 - *Expertenzentrierte/analytische Verfahren* werden typischerweise in frühen Entwicklungsstadien eingesetzt. Dabei werden die Elemente der Benutzerschnittstelle und ihre Auswirkungen analysiert und modelliert. Größtenteils erfolgt die Evaluierung durch *Usability*-Experten, d.h. Nutzer und ein fertiges System werden nicht benötigt. Diese Verfahren werden auch unter dem Begriff „*Usability*-Inspektion“ zusammengefasst (Holzinger, 2005).
 - Bei *nutzerzentrierten/empirischen Verfahren* erfolgt die Datenerhebung über das Beobachten oder „Messen“ von Nutzern. Die gemessenen Variablen repräsentieren die Performanz des Systems oder die Einstellung der Nutzer zum System. Die Implementation des Systems ist zu einem gewissen Maße Voraussetzung (Sturm, 2005). Eine andere Bezeichnung für diese Methoden ist *Usability Testing* (Holzinger, 2005).

Im Folgenden werden einige Verfahren näher beschrieben, anschließend sind in „Tabelle zu Verfahren und deren Eignung“ (Kapitel 5.1), in Anlehnung an Dix et al. (2004), einige etablierte Verfahren im Überblick dargestellt. Den Abschluss dieses Kapitels bilden einige Hinweise und Fragestellungen zur Auswahl geeigneter Prüfaufgaben, die zur Durchführung der vorgestellten Verfahren notwendig sind.

4.2 Expertenzentrierte Verfahren

4.2.1 Cognitive Walkthrough (CW)

Der *Cognitive Walkthrough* ist eine aufgabenorientierte Methode (Holzinger, 2005). Dabei analysieren Experten, üblicherweise Designer oder Psychologen, die Funktionalitäten des Systems auf Basis einer Beschreibung der Schnittstelle, einer Beschreibung der Aufgabe welche der Nutzer ausführen wird, einer Liste von Handlungen, die notwendig sind, um die Aufgabe zu erfüllen, und Informationen über die Nutzer und den Nutzungskontext (Wharton, Riemann, Lewis & Polson, 1994). Grundlage sind Theorien des entdeckenden Lernens und Forschungsergebnisse zum Problemlösen (Wharton et al., 1994). Der Ablauf gliedert sich in fünf Schritte (Wharton et al., 1994):

- Definition der Eingangsgrößen für den *Cognitive Walkthrough*
 - Identifizieren der Nutzer
 - Festlegen der zu evaluierenden Aufgaben
 - Festlegen und Beschreiben der Handlungssequenzen für jede Aufgabe
 - (detaillierte) Beschreibung der Schnittstelle
- Einberufen der Analysten
- Analysieren der Handlungssequenzen für jede Aufgabe
 - Simulation der mentalen Prozesse des Nutzers
- Protokollieren kritischer Informationen
- Überarbeiten der Schnittstelle

Vorteil des CW ist die Unabhängigkeit von Nutzern und einem voll funktionstüchtigem System. Mit der Durchführung eines CW kann der Designer unterstützt werden, die Perspektive eines potentiellen Nutzers einzunehmen. Die effektive Identifikation von Problemen bei der Interaktion von Nutzer und System kann dabei helfen, Nutzerziele und Vermutungen der Nutzer über das System zu definieren (Holzinger, 2005). Nachteile des CW sind das niedrige Level der resultierenden Informationen. Auch kann ein CW abhängig vom zu testenden System sehr langwierig sein und bei falscher Aufgabenselektion verzerrte Ergebnisse liefern (Holzinger, 2005).

4.2.2 Prüflisten

Mit Hilfe von Prüflisten soll die Evaluation von Kommunikationsendgeräten durch Experten sowohl aufgabenunabhängig als auch unter Betrachtung konkreter Aufgaben ermöglicht werden. In einer gegliederten Liste werden dazu Eigenschaften zusammengestellt, die vom betrachteten System gefordert werden. Viele Prüflisten bieten lediglich eine Aufzählung solcher Eigenschaften in Form von Gestaltungsregeln, eine zusätzliche Bewertung durch Angaben über die Erfüllung einzelner Kriterien und deren Gewichtung durch die Prüfer ist möglich. Bei der Mensch-Technik-Interaktion muss sich vor allem die Gestaltung des Dialogs an Aufgaben orientieren. Ergonomische Kriterien, die die menschliche Sensorik und Motorik betreffen, werden jedoch als weitgehend aufgabenunabhängig betrachtet. In einer inzwischen überholten Fassung dieser ITG-Richtlinie von 1994 wurde eine Prüfliste erarbeitet, die sich in zwei Hauptteile – einen aufgabenunabhängigen und einen aufgabenorientierten Teil – gliedert. Ähnlich den Prüflisten ist das Verfahren der heuristischen Evaluation (s. unten) sehr breit angelegt, jedoch sind Prüflisten in der Regel ausführlicher.

4.2.3 Heuristische Evaluation (HE)

Der Begriff „heuristisch“ leitet sich vom Griechischen „heureskein“ ab und bedeutet etwas zu finden bzw. entdecken (Holzinger, 2005). Die heuristische Evaluation ist eine Methode des *Discount Usability Engineering*, ein von Nielsen eingeführter Ansatz mit pragmatischer und kostengünstiger Ausrichtung, der sich der Argumentation entgegenstellt, *Usability*-Evaluation sei zu teuer, zu schwierig und würde zu lange dauern (Nielsen, 1994).

Bei einer HE bewerten mehrere Experten die Benutzerschnittstelle. Überprüft wird, ob gängigen *Usability*-Prinzipien (Heuristiken) Rechnung getragen wird (Nielsen, 1994). In der ursprünglichen Version wird verlangt, dass jeder Evaluator die Schnittstelle für sich allein untersucht und mittels der Heuristiken mehrmals Interaktionen mit der Schnittstelle durchführt. Erst nach Abschluss aller Bewertungen kommunizieren die Experten miteinander, um ein aggregiertes Urteil zu treffen. Dem zugrunde liegt das Ziel, dass ein möglichst unabhängiges, unverzerrtes Urteil der einzelnen Gutachter angestrebt wird (Nielsen, 1994). Das Ergebnis einer HE ist eine Liste von *Usability*-Problemen, die genau beschrieben und begründet sein sollten (Nielsen, 1994). Weiterhin können die Probleme hinsichtlich ihrer Häufigkeit, ihres Einflusses und ihrer Persistenz beurteilt werden (Nielsen, 1994). Gewöhnlich werden ca. 3-5 Experten benötigt, die ca. 60-70 Prozent der Probleme aufdecken. Bei mehr als 10 Gutachtern sind kaum verbesserte Ergebnisse zu erwarten (Nielsen, 1994).

Die Vorteile der HE liegen in der schnellen, einfachen und kostengünstigen Durchführbarkeit sowie in der universellen Anwendbarkeit während des gesamten Designprozesses (Holzinger, 2005). So ist, auch bei noch nicht implementierten Systemen, eine HE möglich, die Experten

bewerten dann den vorliegenden Systementwurf (Nielsen, 1994). Außerdem können mittels HE sowohl größere als auch kleinere *Usability*-Probleme erfasst werden. Jedoch ist es fraglich, inwieweit es den Experten gelingt, sich in die Rolle der Nutzer einzufühlen. Dies ist jedoch ein generelles Problem vieler expertenzentrierter, analytischer Verfahren.

4.2.4 Modellbasierte Evaluation (ME)

Die modellbasierte Evaluation erfolgt auf Basis von Nutzermodellen, mit denen Nutzerverhalten während der Interaktion mit dem System vorhergesagt werden soll (Sturm, 2005). In der einfachsten Form sind dies Modelle zur Dauer und Genauigkeit von Arm- und Mausbewegungen (*Fitt's Law*) oder Ausführungszeiten bei der Tastaturbedienung (*Keystroke Level Model*, oder die Methode *Goal Operators Methods Selection Rules*, GOMS). Sollen komplexere Denk- oder Wahrnehmungsprozesse nachgebildet werden, wird auf Theorien und Befunde der kognitiven Psychologie (Dix et al., 1993) oder Angaben von Experten aus dem jeweiligen Anwendungsbereich zurückgegriffen. Beispiele sind Aufgabenanalysen wie die *Hierarchical Task Analysis* (Annett, 2003), die *Cognitive Complexity Theory* (CCT) von Kieras und Polson (1985), der *Adaptive Control of Thought-Rational* (ACT-R) Ansatz von Anderson und Kollegen (z. B. Anderson & Lebiere, 1998) oder das *State, Operator And Result Modell* (SOAR) (Newell, 1990). Insbesondere die letzteren Verfahren werden eher im akademischen als im angewandten Bereich genutzt.

Die ME kann schon sehr früh im Designprozess angewandt werden und ermöglicht durch die theoretische Fundiertheit einen tiefen Einblick ins Verhalten der Nutzer (Holzinger, 2005). Der Schwachpunkt der ME liegt in dem erforderlichen Zeitaufwand, da jede Aufgabe in ihre Einzelschritte zerlegt werden muss, und dem benötigten, sehr hohen Grad an Expertise (Holzinger, 2005). Dennoch ist sie für *Usability*-Bewertungen in Bereichen, in denen die Nutzergruppe klein oder eine exakte Abschätzung aller möglichen Folgen einer Aktion notwendig ist (Piloten, Kraftwerke), durchaus nützlich.

4.3 Nutzerzentrierte Verfahren

4.3.1 Experiment

Bei experimenteller Evaluation werden spezifische Aspekte des Interaktionsverhaltens eines Systems kontrolliert bewertet (Sturm, 2005). Zu diesem Zweck werden im einfachsten Experimentaldesign eine Hypothese formuliert und mindestens zwei verschiedene experimentelle Bedingungen geschaffen, die sich nur hinsichtlich des zu testenden, manipulierten Faktors unterscheiden. Alle Unterschiede, die sich in der gemessenen, abhängigen Variable ergeben, werden auf Unterschiede hinsichtlich der getesteten, unabhängigen Variable zurückgeführt (Dix, et al. 1993). Mit Experimenten ist es möglich, Daten hoher Qualität zu erheben, da Störvariablen im hohen Maße ausgeschaltet oder kontrolliert werden können. Experimente ermöglichen, eine saubere Durchführung vorausgesetzt, Kausalschlüsse über Mensch-Maschine-Interaktionen und sind zur Theoriebildung und -überprüfung unerlässlich (Hegner, 2003). Der stark kontrollierte Charakter eines Experimentes bringt es jedoch mit sich, dass die Ergebnisse nicht das Nutzerverhalten in natürlichen Umgebungen abbilden (Sturm, 2005). Auch ist das Aufsetzen, Durchführen und Auswerten eines Experimentes zeit- und kostenintensiv.

4.3.2 Feldstudien

Um Daten aus realen Nutzungssituationen zu sammeln, bieten sich Feldstudien an. Diese geben Aufschluss über Dinge, die in Laborexperimenten nur schlecht oder gar nicht untersucht werden können, wie insbesondere reales Nutzungsverhalten (und damit auch die Akzeptanz). Jedoch sind sie für Kausalschlüsse nicht geeignet, da die Bedingungen nicht so kontrollierbar sind wie in einem Laborexperiment. Je nach Grad der Kontrolliertheit kann man Feldstudien als Quasi-Experimente betrachten.

4.3.3 Interviews/Fragebögen

Der Einsatz von Fragebögen und Interviews bietet sich dann an, wenn Daten erhoben werden sollen, die über Performanzdaten hinausgehen. Insbesondere zu Erfassung subjektiver Daten wie Nutzerzufriedenheit, Einstellungen der Nutzer und Ängste der Nutzer wird auf Fragebögen und Interviews zurückgegriffen (Holzinger, 2005). Damit können Konstrukte erfasst werden, die sich anderen Verfahren entziehen. Abgesehen davon kann eine Befragung verhältnismäßig schnell und günstig durchgeführt werden. Anzumerken ist jedoch, dass diese subjektiven Daten durch niedrige Validität gekennzeichnet sind (Holzinger, 2005).

4.3.4 Protokollanalysen und Lautes Denken (*Thinking Aloud*)

Bei Protokollanalysen werden Nutzer und Nutzerverhalten mittels Video-, Audio-, Log-Dateien oder schriftlichem Protokoll festgehalten.

Die wohl bekannteste Methode ist das *Thinking Aloud*, das laute Denken. Die Testpersonen werden, während sie die Schnittstelle benutzen, dazu angehalten, ihr Verhalten und ihre Gedanken zu verbalisieren und laut zu äußern (Holzinger, 2005). Dies kann während des Tests geschehen oder nachträglich als retrospektives lautes Denken, bei dem der Nutzer (Video-) Aufzeichnungen der Testsession kommentiert. Beim retrospektiven lauten Denken ist jedoch mit Verzerrungen durch Erinnerungsfehler zu rechnen. Bei einer weiteren Variante sind mehrere Probanden eingebunden, welche zusammen die Schnittstelle nutzen. Damit soll die Unnatürlichkeit dieser Methode reduziert werden. Hackmann und Biers (1992) zeigten, dass lautes Denken in Zweier-Teams einzelnen Evaluatoren vorzuziehen ist, da hier bessere Ergebnisse zu erwarten sind.

Diese Methode kann sowohl bei freier Exploration der Schnittstelle, bspw. zur Erhebung eines generellen Eindrucks über das System, als auch bei der Bearbeitung konkreter Aufgabenstellungen angewandt werden. Stellen und Elemente des Systems, die ungünstig sind und vom Nutzer fehl interpretiert werden, fallen schnell auf. Jedoch zeigte sich, dass *Thinking Aloud* häufig als unnatürlich und verwirrend empfunden wird (Lin, Choong & Salvendy, 1997). Zudem sind oft Hinweise durch den Versuchsleiter nötig, um die Teilnehmer zum lauten Denken zu animieren. Mitunter fällt den Probanden das ununterbrochene Verbalisieren ihrer Gedanken schwer. (Hegner, 2003). Problematisch sind zusätzlich Verzerrungen durch soziale Erwünschtheit (Hegner, 2003). Hinzu kommen Probleme bei Schnittstellen, die auditive Ein- und Ausgaben verwenden (z. B. Sprachdialogsysteme), da diese Informationen durch das ununterbrochene Verbalisieren verdeckt werden können.

5 Praktischer Leitfaden

5.1 Tabelle zu Verfahren und deren Eignung

	Cognitive Walkthrough	Heuristische Evaluation	Modellbasierte Evaluation	Experiment	Feldstudie	Interviews, Fragebögen	Protokollanalyse/ Thinking Aloud
Stadium im Entwicklungsprozess	Jedes Stadium	Jedes Stadium	Design	Jedes Stadium	Nach Prototyping	Jedes Stadium	Nach Prototyping
Nutzereinbindung	Nein	Nein	Nein	Ja	Ja	Ja	Ja
Stil	Labor	Labor	Labor	Labor	Feld	Labor/Feld	Labor/Feld
Level der Information	Niedrig	Hoch	Niedrig	Hoch und niedrig	Hoch und niedrig	Hoch	Hoch und niedrig
Art der Daten	Qualitativ	Qualitativ	Quantitativ	Qualitativ und quantitativ	Eher qualitativ	Qualitativ und quantitativ	Qualitativ
Ist das Verfahren aufdringlich?	Nein	Nein	Nein	Ja	Ja	Nein	Ja
Ausmaß benötigter Ressourcen	Mittel	Niedrig	Mittel	Hoch	Hoch	Niedrig	Hoch
Aufgabenorientierung	Ja	Nein	Typischerweise ja	Ja und nein	Ja und nein	Ja und nein	Ja und nein
Eignung für Modalitäten	Auditiv, visuell und haptisch/kinästhetisch; multimodal eingeschränkt	Auditiv, visuell, haptisch/kinästhetisch; multimodal eingeschränkt	Auditiv, visuell, haptisch/kinästhetisch; multimodal eingeschränkt	Auditiv, visuell, haptisch/kinästhetisch und multimodal	Auditiv, visuell, haptisch/kinästhetisch und multimodal	Auditiv, visuell, haptisch/kinästhetisch und multimodal	Visuell, haptisch/kinästhetisch und multimodal; ungeeignet für auditive Modalität
Hauptnachteile	Nur angenommene Probleme, hohe Expertise benötigt	Nur angenommene Probleme, passende Heuristiken nicht immer vorhanden, hohe Expertise benötigt	Hohe Expertise benötigt	Sehr aufwendig	Aufwendig und schwer zu kontrollieren, funktionieren der Prototyp notwendig	Oft lange Vorbereitungszeit zur Erstellung valider Fragebögen notwendig	Lautes Denken fällt Nutzern oft schwer, für Sprachdialogsysteme ungeeignet, funktionieren der Prototyp notwendig
Hauptvorteile	Ressourcensparend	Ressourcensparend	Ressourcensparend	Generalisierbare Daten	Daten werden von „echten“ Endnutzern generiert	Einfach anzuwenden	Daten werden von „echten“ Endnutzern generiert

5.2 Ermittlung der Testaufgaben mittels Fragekatalog

Die meisten der hier vorgestellten Verfahren zur Evaluation von Kommunikationseinrichtungen sind zumindest zum Teil aufgabenorientiert, d.h. zu ihrer Durchführung bedarf es eines festen Bestandes an Aufgaben, die mit einer Anwendung der Kommunikationseinrichtung ausgeführt werden können. Der folgende Fragenkatalog dient der Ermittlung von Testaufgaben, die als Grundlage für die Evaluation mit einem der Verfahren herangezogen werden können. Ausgangspunkt der Überlegungen sind zwei grundlegende Fragen:

- Welche Anwendungen sind in das Gerät integriert? Die Kommunikationseinrichtung muss die zu testende Anwendung enthalten. Wenn mehrere Kommunikationseinrichtungen verglichen werden sollen ist zu beachten, ob evtl. manche Anwendungen als Teil einer Anwendung oder als eigenständige Anwendung realisiert sind (z. B. Adressbuch und Telefon).
- Welche dieser Anwendungen sollen bewertet werden? Es müssen Anwendungen ausgewählt werden, für die Testaufgaben definiert werden sollen. Sollen mehrere Kommunikationseinrichtungen verglichen werden, muss beachtet werden, dass andere Anwendungen manchmal aus einer zu testenden Anwendung heraus aufgerufen werden können, wenn dort angesiedelte Funktionen genutzt werden.

5.2.1 Ermittlung der Hauptfunktionen

Eine klare Trennung von Anwendungen kann manchmal schwierig sein. Daher sollten Hauptfunktionen einer Anwendung definiert werden, die nach Möglichkeit keine Funktionen anderer Anwendungen nutzen. Die Ermittlung von Hauptfunktionen einer Anwendung kann über die Beobachtung und Befragung realer Endnutzer oder durch eigene Definition erfolgen. Grundsätzlich müssen Hauptfunktionen einer Anwendung ermittelt werden, um eine sinnvolle Bewertung zu gewährleisten. Zur Ermittlung von Hauptfunktionen hilft die Beantwortung folgender Fragen.

- Welche Funktionen werden am häufigsten ausgeführt? Hauptfunktionen sind Schlüsselfunktionen der Anwendung. Sie identifizieren die essentiellen Aufgaben, die ausgeführt werden.
- Welche Funktionen sind für den Betrieb der Anwendung notwendig? Schritte, die vollzogen werden müssen, um essentielle Aufgaben ausführen zu können, sind zu den Hauptfunktionen zu zählen. Dazu kann z. B. das Ein- und Ausschalten eines Gerätes (z. B. Fernseher) gehören.
- Welche Funktionen sind repräsentativ/typisch für die Anwendung? Manche Funktionen sind charakteristisch für eine Anwendung und nicht unbedingt essentiell aber sinnvoll für die Interaktion. So soll z. B. ein Eintrag im Adressbuch eines Telefons direkt angerufen werden können.

5.2.2 Ermittlung der Neben- und Zusatzfunktionen

Nebenfunktionen sind wichtige aber weniger häufig genutzte Funktionen. In der Regel haben sie einen direkten Bezug zur Hauptfunktion. Die Hauptfunktionen sollten jedoch im Prinzip ohne die Nutzung von Nebenfunktionen ausführbar sein. Beispielsweise ist ein Mobiltelefon

je nach Einstellung durch den Nutzer nur nach Eingabe einer persönlichen Identifizierungsnummer (PIN) nutzbar.

- Welche Nebenfunktionen sollen getestet werden? Je nach Szenario kann eine Nebenfunktion mehr oder weniger wichtig sein. Bei der Auswahl sollte besonders auf den spezifischen Benutzungskontext und die Besonderheiten von bestimmten Benutzergruppen geachtet werden.

5.3 Praktische Hinweise

5.3.1 Generelle Hinweise

Der Zweck einer *Usability*-Evaluation ist es, spätere Probleme bei der Nutzung zu identifizieren. Eine wichtige Voraussetzung dafür ist, dass dem Nutzer die interne Struktur der Anwendung/des Gerätes etc. nicht bekannt ist, denn nur so können Diskrepanzen zwischen seinen Annahmen und Erwartungen und denen der Entwickler identifiziert werden. Dies bedingt, dass die Evaluation ab einem gewissen Punkt ohne aktive Beteiligung der Entwickler stattfinden sollte, da der Einfluss auf den Versuchsablauf und/oder die Versuchspersonen die Ergebnisse stark beeinflussen kann.

Eine Vielzahl von *Usability*-Problemen entsteht immer noch durch schwer erkennbare Schriften/Icons oder unklare Benennungen von Funktionen. Eine erste Möglichkeit, Fehler dieser Art aufzudecken, ist, einen Kollegen unvoreingenommen das aktuelle Layout bewerten zu lassen. Diese ‚interne Evaluation‘ stößt allerdings schnell an ihre Grenzen, da die Kollegen zumeist zu gut mit der Grundstruktur oder der internen Logik des zu bewertenden Geräts/Programms vertraut sind, um es wirklich unbefangen benutzen zu können.

5.3.2 Methodische Hinweise

- Nutzertests sind aufwändiger als Experten-Evaluationen. Gleichzeitig ist es immer wieder überraschend, welche nicht vorausgesehenen Fehler naive Nutzer aufdecken können. Gerade für neue, ungewohnte Produkte sind sie daher auf jeden Fall angebracht.
- Die zum Test ausgewählten Nutzer sollten hinsichtlich ihrer Merkmale wie Alter, Geschlecht, Bildungsstand, Technikerfahrung etc. der späteren Haupt-Nutzergruppe entsprechen.
- Die Qualität von *Usability*-Tests steht und fällt mit der Auswahl der gestellten Aufgaben, die ein Testnutzer lösen soll. Die Aufgaben sollten klar formuliert und repräsentativ für die spätere Nutzung sein. Dabei ist darauf zu achten, dass der Aufgabentext weder unrealistische Szenarien entwirft noch implizit den Lösungsweg schon mit vorgibt (z. B. durch Formulierungen, wie „unter dem Menüpunkt...“).
- Der Versuchsleiter sollte eine erkennbar ‚neutrale‘ Instanz darstellen. Ist dies nicht der Fall, besteht die Gefahr, dass man durch suggestive Fragen oder Hinweise die Versuchsperson auf den richtigen Weg lenkt oder die Versuchsperson die Kommunikationsendeinrichtung ausschließlich positiv bewertet, weil sie glaubt, dies sei sozial erwünscht.

- Ähnlich sind Nutzertests (nicht aber Experten-Evaluationen) bei halbfertigen Funktionen/Interfaces nur begrenzt aussagekräftig bzw. fällt das Urteil häufig zu positiv aus, weil die meisten Nutzer Schwierigkeiten haben, sich nicht implementierte Interaktionsschritte detailliert vorzustellen bzw. implizit davon ausgehen, dass diese problemlos verlaufen werden. In diesem Fall sollte man die Aufgabe so wählen, dass nur bereits funktionierende Komponenten genutzt werden müssen.
- Tests mit Anwendungs-/Geräte-Versionen, die nicht stabil laufen, sind von nur sehr begrenzter Aussagekraft, da ein Nutzer die Systemabstürze entweder in seinem Urteil berücksichtigt (auch wenn man ihn bittet, dies nicht zu tun) oder die Seriosität des Tests insgesamt anzweifelt.
- *Usability*-Tests von Prototypen, an denen so gut wie keine Änderungen mehr vorgenommen werden können, kann man sich sparen. Stellen Nutzer fest, dass Ihre Verbesserungsvorschläge nicht umgesetzt wurden, führt dies häufig zu einer gesteigerten Ablehnung (Reaktanz).

5.3.3 Leitfragen zur Planung einer Evaluationsstudie

Bei der Planung einer Evaluationsstudie ist es ratsam die nachstehenden Fragen (Sonntag et al., 2009) möglichst vollständig und nacheinander zu beantworten:

- Untersuchungsziel: Was soll mit der Evaluationsstudie erreicht werden?
- Untersuchungsobjekt: Was ist das Untersuchungsobjekt (System)?
- Untersuchungszeitpunkt: In welchem Entwicklungsstadium soll das System evaluiert werden?
- Untersuchungsmethode: Welche Methode ist geeignet, um die relevanten Ziele zu erreichen?
- Untersuchungsorgan: Wer sollen die Probanden/Tester sein, und wie sollen sie akquiriert werden?
- Welche Hilfen werden dem Nutzer angeboten? Inwieweit leistet der Versuchsleiter Hilfestellungen?
- Untersuchte Funktionalität: Welche Funktionen sollen getestet werden, und was sind die geeigneten Testaufgaben hierfür?
 - Wann gilt eine Aufgabe als erfolgreich bearbeitet?
 - Wann wird eine Aufgabe abgebrochen?
- Untersuchungsergebnisse: Welche Daten sollen wie erhoben werden und wie sollen diese ausgewertet werden?
- Untersuchungsaufbau: Welche Ausstattung wird für die Evaluation benötigt (zusätzliche Hard- und Software, wie z. B. Kameras und Mikrofone)?
 - Wo und wann soll die Evaluation stattfinden?
 - Wie lang soll jeder einzelne Test dauern?
 - Wer soll die Evaluation durchführen?
 - Wie viele Probanden werden benötigt und wie werden sie vergütet?

5.3.4 Der Umgang mit Versuchspersonen und ihren Daten

Sobald *Usability*-Bewertung auf Versuchspersonen zurückgreift, stellt sich die Frage, welche Versuchsanordnungen ethisch vertretbar sind und wie mit den Teilnehmern sowie ihren Daten umzugehen ist.

Da *Usability*-Tests keine körperliche oder psychische Beeinträchtigungen für die Probanden mit sich bringen werden, sind die Auflagen hier weniger streng als für medizinische oder psychologische Untersuchungen, die in der Regel durch die Ethikkommission der zugehörigen Institution (Universität, Klinik etc.) genehmigt werden müssen. Dennoch sollte man sich auch für den ‚harmloseren‘ Fall von *Usability*-Tests über den Umgang mit Probanden und ihren Daten Gedanken machen, bevor man eine Untersuchung durchführt, denn hier besteht in den meisten Fällen ein grundsätzlicher Konflikt: Als Leiter einer Studie möchte man für den Aufwand, den man betreibt, möglichst viele und genaue Daten erhalten, gleichzeitig steht man aber auch gegenüber den Versuchspersonen, die sich dem Versuchsleiter für eine gewisse Zeit ‚anvertrauen‘, in der Verantwortung.

Es ist selbstverständlich, dass Datenschutzrichtlinien eingehalten und die Persönlichkeitsrechte der Teilnehmer geachtet werden. Die folgenden Punkte sind keine juristische Belehrung und die Auflistung erhebt keinen Anspruch auf Vollständigkeit, sondern beinhaltet lediglich Hinweise, die sich in der Praxis bewährt haben. Als Mindestanforderung kann eine einfache Maxime gelten:

- Die Probanden sollten so behandelt werden, wie man auch selbst in dieser Situation behandelt werden möchte.
- Mit den Daten sollte im weiteren Verlauf nichts gemacht werden, was man mit Daten der eigenen Person nicht auch zulassen würde.

Konkret bedeutet dies:

- Schon im Akquiseschreiben (Aushang, Rundmail o.ä.) sollte sich eine knappe Schilderung des Versuchs befinden oder zumindest keine dem tatsächlichen Ablauf widersprechenden Informationen verbreitet werden.
- Die Probanden müssen sich vor Versuchsbeginn explizit mit der Erfassung, Speicherung und Auswertung ihrer Daten einverstanden erklären.
- Die Teilnehmer müssen spätestens im Anschluss, besser aber noch zu Beginn über den Zweck der Untersuchung und den Auftraggeber aufgeklärt werden.
- Der Versuchsleiter sollte nicht nur wissen, was er macht, er sollte es auf Nachfragen auch kurz verständlich erklären können.
- Alle aufgezeichneten Daten dürfen nicht für unberechtigte Dritte zugänglich sein, auch nicht für andere Versuchsteilnehmer. Nichts wirkt unprofessioneller als herumliegende ausgefüllte Fragebogen anderer Teilnehmer im Untersuchungsraum.
- Die Untersuchungsdaten sollten kodiert sein und keine unmittelbaren Rückschlüsse auf die untersuchte Person erlauben. Dokumente wie Terminpläne und Quittungen, die den Klarnamen der Probanden enthalten, sollten getrennt aufbewahrt werden.
- Jeder Proband wird freundlich und mit Respekt behandelt, und auch auf vermeintlich unsinnige Befürchtungen („Handystrahlung“) wird ernsthaft eingegangen.

- Dem Wunsch von Seiten eines Teilnehmers, den Versuch zu einem bestimmten Zeitpunkt zu beenden, ist unmittelbar Folge zu leisten.
- Den Teilnehmern wird explizit erklärt, dass der Zweck die Bewertung einer Schnittstelle/Kommunikationseinrichtung und nicht die Bewertung einer Person ist.
- Es werden keine Vergleiche der Angaben oder Leistung eines Probanden mit anderen Teilnehmern angestellt („Die meisten andern waren schneller als Sie.“).
- Bei der Versuchsdurchführung sind nicht mehr Personen anwesend als nötig.
- Die Untersuchung dauert nicht länger als vorher angegeben – oder wenn doch, dann nur nach Absprache mit dem Probanden und entsprechender zusätzlicher Vergütung.
- Aufwandsentschädigung etc. sollten wenn möglich unmittelbar im Anschluss an den Versuch und in bar ausgehändigt werden.
- Die Daten werden nicht ohne vorheriges Einverständnis an Dritte weitergegeben.
- Ton- oder Bildaufnahmen, auf denen die Versuchsperson zu identifizieren ist, werden ohne explizites Einverständnis nicht weitergegeben oder Dritten präsentiert. Das gilt auch für Bilder vom Versuchsaufbau, auf denen eine Person zu sehen ist.

Es hat sich bewährt, wenn ein Versuchsleiter zunächst selbst als Versuchsperson in einem geplanten Test fungiert. Auf diese Weise lassen sich mögliche Unannehmlichkeiten für Probanden leicht feststellen und es wird ihm/ihr bewusst, was er/sie den Teilnehmern zumutet. Ebenso sind Kollegen oder Bekannte als Vorversuchspersonen besser geeignet, Schwachpunkte zu identifizieren, da sie von dem Versuchsaufbau in der Regel weniger beeindruckt/ eingeschüchtert sind als “fremde“ Probanden.

Fragebögen und Interviewleitfäden sollten im Vorhinein mit Kollegen oder bei Unsicherheit mit einem Datenschutzbeauftragten durchgesprochen werden, um sicherzustellen, dass man nicht evtl. Fragen stellt, auf die aufgrund ihres intimen Charakters eine unwahre Antwort wahrscheinlich ist. Ähnlich ist bei Anbietern, die solche Informationen versprechen, Misstrauen angebracht.

Anhang A

Beispiel für eine Versuchsbeschreibung + Einverständniserklärung

Liebe(r) Versuchsteilnehmer(in),

in den folgenden 45 Minuten werden Sie verschiedene Audio-Clips hinsichtlich der emotionalen Wirkung auf Sie auf folgenden Skalen bewerten:

...

Uns interessiert, welchen emotionalen Eindruck ein Audio-Clip hervorruft. Es gibt keine richtigen oder falschen Antworten, und ebenso geht es nicht darum, Sie als Person zu bewerten. Alle erfassten Daten werden anonymisiert weiterverarbeitet.

Zu Beginn des eigentlichen Versuchs erhalten Sie noch einmal ausführliche Instruktionen. Diese Kurzbeschreibung dient in erster Linie als Vorab-Information für die Einverständniserklärung.

Falls Sie dazu Fragen haben, wenden Sie sich bitte an den Versuchsleiter.

Einverständniserklärung

Ich bin über den Sinn und die Durchführung der Untersuchung informiert worden und gebe hiermit mein Einverständnis zur elektronischen Speicherung und anonymisierten Weiterverarbeitung meiner Daten.

Anhang B

Beschreibung von Beispielfragebögen

AttrakDiff (Hassenzahl, Burmester, Koller 2003)

Mit diesem Fragebogen werden Qualitätsmerkmale untersucht, die über die reine *Ease of Use* hinausgehen. So werden neben der wahrgenommenen pragmatischen Qualität auch Aspekte der hedonischen Qualität und der Attraktivität des Systems insgesamt untersucht. Der AttrakDiff enthält 28 Items, die in Form eines semantischen Differentials dargeboten werden. Die Nutzer beurteilen das System mit Hilfe gegensätzlicher Adjektivpaare auf einer siebenstufigen Skala. Die 28 Items werden verschiedenen Faktoren bzw. Dimensionen zugeordnet. Hassenzahl et al. (2003) geben vier Dimensionen an, denen die einzelnen Items zugeordnet werden:

- Die *Pragmatische Qualität* (PQ) beschreibt die Benutzerfreundlichkeit des Systems und den Grad, mit dem Nutzer ihre Ziele und Aufgaben mit dem System bewältigen können.
- Die *Hedonische Qualität – Stimulation* (HQ-S) verdeutlicht, in wie weit das System den Nutzer dabei unterstützen kann, sich weiterzuentwickeln, indem es neuartige und anregende Konzepte bietet.
- Die *Hedonische Qualität – Identität* (HQ-I) gibt an, wie gut sich Nutzer mit dem System identifizieren können.
- Die *Attraktivität* (ATT) ist eine globale Dimension, die die wahrgenommene Qualität insgesamt beschreibt.

System Usability Measurement Inventory (SUMI, Kirakowski & Corbett 1996)

Der SUMI wurde 1990 in erster Linie für grafische Schnittstellen entwickelt. Er enthält 50 Aussagen, die auf einer dreistufigen Skala bewertet werden sollen und dient dazu, die Qualität einer Software aus Nutzerperspektive zu untersuchen. Die 50 Items verteilen sich auf die 5 Skalen:

- *Efficiency* beschreibt das Ausmaß, mit dem der Nutzer das Gefühl hat, von der Software unterstützt zu werden.
- *Affect* misst die allgemeine emotionale Reaktion der Nutzer auf die Software.
- *Helpfulness* beschreibt, wie selbsterklärend die Software ist, und spezifischere Aspekte wie die Angemessenheit der Hilfe und der Dokumentation.
- *Control* misst, inwieweit der Nutzer bei der Aufgabenbearbeitung das Gefühl hat, die Software zu kontrollieren, und nicht umgekehrt das Gefühl hat, vom System kontrolliert zu werden.
- *Learnability* misst, wie schnell und wie gut der Nutzer das Gefühl erreicht, das System zu beherrschen und falls notwendig neue Funktionen zu erlernen

Subjective Assessment of Speech System Interfaces (SASSI, Hone & Graham 2000)

Auf Grundlage des SUMI wurde für Sprachdialogsysteme der Fragebogen SASSI entwickelt. Er enthält 34 Items, die auf einer fünfstufigen Likert-Skala bewertet werden. Die einzelnen Items werden dann zu verschiedenen Faktoren zusammengefasst. Die Entwickler des SASSI ermittelten folgende sechs Faktoren:

- *System Response Accuracy* beschreibt, inwieweit Nutzereingaben vom System richtig verstanden wurden, und ob Intention und Erwartungen des Nutzers erfüllt sind.
- *Likeability* enthält die subjektive Meinung des Nutzers über das System und dessen Empfinden während der Interaktion.
- *Cognitive Demand* sagt etwas über die vom Nutzer empfundene geistige Anstrengung während der Interaktion aus.
- *Annoyance* erfasst Urteile darüber, wie stark negative Gefühle beim Nutzer durch die Interaktion hervorgerufen werden.
- *Habitability* gibt Auskunft darüber, wie weit die Vorstellungen des Nutzers vom System mit denen des Entwicklers übereinstimmen. Dies beeinflusst die Fähigkeit des Nutzers, den Umgang mit dem System schnell und problemlos zu erlernen.
- *Speed* bezieht sich auf die vom Nutzer wahrgenommene Geschwindigkeit der Interaktion und des Systems.

System Usability Scale (SUS, Brooke 1996)

Der eindimensionale SUS Fragebogen ist sehr kurz gehalten und misst global die *Ease of Use*. Er umfasst 10 Items die jeweils auf einer 5-stufigen Likertskala beantwortet werden. Mittlerweile wird eine zweifaktorielle Struktur des SUS mit den Dimensionen *Ease of Use* und *Learnability* diskutiert (Lewis & Sauro, 2009).

Beispielitems

Item	Fragebogen	Aspekt gemäß Taxonomie	Skala gemäß Autoren
Umständlich vs. direkt	AttrakDiff	<i>Efficiency</i>	<i>PQ</i>
Schön vs. hässlich	AttrakDiff	<i>Usability</i>	<i>Attraktivität</i>
Lahm vs. fesselnd	AttrakDiff	<i>Joy of Use</i>	<i>HQS</i>
Stilvoll vs. stillos	AttrakDiff	<i>Joy of Use</i>	<i>HQI</i>
Es bedarf zu vieler Einzelschritte, um eine Funktion auszuführen.	SUMI	<i>Efficiency</i>	<i>Efficiency</i>
Die Arbeit mit der Software macht mir Spaß.	SUMI	<i>Joy of Use</i>	<i>Affect</i>
Der erste Umgang mit der Software ist/war voller Probleme.	SUMI	<i>Intuitivity</i>	<i>Helpfulness</i>
Ich habe das Gefühl, die Software zu beherrschen, wenn ich sie benutze.	SUMI	<i>Ease of Use</i>	<i>Control</i>
Ich muss regelmäßig Einzelheiten im Handbuch nachschlagen.	SUMI	<i>Learnability</i>	<i>Learnability</i>
Das System ist genau.	SASSI	<i>Ease of Use</i>	<i>System Response Accuracy</i>
Das System ist angenehm.	SASSI	<i>Joy of Use</i>	<i>Likeability</i>
Ich musste mich sehr auf die Interaktion mit dem System konzentrieren.	SASSI	<i>Efficiency</i>	<i>Cognitive Demand</i>
Die Interaktion mit dem System ist frustrierend.	SASSI	<i>Joy of Use</i>	<i>Annoyance</i>
Es kann schnell passieren, dass man nicht weiß an welcher Stelle man sich in der Interaktion mit dem System befindet.	SASSI	<i>Intuitivness</i>	<i>Habitability</i>
Das System reagierte zu langsam.	SASSI	<i>Efficiency</i>	<i>Speed</i>
Ich finde das System ist einfach zu benutzen	SUS	<i>Ease of Use</i>	/

Anhang C

Aufgaben- und Prüfliste nach VDE-ITG-Richtlinie 9.4. 1.01

Aufgabenliste „Anrufbeantworter“

Aufgabentyp A) Grundfunktionen (Kenntnis des generellen Benutzungsprinzips ist ausreichend)

1. Anrufbeantworterfunktion einschalten und ausschalten
2. Anruf übernehmen, nachdem der Anrufbeantworter aktiviert wurde
3. Mithören und Lautstärke regeln
4. Neue Nachricht anhören und löschen
5. Nachrichten gezielt anhören und löschen
6. Private Notizen aufsprechen und abfragen

Aufgabentyp B) Einstellfunktionen (Hinzuziehung der Bedienungsanleitung zulässig)

1. Eigene Ansage aufsprechen, anhören und speichern
2. Anzahl der Rufzeichen vor Aktivierung der Anrufbeantworterfunktion einstellen
3. Mithörfunktion ein- und ausschalten
4. „Nachricht weiterleiten“ einstellen
5. Uhrzeit und Datum für Zeitstempel einstellen
7. Sicherheitscode für Fernabfrage einstellen

Aufgabentyp C) Fernabfrage – nur relevant, wenn Funktion vorhanden ist (Benutzung ohne Bedienungsanleitung)

1. In das System einwählen und Berechtigung prüfen
2. Anrufbeantworterfunktion von Ferne ein- und ausschalten
3. Neue Nachrichten von Ferne anhören und löschen
4. Ansage ändern

Vorschlag für die Anzahl der auszuwählenden Prüfaufgaben:

Typ A) 4 Aufgaben

Typ B) 3 Aufgaben

Typ C) 3 Aufgaben

Prüfliste (generisch)

ITG-Prüfverfahren		Gerät:									
		Ergebnisblatt für:									
		Prüfer:									
		Datum:									
Aktionen											
1.		6.									
2.		7.									
3.		8.									
4.		9.									
5.		10.									
Urteile und Begründungen		Antworten für die einzelnen Aktionen (Urteil/Begründung):									
Prüffragen:		1	2	3	4	5	6	7	8	9	10
0.	Bekannt/erkennbar, daß Vorbereitung für A1 erforderlich?										
Analyse der korrekten Aktion A1											
1.	Klarer, zielführender Hinweis auf diese Aktion?										
2.	Ist diese Aktion durch Ausprobieren zu ermitteln?										
3.	Trägt diese Aktion erkennbar zum Erreichen des Ziels bei?										
4.	Genügend Zeit für Auswahl und Durchführung dieser Aktion?										
Analyse der Situation nach der Aktion A1											
a) Fehlerfreie Aktion(en)											
5.	Erkennbar, welcher Schritt zum Ziel erfolgte?										
6.	Erkennbar, ob weitere Aktion erforderlich?										
b) Fehlerhafte Aktion(en)											
Fehler 1:											
7.1.1	Erkennbar, daß Fehler gemacht wurde?										
7.1.2	Erkennbar, wie weiter vorzugehen ist?										
Fehler 2:											
7.2.1	Erkennbar, daß Fehler gemacht wurde?										
7.2.2	Erkennbar, wie weiter vorzugehen ist?										
Fehler 3:											
7.3.1	Erkennbar, daß Fehler gemacht wurde?										
7.3.2	Erkennbar, wie weiter vorzugehen ist?										
Gesamtbeurteilung der Aktionen A1 bis An											
8.	Entsprechen die Aktionen A1 bis An den Erwartungen?										
9.	Aufwand für die Aktionen A1 bis An angemessen?										
Fehlerfälle											
		Erläuterung der Codierung (Seite 1):									
		Urteile:									
F11											
F12		J	Ja								
F13		E	Eingeschränkt Nein								
F21		N	Nein								
F22		-	Frage nicht zutreffend								
F23		?	Frage unklar								
F31											
F32											
		Begründungen:									
F33		a	Hinweis/Rückmeldung fehlt								
F41		b	Hinweis/Rückmeldung unklar								
F42		c	Hinweis/Rückmeldung irreführend								
F43		d	Hinweis/Rückmeldung unnötig								
F51											
F52		e	Antwortzeit des Systems zu groß								
F53		f	Zeitfenster für Benutzer zu klein								
F61											
F62		g	Fehlererkennung schwierig/unmöglich								
F63		h	Fehlerbehebung schwierig/unmöglich								
F71		i	Fehlertoleranz ungenügend								
F72											
F73		j	Verstoß gegen Standard, Norm								
F81		k	Inkonsistenz (Begriffe usw.)								
F82		l	Anordnung nicht plausibel								
F83											
F91		m	Handhabung umständlich								
F92		n	Handhabung zeitraubend								
F93		o	Gedächtnisbelastung zu hoch								
F101		p	Abbruchmöglichkeiten ungenügend								
F102		*	Sonstiges...								
F103		.	kein Eintrag								

Anhang D

Instruktion und Protokoll – Heuristische Evaluation

Instruktion

Liebe(r) Versuchsteilnehmer(in),

[...] Nun bitten wir Sie, verschiedene Aufgaben zu bearbeiten und Ihnen dabei auffallende Probleme zu protokollieren. Weiterhin ist es notwendig, das Problem zunächst zu beschreiben und zu begründen. Anschließend geben sie uns bitte eine Einschätzung der Relevanz des Problems. Auch hier soll der Fokus ihrer Beurteilung auf dem System liegen.

Behalten sie während der Protokollierung und Bewertung bitte die folgenden Fragen im Hinterkopf:

- Wird der Benutzer versuchen, den richtigen Effekt zu erzielen?
- Wird der Benutzer erkennen, dass die korrekte Aktion zur Verfügung steht?
- Wird der Benutzer eine Verbindung herstellen zwischen der korrekten Aktion und dem gewünschten Effekt?
- Wenn die korrekte Aktion ausgeführt worden ist: Wird der Benutzer den Fortschritt erkennen?

Als Orientierung sollen Ihnen dabei die folgenden Kriterien dienen:

- Natürlichkeit und Ergonomie
Bsp.: Sind erforderliche Bewegungsabläufe natürlich und ergonomisch?
- Konsistenz und Logik
Bsp.: Sind die Dialoge logisch und widerspruchsfrei?
- Rückmeldung des Systemzustandes
Bsp.: Ist klar, was das System gerade tut?
- Fehlertoleranz/Fehlerprophylaxe/Fehlererkennung/Fehlerbehebung
Bsp.: Ist klar, was das System gerade tut?
- Abgeschlossenheit des Dialogs
Bsp.: Ist klar, ob eine weitere Eingabe zu tätigen ist?
- Erlernbarkeit
Bsp.: Sind die Funktionen schnell zu erlernen?
- Angemessenheit der Modalitäten für Aufgabe und Kontext
Bsp.: Ist die Aufgabe mit der jeweiligen Modalität effizient und effektiv zu bewältigen?

Fallen Ihnen weitere, von diesen Dimensionen nicht abgedeckte Probleme auf, sind Sie gebeten, auch diese nach dem dargestellten Muster zu erfassen und zu beurteilen. An diese Testphase schließt sich eine kurze Pause für Sie an. Zum Schluss sind sie gebeten, die Ergebnisse mit den anderen Teilnehmern zu diskutieren.

Uns sind Ihr spontanes Verhalten und Ihre offene Meinung über das System wichtig. Bei Fragen oder Unklarheiten können sie sich jederzeit an die Versuchsleiter wenden.

Und nun: Viel Spaß beim Versuch!

Beispielseite Protokoll

Bewertungsdimension	Problembeschreibung	Begründung	Problembewertung ²
	_____ _____ _____ _____ _____ _____ _____ _____	_____ _____ _____ _____ _____ _____ _____ _____	<input type="checkbox"/> Kein Problem <input type="checkbox"/> Kosmetisches Problem <input type="checkbox"/> Kleines <i>Usability</i> -Problem <input type="checkbox"/> Großes <i>Usability</i> -Problem <input type="checkbox"/> <i>Usability</i> Katastrophe
Natürlichkeit und Ergonomie	_____ _____ _____ _____ _____ _____ _____ _____	_____ _____ _____ _____ _____ _____ _____ _____	<input type="checkbox"/> Kein Problem <input type="checkbox"/> Kosmetisches Problem <input type="checkbox"/> Kleines <i>Usability</i> -Problem <input type="checkbox"/> Großes <i>Usability</i> -Problem <input type="checkbox"/> <i>Usability</i> Katastrophe

² 0 = Kein Problem, 1 = Nur kosmetisches Problem: bedarf nicht zwingend einer Ausbesserung, 2 = Kleines *Usability*-Problem: Ausbesserung sollte niedrige Priorität haben,

3 = Großes *Usability*-Problem: Ausbesserung sollte hohe Priorität haben, 4 = *Usability*-Katastrophe: Ausbesserung muss zwingend erfolgen

Anhang E

Fallbeispiele

Fallbeispiel 1 – Evaluation von Menüstrukturen auf interaktiven Oberflächen

Hesselmann, T. Flöring, S., Schmitt M. (2009) Stacked Half-Pie Menus – Navigating Nested Menus on Interactive Tabletops, Proceedings of ITS 2009

Funktionsbereich: Informationsabrufanwendung

Nutzungsszenario: Eindeutige Szenarien festzulegen ist hier nicht möglich, da es sich um einen ersten Prototypen handelt, bei dem verschiedene Nutzungskontexte denkbar sind.

Ein-/Ausgabemodalität: unimodales GUI-basiertes System mit Touchscreen

Entwicklungsphase im Usability Engineering Lifecycle: Erste empirische Evaluation

Erfasste Qualitätsaspekte: Cognitive Workload, Ease of Use

Angewendete Verfahren: Fragebögen, Lautes Denken

Fragestellung: Bewertet wurde die Menüstruktur einer interaktiven Oberfläche. Ziel war es festzustellen, ob das System gut genug ist (vgl. Abschnitt 4.1). Entsprechend wurden weder Vergleiche zu anderen Systemen angestellt noch die Einhaltung von Standards überprüft.

Vorgehen: Es wurde ein nutzerzentrierter Labor-Test mit 12 Personen durchgeführt. Die Alterstruktur der Teilnehmer war recht homogen. So wurden Nutzer über 34 und unter 14 Jahren nicht berücksichtigt. Entsprechend sind die Evaluationsergebnisse nicht ohne weiteres auf andere Altersgruppen übertragbar. Die Teilnehmer wurden zunächst mit dem System bekannt gemacht. Anschließend wurden verschiedene Aufgaben mit dem System bearbeitet. Hier waren die Teilnehmer dazu angehalten, mittels lauten Denkens die Aufgabenbearbeitung und dabei auftretende Probleme zu kommentieren. Zur Erfassung der mentalen Beanspruchung wurden die Teilnehmer nach Abschluss jeder Aufgabe darum gebeten, den Nasa-TLX Fragebogen (Hart & Staveland, 1988) auszufüllen. Nach Bearbeitung aller Aufgaben wurde die *Ease Of Use* des Systems über den SUS (*System Usability Scale*, Brooke 1996) abgefragt. Ergebnis der Studie waren die beim lauten Denken identifizierten, d.h. durch die Probanden geäußerten Probleme, sowie die quantitativen Daten der Fragebögen. Anzumerken ist, dass der zu bewertende Prototyp nur einen Aufgabentypus, das Navigieren durch hierarchische Strukturen, unterstützt. Entsprechend entfiel die Bestimmung weiterer Haupt- und Nebenaufgaben.

Fallbeispiel 2 – Evaluation einer webbasierten, mobilen Serviceplattform

Porta, D., Sonntag, D. & Neßelrath, R. (2009) A Multimodal Mobile B2B Dialogue Interface on the iPhone, In Proceedings of 4th Workshop on Speech in Mobile and Pervasive Environments (SiMPE).

Funktionsbereich: Informationsabrufanwendung

Nutzungsszenario: Mobile Nutzung im beruflichen Kontext

Ein-/Ausgabemodalität: Multimodales GUI-basiertes System mit Touchscreen und Sprachein- und -ausgabe

Entwicklungsphase im Usability Engineering Lifecycle: Erste empirische Evaluation

Erfasste Qualitätsaspekte: *Ease of Use*

Angewendete Verfahren: Protokollanalysen (Beobachtung), selbst-konstruierter Fragebogen

Fragestellung: In dieser ersten Studie sollte in erster Linie die Effektivität der Plattform erfasst werden. Wie schon im ersten Fallbeispiel wurden demnach weder Vergleiche zu anderen Systemen angestellt noch die Einhaltung von Usability-Standards überprüft. Im nächsten Schritt ist eine vergleichende Studie allerdings geplant (vgl. Abschnitt 4.1).

Vorgehen: An dem Labortest nahmen 11 Personen teil. Alle Teilnehmer waren Studenten, und obwohl Altersangaben nicht vorliegen ist anzunehmen, dass es sich überwiegend um Teilnehmer im Alter von 19 bis 35 Jahren handelte. Die Generalisierbarkeit der Ergebnisse auf andere Altersgruppen ist fraglich. Testaufgaben waren alle Aufgaben einer von den Entwicklern als relevant beurteilten Interaktionssequenz. Nach einer 5-minütigen Explorationsphase wurden die Probanden gebeten 12 Aufgaben zu bearbeiten. Pro Aufgabe waren zwei Versuche zur Lösung der Aufgabe vorgesehen. Wurde die Aufgabe im ersten Anlauf nicht erfolgreich bearbeitet unterstützte der Versuchsleiter die Probanden mit Lösungshinweisen. Ausgewertet wurden die Rate der erfolgreich bearbeiteten Aufgaben sowie der Fragebogen.

Fallbeispiel 3 – Evaluation eines multimodalen Rauminformationssystems

Schaffer, S., Seebode, J., Wechsung, I., Metze, F. and Möller, S. (2009). *Benutzerstudien zur Bewertung multimodaler, interaktiver Anzeigetafeln in unterschiedlichen Entwicklungsstufen. Workshop-Proceedings der Tagung Mensch und Computer 2009. Logos Berlin, 22-27.*

Funktionsbereich: Informationsabrufanwendung

Nutzungsszenario: Festinstalliertes System im beruflichen Kontext

Ein-/Ausgabemodalität: Multimodales GUI-basiertes System mit Touchscreen, Spracheingabe und Gesichtserkennung

Entwicklungsphase im Usability Engineering Lifecycle: Expertenevaluation, empirische Evaluation, iteratives Design

Erfasste Qualitätsaspekte: *Ease of Use, Joy of Use, Interaction Performace*

Angewendete Verfahren: *Cognitive Walkthrough* mit Interview, Experiment mit Fragebögen und Protokollanalysen (Audioaufnahmen, Log-Daten der Spracherkennung)

Fragestellung: Ein unimodales Rauminformationssystem sollte durch Erweiterung um eine Spracheingabe (deutsch) verbessert werden. Ziel war es festzustellen, ob die Weiterentwicklungen des Systems zu einer Steigerung der vom Nutzer wahrgenommenen Qualität führen. Es wurden drei Systemvarianten miteinander verglichen: (1) das unimodale Basissystem, (2) ein System mit integrierter Sprachsteuerung nach der ersten Weiterentwicklung, (3) ein System mit verbesserter Sprachsteuerung und Gesichtserkennung nach der zweiten Weiterentwicklung.

Vorgehen: Um möglichst viele *Usability*-Probleme vor dem Experiment mit realen Nutzern zu beseitigen, wurde ein *Cognitive Walkthrough* mit sechs *Usability*-Experten durchgeführt. Dabei waren sechs Aufgaben (im Wesentlichen wechseln zwischen verschiedenen Ansichten und suchen nach Buchungen von Mitarbeitern und Räumen) einmal per Touchscreen und einmal per Sprache zu lösen und die aufgetauchten *Usability*-Probleme zu protokollieren. Probleme, die ausgebessert werden konnten, wurden vor der Durchführung des Experimentes mit realen Nutzern behoben.

Danach wurde ein Labor-Test mit 19 weiblichen und 17 männlichen deutschen Muttersprachlern im Alter von 21 bis 39 Jahren durchgeführt. Die Teilnehmer wurden zunächst anhand einer schriftlichen Versuchsbeschreibung mit dem System bekannt gemacht. Anschließend waren dieselben sechs Aufgaben mit dem unimodalen Basissystem zu lösen, die auch schon bei der Expertenevaluation bearbeitet wurden. Nach der Bearbeitung der Aufgaben wurden die Teilnehmer darum gebeten, die Fragebögen AttrakDiff und SASSI (Hone & Graham, 2000) auszufüllen. Beide Fragebögen dienen der Erfassung von *Joy of Use* und *Ease of Use*. SASSI wurde für unimodale sprachbasierte Systeme entworfen. Daher wurden drei Items für Touchscreen und multimodale Bedienung angepasst. Dann waren die Aufgaben nochmals mit dem multimodalen System zu bearbeiten. Dabei wurden Sprachdaten und die Hypothesen der Spracherkennung aufgezeichnet, um daraus die Leistung des Spracherkennungsmoduls abzuleiten. Abschließend wurden die Teilnehmer ein weiteres Mal darum gebeten, die Fragebögen auszufüllen, um das weiterentwickelte System zu bewerten.

Die Ergebnisse beider Studien wurden für das System der zweiten Entwicklungsstufe umgesetzt. Unter anderem wurde eine Gesichtserkennung implementiert. Zur Evaluation des Systems mit Gesichtserkennung erfolgte auch hier zunächst ein *Cognitive Walkthrough* mit sie-

ben *Usability*-Experten. Dabei hatten drei der Experten bereits an der vorangegangenen Evaluation teilgenommen. Die Aufgaben wurden aus der ersten Studie übernommen, um vergleichbare Ergebnisse zu erhalten. Die Teilnehmer wurden angehalten, besonders auf mögliche fehlende Sprachbefehle und die Auswirkungen der Kameraintegration zu achten. Wiederrum wurden die Probleme, die ausgebessert werden konnten, vor der Durchführung des Experimentes mit realen Nutzern behoben.

Wieder wurde ein Labor-Test mit 18 weiblichen und 17 männlichen deutschen Muttersprachlern im Alter von 22 bis 39 Jahren durchgeführt. Der Ablauf und die Datenerhebung des Experiments entsprach größtenteils der vorangegangenen Untersuchung in der ersten Entwicklungsstufe. Subjektive Daten wurden mit Hilfe des AttrakDiff und des für multimodale Systeme modifizierten SASSI-Fragebogens erfasst. Zusätzlich wurden Video- und Log-Daten der Gesichtserkennung aufgezeichnet, um daraus die Leistung des Gesichtserkennungsmoduls abzuleiten. Wiederrum erfolgte eine Adaption des Systems gemäß den Evaluationsergebnissen. Das System selbst wurde nach Abschluss der Evaluationen erfolgreich im laufenden Betrieb eingesetzt.

Fallbeispiel 4 – Evaluation einer multimodalen Fernbedienung

Wechsung, I., & Naumann, A. B. (2009). *Evaluating a Multimodal Remote Control: The Interplay Between User Experience and Usability*. In *Proceedings of the First IEEE International Workshop on Quality of Multimedia Experience (QoMEX' 2009)* (pp. 19-22). New York: IEEE.

Funktionsbereich: Medienaufzeichnungs- und Abspielanwendung

Nutzungsszenario: Festinstalliertes System im privaten Kontext

Ein-/Ausgabemodalität: Multimodales System mit Tasten, Spracheingabe und Bewegungssteuerung

Entwicklungsphase im Usability Engineering Lifecycle: Sämtliche Phasen

Erfasste Qualitätsaspekte: *Ease of Use, Joy of Use*

Angewendete Verfahren: Heuristische Evaluation, Experiment und Feldstudie jeweils mit Fragebögen und Protokollanalysen

Fragestellung: Für eine IPTV-basiertes Unterhaltungssystem wurde eine neue multimodale Fernbedienung entwickelt. Anforderung war dabei, dass keine Änderungen am Gesamtsystem vorgenommen werden. Ziel war es, die bisherige Fernbedienung zu verbessern, entsprechend wurden zum Teil vergleichende Untersuchungen durchgeführt.

Vorgehen: In einem Expertenworkshop wurden die notwendigen Funktionalitäten und das multimodale Bedienkonzept der neuen Fernbedienung festgelegt. Berücksichtigt wurden Funktionalitäten der bisherigen, unimodalen Fernbedienung sowie modalitätsspezifische Vor- und Nachteile. Danach wurde in mehrere Iterationen eine Spezifikation erarbeitet. Dabei wurden drei verschiedene Designvorschläge als Modelle umgesetzt. Anhand ergonomischer und gestalterischer Kriterien wurde nach Expertendiskussion ein Modell ausgewählt, welches im Anschluss als erster Prototyp realisiert wurde. Dieser Prototyp wurde einer Expertenevaluation mit fünf Teilnehmern unterzogen. Hier verwendete Methoden umfassten den *Cognitive Walkthrough* mit Elementen der Heuristischen Evaluation (siehe Anhang D). Sämtliche Aufgaben wurden mit allen Eingabemodalitäten bearbeitet. Unter Berücksichtigung der in der Expertenevaluation aufgetretenen Probleme wurde im nächsten Schritt eine Nutzerstudie durchgeführt. Die Studie war als klassisches Experiment angelegt, die Versuchsumgebung wurde jedoch so natürlich wie möglich gehalten, um eine Wohnzimmeratmosphäre zu erzeugen.

Getestet wurden Haupt- und Nebenfunktionen der multimodalen Fernbedienung. Als Kontrollbedingung wurde zudem die vorherige, unimodale Fernbedienung mit untersucht. Es wurden junge und alte Nutzer getestet, um Alterseffekte zu ermitteln. Insgesamt nahmen 16 junge und 16 ältere Nutzer an der Untersuchung teil. Das Geschlecht der Nutzer wurde in beiden Gruppen ausbalanciert. Die Hälfte der Nutzer testete die neue Fernbedienung, die andere Hälfte die alte.

Nach der Begrüßung der Teilnehmer wurde das Unterhaltungssystem sowie die Fernbedienung kurz erklärt. Im Anschluss wurden die Probanden gebeten, die Fernbedienung kurz zu testen, dabei war das Unterhaltungssystem jedoch noch nicht in Betrieb. Danach sollten neun verschiedene Aufgaben bearbeitet werden. Die Aufgaben entsprachen größtenteils den Aufgaben der Expertenevaluation. Die Eingabemodalität konnte von den Teilnehmern selbst gewählt werden. Erhoben wurde die Aufgabenbearbeitungszeit sowie - nach Abarbeitung sämt-

licher Aufgaben – der SUS- (Brooke, 1996) sowie der AttrakDiff- Fragebogen (Hassenzahl et al., 2003). Die Auswahl der Fragebögen erfolgte nach mehreren Gesichtspunkten. Beide Fragebögen wurden im selben Rahmenprojekt schon zur Evaluation anderer Systeme genutzt, somit wurde eine gewisse Vergleichbarkeit der Ergebnisse sichergestellt. Weiterhin ist insbesondere der AttrakDiff-Fragebogen auch für multimodale Schnittstellen geeignet. Während der SUS die *Ease of Use* des Systems misst, werden mit dem AttrakDiff auch hedonische Aspekte abgedeckt.

Nach Abschluss der Laborstudie wurden die Ergebnisse in einer weiteren Iteration für den nächsten Prototyp verwendet. Der zweite Prototyp wurde in einer Feldstudie mit realen Nutzern über mehrere Wochen getestet. Detaillierte Interviews mit den Nutzern wurden beim Start sowie am Ende der Feldphase durchgeführt. Weiterhin waren die Teilnehmer gebeten, ein Tagebuch zu führen, um die Nutzung und insbesondere auftretende Probleme oder besonders gelungene Elemente zu dokumentieren. Zusätzlich wurden während des gesamten Zeitraums Log-Daten angelegt. Ziel war es, zum einen Nutzungsmuster für einzelne Modalitäten zu erheben, und zum anderen technische Probleme zu erfassen.

Nach Abschluss und Auswertung der Feldphase wurde schließlich die Produktentwicklung angestoßen.

Anhang F

Liste der englischen Termini und deren Übersetzung

- Acceptability: Akzeptanz
- Adaptive Control of Thought–Rational (ACT-R): Eine Architektur zur Modellierung der menschlichen kognitiven Prozesse
- Aesthetics: Ästhetik
- Agent Factors: technische (nicht funktionale) Qualitätsfaktoren des Systems
- Analysis: Analyse
- Appeal: Eindruck, Wirkung (setzt sich aus Ästhetik und Persönlichkeit eines Systems zusammen)
- Apperance: äußerliches Erscheinungsbild des Systems
- Avatar: Avatar, eine virtuelle Figur
- Cognitive Complexity Theory (CCT): „Kognitive Komplexitätstheorie“, Methode zur Aufgabenanalyse
- Cognitive Walkthrough: kognitiver Durchgang, Durchdenken eines Problems
- Cognitive Workload: kognitive Beanspruchung
- Cooperativity: Kooperativität
- Ease of Use: Einfachheit der Nutzung
- Effectiveness: Effektivität
- Efficiency: Effizienz
- Empirical Testing: empirische Untersuchung
- Expert Evaluation: Expertenevaluation
- Feedback from field: Rückmeldung aus dem Feld
- Fitt's Law: „Fitts' Gesetz“, Modell zur Bestimmung von Dauer und Genauigkeit von Arm- und Mausbewegungen
- Functional Factors: funktionale Qualitätsfaktoren
- Goal Operators Methods Selection Rules (GOMS): Methode, die unterschiedliche Modelle der Mensch-Computer-Interaktion z. B. zur Bestimmung von Ausführungszeiten integriert
- Hendonic: hedonisch
- Hierarchical Task Analysis: „Hierarchische Aufgabenanalyse“, Methode zur Aufgabenanalyse
- Icon: Symbolbild, Piktogramm
- Influencing Factors: Einflussfaktoren
- Input Quality: Eingabequalität
- Interaction Performance: umfasst Leistungsindikatoren zur Beschreibung der Interaktion zwischen Nutzer und System.
- Intuitivity: Intuitivität
- Iterative Design: iteratives Design, zyklisches Re-Design
- Joy of Use: „Spaß bei der Nutzung“, das bewusst positive Erleben der Qualität der Interaktion
- Keystroke Level Model: Modell zur Bestimmung von Ausführungszeiten bei der Tastaturbedienung
- Learnability: Erlernbarkeit
- Output Quality: Ausgabequalität
- Pramatic: pragmatisch
- Prototyping: Entwicklung eines Prototypen

- Quality of Experience: Qualitätsbewertung aus Nutzersicht, umfasst die wahrgenommenen Qualitätsaspekte eines Systems
- Quality of Service: Die Servicequalität, umfasst Aspekte der Leistung des Systems.
- State, Operator And Result Modell (SOAR): Eine Architektur zur Modellierung der menschlichen kognitiven Prozesse
- System Personality: Persönlichkeit des Systems
- Task Factors: Aufgabenbezogene Qualitätsfaktoren
- Thinking Aloud: lautes Denken
- Usability Engineering Lifecycle: Berücksichtigung der Methoden des Usability Engineering in allen Phasen der Systementwicklung zur optimalen Systemgestaltung
- Usability Testing: Untersuchung der Gebrauchstauglichkeit
- Usability: Gebrauchstauglichkeit, manchmal auch „Benutzerfreundlichkeit“
- Usability-Engineering: Methoden zur Messung und Bewertung der Gebrauchstauglichkeit interaktiver Systeme
- Usefulness: Nützlichkeit
- User Experience: Nutzererlebnis, Nutzungserlebnis
- Utility: Nutzen (eines Systems)

Literatur

- Anderson, J. R., Lebiere, C.: Atomic Components of Thought. Hillsdale NJ: Erlbaum, 1998.
- Annett, J.: Hierarchical Task Analysis. In: Holnagel, E., Handbook of Cognitive Task Design, Chapter 2, pp17-35, Mahwah NJ: Lawrence Erlbaum, 2003.
- Baber, C.: Computing in a Multimodal World. In: Proc. UAHCI'01 (HCI International 2001), New Orleans, August 8-10, Mahwah (NJ): Lawrence Erlbaum, pp. 232-236, 2001.
- Bernsen, N. O.: Defining a Taxonomy of Output Modalities From an HCI Perspective. Computer Standards and Interfaces, 18, pp. 537-553, 1997.
- Bevan, N.: What Is the Difference Between the Purpose of Usability and User Experience Evaluation Methods. In: Proceedings of the Workshop UXEM'09 (INTERACT'09), Uppsala, Sweden, 2009.
- Bortz, J., Döring, N.: Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler, 4. Auflage. Berlin: Springer, 2006.
- Brooke, J.: SUS: A "Quick and Dirty" Usability Scale. In: Usability Evaluation in Industry, P. Jordan, B. Thomas, B. Weerdmeester, I. McClelland (eds.), London: Taylor & Francis, pp. 189-194, 1996.
- Charwat, H. J.: Lexikon der Mensch-Maschine-Kommunikation. München: R. Oldenbourg Verlag, 1992.
- Davis, F.: Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. MIS Quarterly, Vol. 13 (3), pp. 319-340, 1989.
- Deutsches Institut für Normung, DIN 55350-11:2008-05 (D): Begriffe zum Qualitätsmanagement – Teil 11: Ergänzung zu DIN EN ISO 9000:2005. Deutsches Institut für Normung, Berlin: Beuth-Verlag.
- Dix, A., Finlay, J., Abowd, G., Beale, R.: Human-Computer Interaction, 3rd edition. New York: Prentice Hall, 2004.
- Hart, S. G., Staveland, L. E.: Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In: Human Mental Workload, P. A. Hancock and N. Meshkati (eds.), North-Holland: Elsevier Science Publishers, pp. 139-184, 1988.
- Hassenzahl, M.: Hedonic, Emotional, and Experiential Perspectives on Product Quality. In: C. Ghaoui (Ed.), Encyclopedia of Human Computer Interaction. Idea Group. Pp. 266-272, 2006.
- Hassenzahl, M.: Aesthetics in Interactive Products: Correlates and Consequences of Beauty. In: Product Experience (H. N. J. Schifferstein & P. Hekkert, eds.), San Diego, CA: Elsevier, pp. 287-302, 2008.

- Hassenzahl, M., Burmester, M., Koller, F.: AttrakDiff: Ein Fragebogen zur Messung wahrgenommener hedonischer und pragmatischer Qualität [A questionnaire for measuring perceived hedonic and pragmatic quality]. In: Mensch & Computer 2003. Interaktion in Bewegung, J. Ziegler, G. Szwillus (eds.), Stuttgart: B.G. Teubner, pp. 187-196, 2003.
- Hassenzahl, M., Monk, A.: The Inference of Perceived Usability From Beauty. *Human-Computer Interaction*, Vol. 25 (3), pp. 235-260, 2010.
- Hegner, M.: Methoden zur Evaluation von Software. IZ-Arbeitsbericht Nr. 29, Hemsens, 2003.
- Hekkert, P.: Design Aesthetics: Principles of Pleasure in Product Design. *Psychology Science*, Vol. 48, 2006.
- Hesselmann, T. Flöring, S., Schmitt M.: Stacked Half-Pie Menus – Navigating Nested Menus on Interactive Tabletops. In: Proc. of ITS 2009, 2009.
- Holzinger, A.: Usability Engineering Methods for Software Developers. *Commun. ACM*, Vol. 48 (1), pp. 71-74, 2005.
- Hone, K. S., Graham, R.: Towards a Tool for the Subjective Assessment of Speech System Interfaces (SASSI). *Natural Language Engineering*, Vol. 6 (3/4), pp. 287-303, 2000.
- Ilmonen, T.: Tools and Experiments in Multimodal Interaction. PhD Thesis, Helsinki University of Technology, 2006.
- Informationstechnische Gesellschaft im Verein Deutscher Ingenieure, VDE/ITG 9.4.1.01: ITG-Prüfverfahren zur Bewertung der Benutzungsfreundlichkeit von Telekommunikationsendgeräten und -systemen, Berlin – Offenbach: VDE Verlag, 1995.
- International Organization for Standardization, ISO 9241-11: Ergonomic Requirements for Office Work with Visual Display Terminals (VDTs) – Part 11: Guidance on Usability. Geneva, Switzerland: International Organization for Standardization, 1998.
- International Organization for Standardization, ISO 9241-210: Ergonomics of Human-system Interaction – Part 210: Human-centred Design for Interactive Systems. Geneva, Switzerland: International Organization for Standardization, 1998.
- Kieras, D., Polson, P.: An Approach to the Formal Analysis of User Complexity. *Int. J. Man-Machine Stud.*, Vol. 22(4), pp. 365-394, 1985.
- Kirakowski, J.: The Software Usability Measurement Inventory: Background and Usage. In: Jordan, P. W., Thomas, B., Weerdmeester, B. A. und McClelland, I. L. (eds.), *Usability Evaluation in Industry*, pp. 169-177, Brighton: Taylor & Francis, 1996.
- Kirakowski, J., Corbett, M.: SUMI: The Software Usability Measurement Inventory. *British Journal of Education Technology*, Vol. 24 (3), pp. 210-212, 1993.
- Lewis, J. R., Sauro, J.: The Factor Structure of the System Usability Scale. In: Proc. HCI International 2009, San Diego CA, 2009.

- Lin, H. X., Choong, Y., Salvendy, G.: A Proposed Index of Usability: A Method for Comparing the Relative Usability of Different Software Systems. *Behavior and Information Technology*, Vol. 16 (4/5), pp. 267-278, 1997.
- Mahlke, S., Lindgaard, G.: Emotional Experiences and Quality Perceptions of Interactive Products. In: *Proceedings of the 12th International Conference on Human-Computer Interaction: Interaction Design and Usability. Lecture Notes in Computer Science* (J. A. Jacko, ed.), Berlin: Springer, pp. 164-173, 2007.
- Mohs, C., Kindsmüller, M. C., Israel, J. H., Meyer, H. A. & IUUI Research Group: IUUI – Intuitive Use of User Interfaces: Auf dem Weg zu einer wissenschaftlichen Basis für das Schlagwort „Intuitivität“. *MMI-interaktiv Journal*, 11, pp. 75-84, 2006.
- Möller, S., Engelbrecht, K.-P., Kühnel, C., Wechsung, I., Weiss, B.: Evaluation of Multimodal Interfaces for Ambient Intelligence. In: *Human-Centric Interfaces for Ambient Intelligence* (H. Aghajan, R. López-Cózar Delgado and J. C. Augusto, eds.), pp. 347-370, Amsterdam: Elsevier, 2009.
- Möller, S., Krebber, J., & Smeele, J.: Evaluating the Speech Output Component of a Smart-Home System. *Speech Communication* 48 (1), pp. 1-27, 2006.
- Möller, S.: *Quality Engineering. Qualität kommunikationstechnischer Systeme*. Heidelberg: Springer, 2010.
- Newell, A.: *Unified Theories of Cognition*. Cambridge MA: Harvard University Press, 1990.
- Nielsen, J.: *Usability Engineering*. Boston: Academic Press, 1993.
- Nielsen, J.: Heuristic Evaluation. In: Nielsen, J., Mack, R. L. (eds.), *Usability Inspection Methods*, New York NY: John Wiley & Sons, 1994.
- Norman, D. A.: Emotion and Design: Attractive Things Work Better. *Interactions Magazine*, ix (4), pp. 36-42, 2002.
- Porta, D., Sonntag, D., Neßelrath, R.: A Multimodal Mobile B2B Dialogue Interface on the iPhone. In: *Proc. of 4th Workshop on Speech in Mobile and Pervasive Environments (SiMPE)*, 2009.
- Preece, J., Rogers, Y., Sharp, H., Benyon, D., Holland, S., Carey, T.: *Human-Computer Interaction*. Wokingham, UK: Addison-Wesley, 1994.
- Schaffer, S., Seebode, J., Wechsung, I., Metze, F., Möller, S.: Benutzerstudien zur Bewertung multimodaler, interaktiver Anzeigetafeln in unterschiedlichen Entwicklungsstufen. In: *Workshop-Proceedings der Tagung Mensch und Computer 2009*, Berlin: Logos, pp. 22-27, 2009.
- Schleicher, R., Trösterer, S.: The ‘Joy-of-Use’-Button: Recording Pleasant Moments While Using a PC. In: *Human-Computer Interaction – INTERACT 2009, Vol. 5727/2009*, Heidelberg: Springer, 2009.

Schomaker, L., Nijtmans, J., Camurri, A., Lavagetto, F., Morasso, P., Benoit, C., Guiard-Marigny, T., Le Goff, B., Robert-Ribes, J., Adjoudani, A., Defee, I., Munch, S., Hartung, K., Blauert, J.: A Taxonomy of Multimodal Interaction in the Human Information Processing System. Multimodal Integration for Advanced Multimedia Interfaces (MIAMI). ESPRIT III, Basic Research Project 8579, 1995.

Sonntag, D., Jacobs, O., Weihrauch, C.: Usability Guidelines for Use Case Applications. Theseus Report CTC WP4, Task 4.1, MS3, April 30, Deutsches Forschungsinstitut für Künstliche Intelligenz, Saarbrücken, 2009.

Sturm, J. A.: On the Usability of Multimodal Interaction for Mobile Access to Information Services. PhD Thesis, Nijmegen: PrintPartners Ipskamp, 2005.

Tractinsky, N.: Aesthetics and Apparent Usability: Empirically Assessing Cultural and Methodological Issues. In: Proc. CHI 1997, pp. 115-122, 1997.

Tractinsky, N., Katz, A. S., Ikar, D.: What is Beautiful is Usable. *Interacting with Computers*, Vol. 13 (2), pp. 127-145, 2000.

Wechsung, I., Naumann, A.B.: Evaluating a Multimodal Remote Control: The Interplay Between User Experience and Usability. In: Proc. of the First IEEE International Workshop on Quality of Multimedia Experience (QoMEX' 2009), pp. 19-22, New York NY: IEEE, 2009.

Wharton, C., Riean, J., Lewis, C., Polson, P.: The Cognitive Walkthrough Method: A Practitioner's Guide. In: Nielsen, J. and Mack, R. L. (eds.), *Usability Inspection Methods*, New York NY: John Wiley & Sons, pp. 105-140, 1994.

Ausgearbeitet von:

I. Wechsung, J. Seebode, S. Schaffer, R. Schleicher, S. Möller, Deutsche Telekom Labs, TU Berlin