

Messung und Bewertung der Usability in Smart Home-Umgebungen

Inhaltsverzeichnis

0	Vorbemerkungen	3
1	Einführung und Definitionen	4
1.1	Gültigkeitsbereich	4
1.2	Interaktion mit Diensten im Smart Home	5
1.3	Qualitätsaspekte von Nutzerschnittstellen	8
1.3.1	Ease of Use	8
1.3.2	Joy of Use	9
1.4	Usability von Nutzerschnittstellen im Smart Home	9
1.5	Usability Engineering Lifecycle	12
2	Auswahl von Evaluationskriterien	14
2.1	Evaluationskriterien bei unterschiedlichen Domänen	14
2.1.1	Kommunikation	14
2.1.2	Sicherheit	14
2.1.3	Energie-Management	15
2.1.4	Gerätesteuerung	15
2.1.5	Entertainment	15
2.1.6	Gesundheit und Fitness	16
2.2	Evaluationskriterien nach Eigenschaften des Dienstes	16
2.2.1	Intelligente Dienste	16
2.2.2	Adaptive Schnittstellen	17
2.2.3	Persuasive Schnittstellen	18
2.2.4	Dienste mit Sensoren	19
2.2.5	Dienste mit Aktuatoren	19
2.2.6	Generische Ein- und Ausgabegeräte	20
2.2.7	Spracheingabe und -ausgabe	20
2.2.8	Dienste für Senioren	21
3	Evaluationsverfahren	23
3.1	Übersicht und Einführung	23
3.2	Tests in Labor und Feld	25
3.3	Versuchsarten	28
3.3.1	Expertenzentrierte analytische Verfahren	28
3.3.2	Nutzerzentrierte empirische Verfahren	30
3.4	Messinstrumente	34
3.4.1	Logfiles	34
3.4.2	Fragebögen	34
3.4.3	Interviews	34
3.4.4	Nutzungstagebücher	34

4	Praktische Hinweise	36
4.1	Methodische Hinweise	36
4.2	Leitfragen zur Planung einer Evaluationsstudie.....	37
4.3	Ermittlung der Testaufgaben mittels Fragenkatalog	37
4.3.1	Ermittlung der Hauptfunktionen	37
4.3.2	Ermittlung der Neben- und Zusatzfunktionen.....	38
4.4	Der Umgang mit Versuchspersonen und ihren Daten.....	38
4.5	Besondere Nutzergruppen	40
4.5.1	Empirische Verfahren mit Senioren.....	40
4.5.2	Empirische Verfahren mit Kindern	41
5	Anhang	43
5.1	Beispiele für eine Versuchsbeschreibung und Einverständniserklärung.....	43
5.1.1	Beispiel: Versuchsbeschreibung.....	43
5.1.2	Beispiel: Einverständniserklärung.....	43
5.2	Beschreibung von Beispielfragebögen	44
5.2.1	AttrakDiff (Hassenzahl, Burmester, Koller 2003)	44
5.2.2	System Usability Measurement Inventory (SUMI, Kirakowski & Corbett 1996).....	44
5.2.3	Subjective Assessment of Speech System Interfaces (SASSI, Hone & Graham 2000).....	44
5.2.4	System Usability Scale (SUS, Brooke 1996)	45
5.3	Beispielitems	46
5.4	Instruktion und Protokoll – Heuristische Evaluation	47
5.4.1	Beispiel: Instruktion	47
5.4.2	Beispielseite Protokoll.....	48
5.5	Fallbeispiel 1: Evaluation der Akzeptanz von technischen Unterstützungssystemen im Smart Home	49
5.6	Fallbeispiel 2: Adaptives Exer-Serious Game.....	51
5.7	Fallbeispiel 3: Evaluation einer Mobiltelefon-basierten Nutzschnittstelle zur Steuerung von Smart Home-Funktionen	52
5.8	Fallbeispiel 4: Evaluation einer multimodalen Fernbedienung.....	54
5.9	Tabelle der Abkürzungen sowie englischen Termini und ihrer Übersetzung	56
6	Literatur	60

Messung und Bewertung der Usability in Smart Home-Umgebungen

0 Vorbemerkungen

Die im Folgenden vorgestellte Richtlinie beschreibt Verfahren zur Messung der Gebrauchstauglichkeit (Usability) von Nutzerschnittstellen im Smart Home. Sie richtet sich insbesondere an Praktiker, also z. B. Ingenieure und Informatiker, die mit der Entwicklung und Evaluierung von Diensten für das Smart Home befasst sind, und soll ihnen eine konkrete Handreichung zur Bewertung verschiedener Aspekte der Usability geben. Entsprechend liegt der Fokus dieser Richtlinie, im Gegensatz zu anderen Richtlinien, Empfehlungen oder Normen (bspw. von der DIN, ISO, ETSI oder ITU-T), nicht auf der Erarbeitung von Definitionen oder der rein funktionalen Prüfung von Hardware- oder Softwarefunktionen, sondern auf der Auswahl passender Messkriterien und Evaluationsverfahren. Da im Smart Home zunehmend Dienste entwickelt werden, die mittels Sensoren Eigenschaften der Umgebung erfassen und darauf basierend sich selbst oder die Umgebung adaptieren, wird den Nutzerschnittstellen solcher „intelligenten“ Dienste besondere Beachtung geschenkt.

Soll eine Schnittstelle nach der vorliegenden Richtlinie evaluiert werden, müssen zunächst Evaluationskriterien festgelegt werden. Hierzu können die auf den Dienst zutreffenden Abschnitte in Kapitel 2 (Auswahl von Evaluationskriterien) konsultiert werden. Anschließend sollte mit Hilfe der Übersicht über Evaluationsmethoden in Kapitel 3 (Evaluationsverfahren) ein geeignetes Messverfahren für die an die Evaluation gestellte Fragestellung identifiziert werden. Vor der ersten Nutzung der Richtlinie empfiehlt sich die Lektüre von Kapitel 1 (Einführung und Definitionen), in dem das der Richtlinie zugrunde liegende Verständnis von Usability und Diensten im Smart Home dargelegt wird. Für weniger erfahrene Praktiker finden sich zudem Erklärungen zu den Messmethoden (Kapitel 3), eine Reihe praktischer Hinweise zur Durchführung von Evaluationsverfahren (Kapitel 4), sowie Beispiele für Evaluations-Studien (Anhang).

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 finden sich im Anhang.

1 Einführung und Definitionen

1.1 Gültigkeitsbereich

Die vorliegende Richtlinie widmet sich der Evaluation der Usability von Nutzerschnittstellen zu Diensten im Smart Home.

In der vorliegenden Richtlinie wird unter Smart Home ein Zuhause, d. h. ein Haus oder eine Wohnung, verstanden, welches mit diversen miteinander kommunizierenden datenverarbeitenden Geräten (Sensoren, Aktuatoren, Ein- und Ausgabegeräte, sowie Haushaltselektrogeräte) ausgestattet ist. Die Geräte sind dazu in ein Netzwerk eingebunden, dass über Protokolle einen Informationsfluss erlaubt. Dies ermöglicht einen Datenaustausch zwischen den Geräten, aber auch die Steuerung eines Gerätes über das Netzwerk.

Ein Dienst ist eine Anwendung, die in dem Netzwerk Informationen oder Funktionen bereitstellt. In dezentralen Netzwerken werden Dienste durch einzelne Geräte bereitgestellt, während in einem zentral organisierten Netzwerk eine Steuerungskomponente zentrale Dienste bereitstellt. Es sind auch Mischformen möglich, entscheidend ist jedoch, dass ein Dienst nicht zwangsläufig nur auf die Informationen und Steuerungsmöglichkeiten eines Gerätes zugreifen kann, sondern auch geräteübergreifend arbeiten kann. Zusätzlich können im Netzwerk auch übergeordnete Dienste, z. B. zur Integration neuer Dienste in das Netzwerk, bereitgestellt werden.

Während Nutzerschnittstellen in der Regel Ein- und Ausgabemöglichkeiten für den Informationsaustausch zwischen Nutzer und Dienst bereitstellen, kommen im Smart Home zunehmend Dienste vor, die aus Sensordaten auf Anforderungen des Nutzers schließen und die Umgebung entsprechend dieser Anforderungen verändern, oder sich selbst an diese Anforderungen adaptieren. Obwohl man in diesen Fällen nicht von einer Nutzerschnittstelle sprechen würde, fällt diese Form der Interaktion in den Gültigkeitsbereich der vorliegenden Richtlinie.

Insbesondere bei den Themenfeldern Gesundheit und Energie-Management werden heute häufig sog. persuasive Nutzerschnittstellen geplant. Bei persuasiven Schnittstellen handelt es sich um Nutzerschnittstellen, die Entscheidungen des Nutzers (z. B. ob er Sport treibt oder Energie spart) positiv beeinflussen sollen. Aufgrund der Bedeutung dieses Konzeptes im Smart Home-Bereich werden persuasive Nutzerschnittstellen hier explizit eingeschlossen.

Ein breites Anwendungsfeld für Technologien im häuslichen Umfeld bilden sog. „Ambient Assisted Living“-Dienste (AAL-Dienste), die die Betreuung pflegebedürftiger Personen erleichtern sollen. Im Bereich AAL gelten eine Reihe besonderer Qualitätskriterien, nicht zuletzt, da hier in der Regel Pflegedienste als zusätzlicher Nutzer mit der Rolle „Betreuer“ eingebunden werden. Aus diesem Grund wird der Bereich AAL in dieser Richtlinie ausgeklammert. Es soll jedoch an dieser Stelle auf die Publikation „Qualitätskriterien im Umfeld von AAL“ (BMBF/VDE Innovationspartnerschaft AAL, 2011) verwiesen werden.

Nicht im Gültigkeitsbereich der Richtlinie liegt die Bewertung technischer Faktoren, die sich nur insofern auf die Gebrauchstauglichkeit eines Dienstes auswirken, als sie grundlegende

Voraussetzung für dessen Nutzung darstellen. Dies sind z. B. die Energieeffizienz eines Eingabegerätes oder die Bandbreite einer Netzwerkverbindung.

1.2 Interaktion mit Diensten im Smart Home

Abbildung 1 zeigt schematisch Akteure und beteiligte Komponenten bei der Interaktion im Smart Home. Zunächst wird unterschieden zwischen dem *Nutzer*, dem *Smart Home*, und einem *Dienst*, der in das Smart Home integriert werden soll. Neben diesem Dienst steht eine Vielzahl weiterer Dienste im Smart Home zur Verfügung, die jedoch aus der gewählten Perspektive als Teil des bereits existierenden Smart Home, d. h., der Umgebung, in die der neue Dienst integriert werden soll, betrachtet werden können. Da alle Dienste beliebig miteinander vernetzt sein können, besteht kein prinzipieller Unterschied zwischen über- und untergeordneten Diensten, so dass übergeordnete Dienste in der Grafik nicht separat aufgeführt werden.

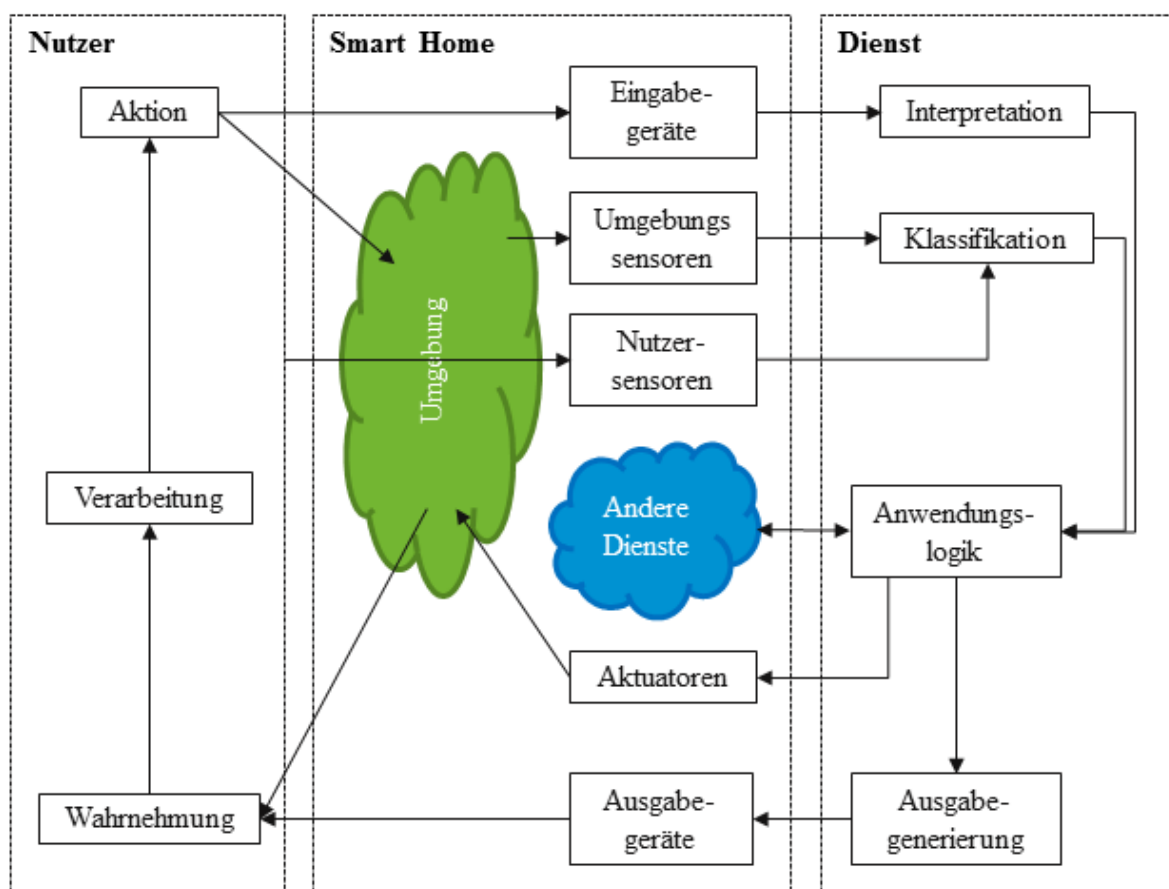


Abbildung 1 Vereinfachte und verallgemeinerte Darstellung der Interaktion von Nutzern mit Diensten im Smart Home

Die Bedienung der verfügbaren Dienste kann über generische *Ein-* und *Ausgabegeräte* erfolgen, die bereits im Smart Home verfügbar sind. Ein generisches Eingabegerät könnte z. B. ein im Haus verbautes Mikrophon sein, während ein Smartphone sowohl als generisches Eingabegerät, als auch als Ausgabegerät verwendet werden könnte. Dienste zur Steuerung von Geräten können neben diesen vom Smart Home bereitgestellten Nutzerschnittstellen häufig auch über im jeweilige Gerät verbaute (und damit spezifische) Nutzerschnittstellen gesteuert werden. Da alle Geräte miteinander vernetzt sind, können Ein- und Ausgaben auch simultan über mehrere generische Geräte und/oder die spezifische Nutzerschnittstelle erfolgen.

Informationen aus der Umgebung, bzw. über die Umgebung (z. B. Helligkeit im Raum) werden, sowohl vom Nutzer als auch vom Dienst, bei der Interaktion genutzt. Systemseitig werden dazu *Umgebungssensoren* eingesetzt. Der Nutzer wiederum nimmt die Umgebung mit seinen Sinnen wahr. Zusätzlich zu Umgebungssensoren kann das System über Sensoren verfügen, die Informationen über den Zustand des Nutzers, wie z. B. Vitaldaten (Puls u. ä.) erfassen (im Bild: *Nutzersensoren*). Andererseits können sowohl Nutzer als auch Dienst direkt auf die Umgebung einwirken (z. B. ein Fenster öffnen). Systemseitig werden dazu *Aktuatoren* verwendet. Der Dienst kann darüber hinaus mit anderen, im Smart Home verfügbaren, Diensten vernetzt sein und über diese weitere Änderungen der physischen oder virtuellen Umgebung erwirken (z. B. könnte ein Dienst einen Wäschetrockner abhängig von der Last im Stromnetz einschalten).

Änderungen an der Umgebung können durch Nutzer und System beobachtet werden, so dass auch über Sensoren und Aktuatoren eine Interaktions-Schleife geschlossen werden kann. Jedoch ist die Interaktion in diesem Fall nicht immer an den Interaktionspartner gerichtet. Für die weiteren Betrachtungen ist es daher sinnvoll, zwischen expliziten und impliziten Ein- und Ausgaben zu entscheiden (Schmidt, 2000).

Bei einer expliziten Eingabe ist die Aktion des Nutzers direkt an das System gerichtet, um eine bestimmte Systemreaktion zu erzielen (z. B. berührt der Nutzer einen Touchscreen, um einen Button zu aktivieren). Eine implizite Eingabe liegt vor, wenn die Nutzeraktion nicht direkt an das System gerichtet ist, aber trotzdem vom System registriert wird (z. B. öffnet der Nutzer die Haustür und aktiviert dabei einen Mikroschalter im Türrahmen). Explizite Eingaben werden über Eingabegeräte gemacht, während implizite Nutzereingaben über Sensoren erfolgen. Kennt der Nutzer die Verarbeitungsmechanismen des Dienstes, kann er jedoch auch über Sensoren explizite Eingaben machen (z. B. einen Lichtsensor abdunkeln, um das Licht einzuschalten). Ebenso kann ein Eingabegerät durch explizite Nutzeraktionen implizierte Informationen (z. B. Nähe des Nutzers zum Eingabegerät) wahrnehmen, somit also implizite Eingaben erlauben. In **Tabelle 1** sind verschiedene mittels Sensoren oder Eingabegeräten erfassbare Daten und ihre typische Zuordnung zur Eingabeart gelistet. Beschreibungen zur technischen Umsetzung der in Tabelle 1 gelisteten Möglichkeiten gibt Wilson (2007).

Analog ist eine explizite Systemausgabe direkt an den Nutzer gerichtet (z. B. eine Nachricht erscheint auf dem Bildschirm). Eine implizite Ausgabe ist eine Systemreaktion die nicht direkt an den Nutzer gerichtet ist, von diesem aber wahrgenommen werden kann (z. B.: das Deckenlicht des Raumes, den der Nutzer betritt, geht an). Ausgaben mittels Ausgabegeräten sind in jedem Fall explizit, während die Verwendung von Aktuatoren explizit (z. B. Lichtsignal bei Anruf) oder implizit (z. B. Anpassung der Helligkeit) sein kann. Nicht alle Systemreaktionen, die durch Aktuatoren erfolgen, werden vom Nutzer unmittelbar wahrgenommen (z. B.: Aktivierung der Warmwasserzirkulation). In diesen Fällen bezeichnen wir die Systemreaktion nicht als Ausgabe.

Tabelle 1 Systemseitig erfassbare Nutzeraktionen oder -eigenschaften und deren Zuordnung zur Art der Eingabe nach Wilson (2007)

	Art der Eingabe		Messgegenstand Nutzer		
	explizit	implizit	Zustand	Aktion	Umwelt
Anwesenheit/Bewegung		✓	✓		
Abstandsmessung		✓	✓		
Position		✓	✓		
Bewegung (und deren Richtung)		✓	✓	✓	
Touch	✓			✓	
Blickerfassung und Eye-Tracking		✓		✓	
Sprache	✓			✓	
Gesten	✓			✓	
Identität (Nutzer und Objekte)		✓	✓		✓
Affekt		✓	✓		
Kontext		✓	✓	✓	✓
Gehirn-Computer-Schnittstelle	✓			✓	

Einige Dienste im Smart Home verarbeiten viele unterschiedliche Informationen von verschiedenen Sensoren und Eingabegeräten, und nutzen mehrere unterschiedliche Ausgabegeräte und Aktuatoren. Um zu einer nützlichen Systemantwort zu kommen, bedarf es in diesen Fällen einer komplexen Anwendungslogik, die häufig nur mit Methoden der künstlichen Intelligenz verarbeitet werden kann (z.B. Machine-Learning oder probabilistische Modellierung).

Informationen aus Eingabegeräten und Sensoren werden typischerweise vorverarbeitet. Je nach Eingabegerät kann eine *Interpretation* der Nutzereingabe erforderlich sein (z.B. bei Sprache). Dies gilt insbesondere, wenn die Informationen aus unterschiedlichen Eingabegeräten zu einer sinnvollen Einheit zusammengeführt werden müssen. Verfahren zur Interpretation von Eingaben können äußerst komplizierte logische Zusammenhänge aufweisen und werden häufig von Experten spezifiziert. Sensordaten können vor der Verarbeitung *vorklassifiziert* werden. Klassifikation meint hier, dass kontinuierliche und möglicherweise mehrdimensionale Eingangssignalen in eine beschränkte Anzahl bekannter Klassen eingeordnet werden, häufig mit maschinellem Lernen. Durch die Vorklassifikation kann die Anwendungslogik vereinfacht werden und Verarbeitungsregeln lassen sich leichter von Hand definieren, da die vorklassifizierten Daten leichter zu interpretieren sind als die Rohdaten.

Die Systemausgabe wird ggf. vom Dienst für ein generisches Ausgabegerät zur Verfügung gestellt (z.B. Text für TTS, HTML für Bildschirmausgabe). Dazu muss die auszugebende Information für die entsprechende Modalität aufbereitet und ggf. in ein standardisiertes Format umgewandelt werden (im Bild: *Ausgabegenerierung*).

Die Darstellung in Abbildung 1 ist stark vereinfacht und deckt nicht alle möglichen Fälle ab. So kann eine Modalität oder ein Sensor mehrere Vorverarbeitungsschritte erfordern, bei denen Klassifikation und Interpretation gemischt sein können. Ein typisches Beispiel für mehrere Vorverarbeitungsschritte ist die Sprachverarbeitung, bei der zunächst das akustische Signal in Text umgesetzt wird (Klassifikation), und anschließend der Text in eine semantische Information, z.B. einen Sprechakt (Interpretation), überführt wird. Auch nutzt nicht jeder Dienst

alle im Smart Home zur Verfügung stehenden Geräte. Die Grafik soll in erster Linie die unterschiedlichen Möglichkeiten der Interaktion verdeutlichen.

1.3 Qualitätsaspekte von Nutzerschnittstellen

Usability (im Sinn von Gebrauchstauglichkeit) beschreibt gemäß der ISO-Norm 9241-11 (ISO, 1998) „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 die wahrscheinlich am weitesten verbreitete Definition von Usability darstellt, sind in der Literatur vielfältige weitere Definitionen zu finden.

Darüber hinaus sei erwähnt, dass in jüngster Zeit der Begriff User Experience (UX), das Nutzungserlebnis, an Popularität gewinnt. In der ISO-Norm 9241-210 (ISO, 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 kann User Experience gemäß Bevan (2009) auf unterschiedliche Art und Weise verstanden werden: 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 umfassendere Darstellung der Zufriedenheitskomponente der Usability. In der vorliegenden Richtlinie wird der letztgenannten 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 durch den Nutzer. Jeder dieser Aspekte lässt sich durch weitere Kriterien beschreiben, die im Weiteren dargestellt werden.

Folglich wird unsere Definition von Usability nicht ausschließlich auf Performanzparameter bezogen und entspricht damit trotz Überschneidungen nicht dem Konzept der „Fitness for Use“ nach DIN 55350-11 (DIN 55350-11:2008-05).

1.3.1 Ease of Use

Die (wahrgenommene) Ease of Use beschreibt inwiefern Nutzer annehmen, dass das Benutzen eines Systems frei von Anstrengung sein wird (Davis, 1989). Ease of Use umfasst u. a. die Aspekte Effektivität, Effizienz, Intuitivität und Erlernbarkeit.

Effektivität bezieht sich auf die „Genauigkeit und Vollständigkeit, mit der Benutzer ein bestimmtes Ziel erreichen“ (nach ISO 9241-11 1998). 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).

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.

1.3.2 Joy of Use

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

Weiterhin von Bedeutung ist die Discoverability des Systems. Die Discoverability beschreibt nach Wechsung (2014), inwiefern das System zur persönlichen Entwicklung des Nutzers, d. h. zur Vermehrung seines Wissens und der Entwicklung seiner Fertigkeiten beiträgt, indem es ihm Neues zu entdecken gibt. Dieses Konzept steht also in direktem Verhältnis zu Hassenzahls Konzept der Stimulation (Hassenzahl, 2003).

Schlussendlich kann auch die Persönlichkeit des Systems bewertet werden. Dieses Konzept bezieht sich auf die Faktoren, die das Verhalten des Systems als Interaktionspartner betreffen (Möller et al., 2006). Die Eigenschaft „Persönlichkeit“ ist vor allem bei Systemen wichtig, die menschliches Verhalten imitieren, bspw. natürlich-sprachliche Systeme oder solche, die einen Avatar als Systemausgabe verwenden.

1.4 Usability von Nutzerschnittstellen im Smart Home

Die Gestaltung von Nutzerschnittstellen für ein Smart Home weist einige Besonderheiten auf, aus denen sich auf eine besondere Rolle einiger Bewertungskriterien im Smart-Home schließen lässt:

- Potenziell werden mehrere Dienste gleichzeitig oder in kurzer Abfolge genutzt; wobei möglicherweise dieselben Ein- und Ausgabegeräte verwendet werden.
- Mehrere Nutzer (Bewohner, aber auch Gäste) teilen sich die zur Verfügung stehenden Ein- und Ausgabegeräte und Dienste.
- Schnittstellen im Smart Home werden von den Nutzern möglicherweise häufig und über lange Zeiträume hinweg genutzt.
- Dienste können von verschiedenen Nutzern gleichzeitig genutzt werden (Multi-User-Szenarien).
- Die Bandbreite unterschiedlicher Nutzer reicht von Kleinkindern bis hin zu Senioren, die sehr unterschiedliche körperliche und kognitive Fähigkeiten haben können.

- Der Dienst kann die Umgebung verändern und Veränderungen der Umgebung wahrnehmen. Durch diese Eigenschaft kann es zu den oben beschriebenen Formen der impliziten Interaktion kommen.
- Bei der Vorverarbeitung von Eingangssignalen (Interpretation von Nutzereingaben, Klassifikation von Sensordaten, Daten-Fusion) können je nach Modalität oder Sensor unterschiedliche Arten von Fehlern auftreten. Insbesondere beim gleichzeitigen Auftreten mehrerer Fehler ist häufig schwer vorhersagbar, welche Kombinationen von Fehlern auftreten können, und wie dies den Interaktionsverlauf beeinflusst.
- Die Verarbeitung unterschiedlicher Eingabesignale erfordert häufig ein komplexeres Interaktionsmanagement. Auch dabei können leicht unvorhergesehene Fehler auftreten.
- Im Netzwerk stehen möglicherweise weitere, übergeordnete Dienste, wie z.B. die regelbasierte Programmierung von Diensten zur Verfügung.
- Die Dienste werden i.d.R. in ein bestehendes Netzwerk integriert, das die Kommunikation mit anderen Diensten regelt.

Aus diesen Besonderheiten lässt sich ableiten, dass folgende Bewertungskriterien eine besondere Rolle bei der Evaluierung der Usability von Diensten im Smart Home spielen:

- Konsistenz: Die Steuerung unterschiedlicher Dienste über unterschiedliche Eingabegeräte soll konsistent sein, so dass Wechsel zwischen Aufgaben oder Eingabegeräten ohne großen Lernaufwand stattfinden kann.
- Kontrolle: Der Nutzer soll in letzter Instanz entscheiden können oder zumindest informiert werden, welche Aktionen in der Umgebung oder auf anderen Diensten ausgeführt werden. Dazu sollten alle Funktionen auch manuell steuerbar sein.
- Transparenz: Das Verhalten des Systems sollte für den Nutzer so transparent sein, dass er Fehler erkennen und korrigieren kann. Dies gilt insbesondere für adaptive Dienste, die ihre Nutzerschnittstelle dynamisch ändern können.
- Aufdringlichkeit: Wenn der Dienst über Aktuatoren Änderungen an der Umgebung des Nutzers vornimmt, sollen diese den Nutzer nicht bei anderen Tätigkeiten stören oder davon ablenken, sofern dies nicht für die Aufgabe des Dienstes erforderlich ist. Dies gilt auch für die Ausgabe von Informationen über Ausgabegeräte.
- Intelligenz: Dienste, die ein komplexes Interaktionsmanagement haben, können hinsichtlich ihrer Intelligenz bewertet werden.
- Personalisierbarkeit: Der Dienst und die Schnittstelle sollen sich an unterschiedliche Personen (deren Vorlieben und Gewohnheiten) anpassen können oder durch die Nutzer selbst angepasst werden können.
- Barrierefreiheit: Es muss berücksichtigt werden, dass besondere Nutzergruppen Zugang zum System haben können, wie z. B. Kleinkinder, Senioren, Gäste oder Service-dienstleister.
- Tauglichkeit für Nutzung durch mehrere Nutzer (Multi-User-Szenarien): Bei einigen Schnittstellen ist eine gleichzeitige Bedienung durch mehrere Nutzer wünschenswert.
- Vorhandensein von Expertenmodus: Wegen der langen Lernzeiten soll die Schnittstelle soll auch für geübte Nutzer effizient zu bedienen sein. Geübte Nutzer kennen z. B. häufig die Aufgabenabläufe und brauchen daher weniger ausführliche Erklärungen.
- Vertrauen/Sicherheit: Der Nutzer soll *wissen*, dass er die Kontrolle hat, um Gefahren einschätzen zu können.

- Performanz und Robustheit von Sensordatenverarbeitung, Eingabeverarbeitung und Anwendungslogik: Diese kann sich massiv auf die o.g. Kriterien auswirken und sollte deshalb während der Entwicklung der Schnittstelle mit erfasst werden.
- Integration mit anderen Diensten und dem Heimnetzwerk: Die Usability beginnt in der Regel bei der Integration des Dienstes in das heimische Netzwerk.

Dienste im Smart Home durchdringen den Alltag der Nutzer und erfordern daher eine besondere Beachtung ethischer Implikationen der Technikgestaltung. In Anlehnung an das MEESTAR-Modell (Manzeschke et al., 2013) lassen sich ethische Fragestellungen zu sieben Dimensionen zusammenfassen:

- Fürsorge/Verantwortung gegenüber den Nutzern: Die Dienste sollen tatsächliche Bedürfnisse der Nutzer adressieren und zu deren positiver physischer und psychischer Entwicklung beitragen.
- Selbstbestimmung/Autonomie der Nutzer: Der Nutzer soll durch die Technik in seinen eigenen Zielen unterstützt werden und Entscheidungen in letzter Instanz selbst bestimmen können. Die Technik soll dazu so transparent und verständlich sein, dass dem Nutzer die Verantwortung für Konsequenzen der Techniknutzung übertragen werden kann.
- Sicherheit: Die Vernetzung von Diensten, die teilweise zentral für mehrere Haushalte zur Verfügung gestellt werden, ermöglicht es theoretisch externen Personen, in Abläufe einzugreifen und Störungen zu verursachen oder Daten zu stehlen. Die Sicherheit vor körperlichem Schaden oder Schäden am Haus oder Einrichtungsgegenständen darf weder durch solche Eingriffe, noch durch Fehlfunktionen, gefährdet werden. Neben der objektiven Sicherheit sollte auch dem subjektiven Sicherheitsgefühl der Nutzer Beachtung geschenkt werden.
- Privatheit: „Privatheit soll einen unverletzlichen Raum um Personen schaffen.“ (Manzeschke et al., 2013, S. 16). Zu Privatheit gehört z. B. ein Recht auf Einsamkeit, Distanziertheit, Isolation, sowie Anonymität und Intimität mit Freunden oder der Familie. Damit ist das Zuhause für die meisten Menschen der Ort, der Privatheit am ehesten ermöglicht. Bei der Technikgestaltung sollte dies berücksichtigt werden, indem Privatheit nicht gestört wird. Ein wichtiger Aspekt ist zudem, dass der Nutzer die Kontrolle darüber haben muss, wer Daten aus der Privatsphäre einsehen kann. Aus den im Smart Home anfallenden Daten (auch, aber nicht nur von Sensoren) können in der Regel Rückschlüsse auf die Privatsphäre der Nutzer gezogen werden. Daher sollten nur Daten gespeichert werden, die für die korrekte Funktion des Dienstes erforderlich sind. Die Art der gesammelten Daten, deren Verbleib und Weiterverarbeitung müssen für den Nutzer transparent sein.
- Gerechtigkeit: Die Technik sollte im Idealfall allen potenziellen Nutzern zur Verfügung stehen, also z. B. frei von körperlichen, intellektuellen, sozialen oder finanziellen Barrieren sein, so dass Ungleichheiten in der Gesellschaft nicht verstärkt werden. Zur Gerechtigkeit gegenüber späteren Generationen gehört, dass Produkte nachhaltig sind oder nachhaltiges Verhalten fördern.

- Teilhabe: Technik im Smart Home kann die soziale und kulturelle Mitbestimmung sowie die Teilnahme am öffentlichen Leben beeinflussen, z. B. indem Zugang zu Informationen oder sozialen Medien bereitgestellt werden. Technik sollte sich nicht hemmend auf die Teilhabe der Nutzer an der Gesellschaft auswirken.
- Selbstverständnis: „Der Begriff des Selbstverständnisses beschreibt die Bewertung und Wahrnehmung eines Subjektes gegenüber sich selbst.“ (Manzeschke, 2013, S. 19). Dieses sollte durch Technik nicht gestört werden. Daher ist zu fragen, inwieweit eine Technologie das Selbstbild der Nutzer ändern kann (z. B. kann eine Optimierung von Tagesabläufen dazu führen, dass Perfektionierung der Persönlichkeit ein überhöhtes Gewicht bekommt). Zum anderen kann Technik zum Ausdruck eines Selbstbilds beitragen, was zunächst eine Individualisierung der Technik erfordert.

1.5 Usability Engineering Lifecycle

Um eine optimale Usability zu gewährleisten, müssen dafür bereits vor dem Systementwurf Vorkehrungen getroffen werden, vgl. Nielsen (1993). So müssen die Eigenschaften und Fähigkeiten der Nutzer, Anforderungen an die mit der Schnittstelle zu erledigende Aufgabe und typische Abläufe bei der Durchführung dieser, sowie der Nutzungskontext analysiert werden. Erste Designs für die Schnittstelle sollten die dabei gewonnenen Erkenntnisse berücksichtigen. Zudem werden basierend auf den Ergebnissen dieser Aufgabenanalyse Usability-Metriken und ggf. Zielwerte für diese Metriken festgelegt. Nach erfolgter Implementierung dienen die Ziele der Usability-Metriken einer quantitativen Überprüfung des Erfolgs. Auch der finanzielle Rahmen für den Einsatz von Usability Engineering Praktiken kann abgesteckt werden. Dabei sollte auch berücksichtigt werden, dass Teile der Ressourcen die sonst für Fehlerbehebungen eingeplant werden, bereits an dieser Stelle genutzt werden können, da Usability-Probleme die von Anfang verhindert werden später nicht als Fehler („Bug“) behoben werden müssen. So sind nach Lederer (1992) die 4 wichtigsten Gründe (aus insgesamt 24) für die Unterschätzung des Budgets bei Entwicklungsprojekten: häufige Änderungswünsche durch den Nutzer, übersehene Aufgaben, fehlendes Verständnis der Nutzer für ihre eigenen Anforderungen, sowie unzureichende(s) Kommunikation und Verständnis bei der Nutzeranalyse. Die Anwendung von passenden Methoden des Usability Engineering kann solche Probleme verhindern oder abschwächen (Nielsen, 1993, S. 5). Beispiele dafür, wie Einsparungen durch Usability-Aktivitäten geschätzt werden können, finden sich bei Nielsen (1993, Kapitel 4.3 Goal Setting).

Nach dem Abschluss der Aufgabenanalyse folgt das Systemdesign. Das Systemdesign ist von zentraler Bedeutung, weil hier die Grundlagen für eine gute Mensch-Maschine-Interaktion gelegt werden. Ein oder mehrere Design-Vorschläge werden 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. Insbesondere vor der Realisierung der Designs als Software sind diese Praktiken sehr preiswert einzusetzen, da Änderungen, mit denen gefundene Usability-Probleme behoben werden sollen, keine Änderungen an einer Implementierung zur Folge haben.

Die weitere Entwicklung der Prototypen erfolgt meist in mehreren Zyklen von Re-Design und Evaluierung (siehe **Abbildung 2**), da durch Änderungen neue Probleme aufgedeckt oder sogar geschaffen werden können. Sobald die Zielsetzungen für die Usability-Metriken erreicht sind, kann die neue Nutzerschnittstelle im Feld (d.h. in der Praxis) getestet bzw. benutzt werden. Rückmeldungen aus dem Feld gestatten eine weitere iterative Optimierung und liefern Ideen für zukünftige Generationen von diesem und anderen Diensten.

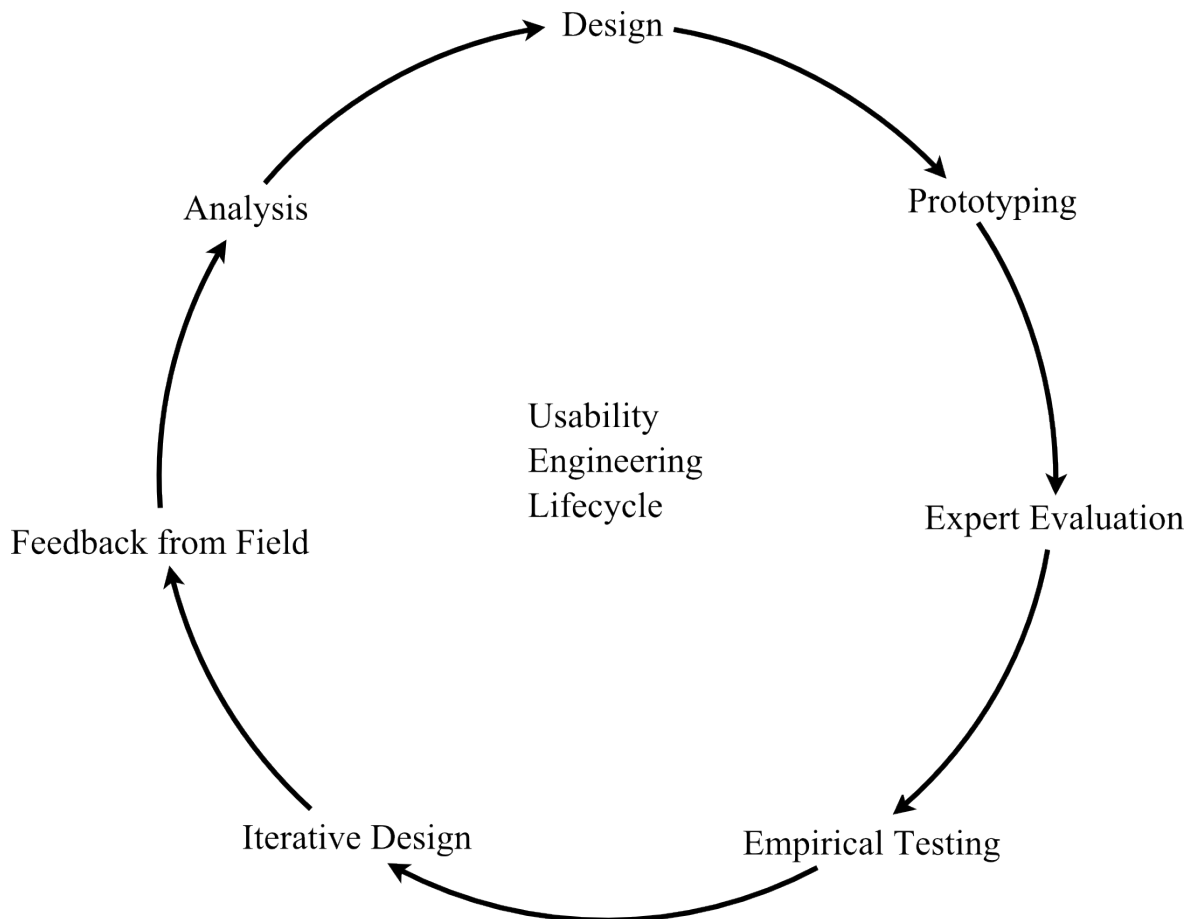


Abbildung 2 Schema des Usability Engineering Lifecycle nach Nielsen, 1993

2 Auswahl von Evaluationskriterien

Im Smart Home findet sich eine große Bandbreite an Diensten aus unterschiedlichen Domänen wie z. B. Entertainment, Kommunikation, Sicherheit, Energie-Management oder Gesundheit. In jeder Domäne ist eine Vielzahl unterschiedlicher Dienste denkbar, die unterschiedliche Ziele oder Bedürfnisse der Nutzer innerhalb dieser Domäne adressieren. Für die Evaluation sind diese Domänen insofern wichtig, als dass sie die Eigenschaften der Nutzer und des Nutzungskontextes eines Dienstes mitbestimmen, und sich somit auf die Relevanz einzelner Qualitätsaspekte, und in einigen Fällen auf die Eignung der Messverfahren auswirken. An dieser Stelle werden zunächst für die einzelnen Domänen jeweils Hinweise gegeben, welche Qualitätsaspekte besonders zu beachten sind. Anschließend werden für einige spezielle Eigenschaften von Schnittstellen die bei der Evaluierung zu beachtende Besonderheiten erläutert.

Soll die Nutzerschnittstelle zu einem Dienst evaluiert werden, sollte zunächst die Domäne identifiziert und die Anmerkungen zu den Domänen in dem jeweiligen Abschnitt konsultiert werden. Anschließend sollten die Abschnitte zu den Eigenschaften der Dienste, sofern sie auf das zu testende System zutreffen, gelesen werden. Die Domänen sind nicht streng komplementär. So trägt z. B. eine Gerätesteuerung auch zum Komfort des Nutzers bei, und möglicherweise auch zur Energie-Verwaltung. Fällt ein Dienst in mehrere Kategorien, sollten die Anmerkungen aller betreffenden Kategorien berücksichtigt werden.

2.1 Evaluationskriterien bei unterschiedlichen Domänen

2.1.1 Kommunikation

Zu dieser Domäne zählen Dienste, die die Kommunikation von Menschen untereinander vermittelnd ermöglichen (z. B. Telefon, E-Mail). Kommunikation ist ein Grundbedürfnis, und vielfach werden Kommunikationsgeräte und Dienste von mehreren Personen im Haushalt geteilt. Besondere Beachtung muss daher der Barrierefreiheit der am häufigsten genutzten Funktionen, sowie der möglichen Nutzung durch mehrere Nutzer geschenkt werden. Komplexere Schnittstellen sollten z. B. personalisierbar sein. Durch Multi-User-Funktionen kann die Nutzung erleichtert werden. Die Kommunikation muss sicher vor Datenverlust sein, und die Privatsphäre der Nutzer sollte gewahrt bleiben. Für genauere Informationen zur Bewertung von Kommunikationsendeinrichtungen kann die dazu erhältliche VDE-ITG-Richtlinie ITG 2.1-01 (ITG, 2011) konsultiert werden.

2.1.2 Sicherheit

Zu dieser Kategorie zählen Systeme, die den Haushalt vor Einbruch und Zerstörung sichern (z. B. Einbruchalarm, Feuermelder, Anwesenheitssimulation). Da solche Schadensfälle nur selten eintreten, sollte für den Nutzer aus der Nutzerschnittstelle erkennbar sein, dass die Geräte einwandfrei funktionieren (z. B. zeigt eine grüne Status-LED an, dass alles einwandfrei funktioniert). Auch bei einwandfreier Funktion sollte getestet werden, ob die Zuverlässigkeit des Systems vom Nutzer korrekt eingeschätzt wird. Während administrative Einstellungen in der Regel von Experten (oder einem bestimmten Bewohner) durchgeführt werden, sollten andere Einstellungen auch für „naive“ Bewohner leicht zu bewältigen sein (z. B. Abschalten des Alarms im Falle eines Fehlalarms).

2.1.3 Energie-Management

Zum Energie-Management zählen Dienste zur Steuerung und Datenverwaltung aller Energiebezogenen Einrichtungen (z. B. Strom, Heizung). Da sich in einem Haushalt mehrere Geräte unterschiedlicher Anbieter für die unterschiedlichen Energieformen befinden können, sollte auf die Konsistenz der Benutzerführung mit Geräten anderer Hersteller und auf die einfache Integrierbarkeit neuer Geräte geachtet werden. Sind Dienste nicht konsistent oder schwer zu integrieren kann es leicht zu Nutzungsfehlern kommen. Zudem sollte die Schnittstelle dem Nutzer in geeigneter Weise vermitteln, welchen Einfluss Änderungen der Einstellungen haben, da dies oft erst längerfristig zu erkennen ist. Da die Nutzerschnittstellen möglicherweise selten, aber über sehr lange Zeiträume hinweg genutzt werden, sollten diese intuitiv, aber auch für geübte Nutzer effizient zu bedienen sein. In manchen Fällen kann es sinnvoll sein, Profile für unterschiedliche Nutzer zu ermöglichen. Eine Besonderheit dieser Domäne ist, dass Verbrauchszahlen regelmäßig zur Abrechnung an einen Dienstleister übermittelt werden müssen, was unter Umständen durch die Gestaltung der Schnittstelle vereinfacht werden kann.

2.1.4 Gerätesteuerung

Die Domäne Gerätesteuerung umfasst Dienste zur automatischen Steuerung von Haushaltsgeräten, wie Lampen, Rollläden, weißer und brauner Ware. In dieser Domäne ist es wichtig, dass die Steuerungen für unterschiedliche Geräte möglichst konsistent sind, damit Nutzungsfehler vermieden werden. Durch Integration der Nutzerschnittstellen für verschiedene Dienste kann ein erheblicher Mehrwert erreicht werden (Beispiel: mit einem einzigen Knopfdruck lässt sich der Fernseher einschalten, das Licht dimmen, und der Sessel ideal einstellen). Die Nutzerschnittstelle sollte auf Effizienz und Effektivität hin optimiert werden, da sie sehr häufig genutzt wird. Es sollte berücksichtigt werden, dass alle im Haushalt lebenden Personen (auch Kinder und Senioren) die zu steuernden Geräte nutzen und dabei möglicherweise Interessenkonflikte entstehen. Durch Personalisierung (z. B. nutzerspezifische Makros) wird die Effizienz der Lösungen verbessert werden. Häufig soll das System Aufgaben teilweise autonom durchführen, in diesem Fall sollte der Abschnitt zu Intelligenten Diensten (2.2.1) zu Rate gezogen werden. Wird die Gerätesteuerung teilweise automatisiert, sollte zudem evaluiert werden, wie stark das Verhalten des Dienstes den Nutzer im schlimmsten Fall stören kann. Das Verhalten des Gesamtsystems sollte für den Nutzer verständlich und nachvollziehbar sein, so dass er Einstellungen optimieren kann. Da Dienste aus der Domäne Gerätesteuerung das Erleben des Zuhauses fundamental ändern können, sollten auch Aspekte der Joy-of-Use betrachtet werden.

2.1.5 Entertainment

Zur Domäne Entertainment zählen Dienste zur Unterhaltung der Bewohner (z. B. IP-TV, Mediaserver). Die Nutzerschnittstellen dieser Domäne sollte auf Effizienz und Effektivität hin optimiert werden, da sie sehr häufig genutzt werden. Es sollte berücksichtigt werden, dass viele verschiedene Personen die Geräte aus dieser Domäne bedienen können (Kinder, Senioren, Gäste, usw.). Wie in der Domäne Gerätesteuerung kann auch hier die Effizienz durch Personalisierung (z. B. vereinfachter Zugang zu präferierten Programmen) verbessert werden. Der Zugang von Kindern zu nicht altersgerechten Medien sollte von Eltern kontrolliert wer-

den können. Werden Inhalte dynamisch in die Schnittstelle integriert, sollte die Bedienbarkeit von den Inhalten unabhängig sein (z.B. sollten Sonderzeichen korrekt dargestellt werden; lange Titel sollten immer erkennbar und unterscheidbar sein).

2.1.6 Gesundheit und Fitness

Dienste im Bereich Gesundheit und Fitness dienen der Planung von Trainingseinheiten, sowie zur Kontrolle von Trainingserfolgen und Gesundheitszustand. Die Kontrollfunktionen sollten Informationen in für den Nutzer aufschlussreicher Form bereitstellen. Messfehler dürfen keine fatalen Konsequenzen haben (z.B. sollte eine Untersuchung durch einen Arzt im Zweifel empfohlen werden). Dies gilt auch für persuasive Schnittstellen (z.B. sollte das System den Nutzer nicht verleiten, den Körper überzustrapazieren). Die Richtlinien zur Gestaltung von AAL-Diensten sollten konsultiert werden (z.B. BMBF/VDE Innovationspartnerschaft AAL, 2011).

2.2 Evaluationskriterien nach Eigenschaften des Dienstes

2.2.1 Intelligente Dienste

Wie sich bereits im Begriff „Smart Home“ andeutet, ist die Entwicklung entsprechender Dienste häufig mit dem Ziel verbunden, dass diese sich intelligent verhalten. Mit *intelligent* ist in diesem Zusammenhang gemeint: „Macht den Eindruck, als ob es intelligent ist“. Dabei lässt sich feststellen, dass intelligent erscheinendes Verhalten häufig auf spezifischeren Eigenschaften des Dienstes basiert, die in den folgenden Abschnitten näher erläutert werden:

- Adaptivität
- Wahrnehmung des Kontexts mittels Sensoren
- Autonome Veränderung der Umgebung mittels Aktuatoren
- Fähigkeit, auf Spracheingabe adäquat zu reagieren

Hat der Dienst eine der genannten Eigenschaften sollten die entsprechenden Abschnitte in diesem Kapitel konsultiert werden. Zusätzlich kann ein Dienst „intelligent“ sein, weil er auf Methoden der künstlichen Intelligenz zurückgreift, um eine große Menge möglicherweise recht heterogener Daten zu verarbeiten und darauf basierend eine nützliche Ausgabe zu generieren. Dieser Aspekt betrifft jedoch einen anderen Teil des Dienstes und nicht die Nutzerschnittstelle.

Als intelligent wahrgenommene Dienste oder Nutzerschnittstellen bekommen durch die Nutzer möglicherweise eine Persönlichkeit zugeschrieben, insbesondere wenn die Schnittstelle einen Menschen imitiert (Möller, 2006). Damit einhergehend kann es sinnvoll sein, Aspekte wie Natürlichkeit, Sympathie oder Charaktereigenschaften bei der Evaluierung des Dienstes zu erfassen.

Wenn bei der Interpretation von Eingaben- oder Sensordaten komplexe, intelligente Verfahren zum Einsatz kommen, wie bei Spracheingabe oder auch textbasierten natürlich-sprachlichen Schnittstellen und Gestenerkennung, kann es zweckdienlich sein, die Wizard-of-Oz-Methodik bei der Evaluierung zu verwenden (siehe auch Abschnitt 2.2.7 „Spracheingabe und -ausgabe“).

2.2.2 Adaptive Schnittstellen

Adaptive Schnittstellen sind dadurch gekennzeichnet, dass sie sich selbstständig an den Kontext sowie die individuellen Charakteristiken des Nutzers anpassen können. Während unter Kontext sowohl der Umgebungskontext (Beschaffenheit der Umgebung) als auch der Benutzungskontext (Anforderungen und Ressourcen) zu verstehen ist, zählen zu den individuellen Charakteristiken sowohl persönlichen Vorlieben, Bedürfnisse als auch Fähigkeiten der Nutzer. Nach Specht (1998) besteht der Prozess der Adaption im Wesentlichen aus der Diagnose und der eigentlichen Anpassung (im engeren Sinne der Adaptionprozess, siehe **Abbildung 3**). Ein adaptives System leitet also aufgrund der Diagnose bzw. Beobachtung (Afferenz) und Auswertung bzw. Modellierung (Inferenz) der Daten (Adaptionsinformationen) automatisch eine Adaption (Adaptionsprozess) ein und passt daraufhin Inhalte oder/und die Nutzerschnittstelle (Adaptionsmittel) entsprechend an (Efferenz).

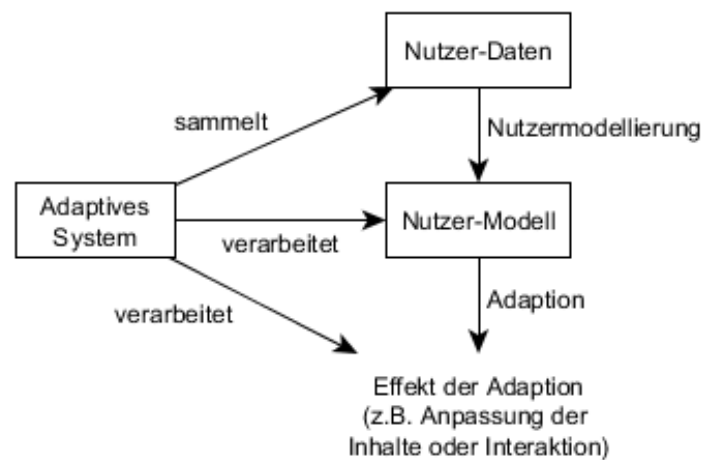


Abbildung 3 Grundlegende Funktionsweise adaptiver Systeme (Brusilovsky, 1996)

Bezogen auf das Smart Home kann die Anpassung (Adaption) sowohl die Nutzerschnittstelle, als auch die Umgebung betreffen. Als Beispiel sei das Smart Home von Bill Gates erwähnt, welches basierend auf einer an der Kleidung befestigten Anstecknadel beim Wechsel in einen anderen Raum des Hauses selbstständig die Zimmertemperatur einstellt, die Lieblingsmusik abspielt und die Beleuchtung entsprechend den individuellen Präferenzen anpasst.

Besondere Herausforderungen beim nutzergerechten Design adaptiver Systeme sind nach Jameson (2008):

- Verminderte Vorhersagbarkeit und Verständlichkeit (bzw. Transparenz)
- Verminderte Steuerbarkeit
- Aufdringlichkeit
- Übertretung der Privatsphäre
- Verminderte Breite des Erlebens (dadurch, dass das System Aufgaben übernimmt oder Inhalte filtert, kommt der Nutzer mit bestimmten Abläufen oder Inhalten nicht mehr in Kontakt und hat dadurch auch weniger Erfahrung damit)

Um diese Herausforderungen zu adressieren existiert eine Reihe präventiver oder kompensierender Maßnahmen. Allerdings wirken sich diese Maßnahmen häufig negativ auf andere Faktoren aus, z.B. kann eine Maßnahme zur Verbesserung der Vorhersagbarkeit dazu führen, dass die Aufdringlichkeit gesteigert wird. Ein Mittelweg muss in der Designphase gefunden werden oder durch den Nutzer konfigurierbar sein (siehe dazu eine ausführlichere Diskussion bei Jameson, 2008).

Bei der Evaluation von adaptiven Systemen oder von Systemen mit adaptiven Schnittstellen ist zu beachten, dass traditionelle Evaluationsansätze aus dem Bereich Human-Computer Interaction (HCI) nicht ohne weiteres angewendet werden können. Der Grundgedanke der HCI-Evaluationsansätze ist es, dass der Zustand und das Verhalten eines Systems gänzlich und allein von direkten und expliziten Eingaben eines Nutzers beeinflusst werden. Adaptive Systeme hingegen weisen ein selbstständiges (nicht notwendigerweise durch den Nutzer kontrolliertes) Verhalten auf, welches traditionelle Evaluationsansätze unzureichend bis gar nicht adressieren. Über die letzten Jahre sind einige vielversprechende Ansätze für die Evaluation von adaptiven Systemen entwickelt worden. Allen gemein ist die Idee, das System nicht als Ganzes zu evaluieren, sondern es in seine Bestandteile zu zerlegen und diese dann einzeln zu bewerten. Diese Methodik wird als „Layered Evaluation“ (Paramythis et al, 2010) bezeichnet. Eine ausführlichere Diskussion geeigneter Evaluationsmethoden für adaptive Systeme findet sich bei Van Velsen et al. (2008).

2.2.3 Persuasive Schnittstellen

Ziel einer persuasiven Schnittstelle ist es, beim Benutzer eine Änderung der Einstellung oder des Verhaltens im Hinblick auf die zu erledigende Aufgabe zu bewirken. Zum Beispiel könnte eine Schnittstelle zum Ablesen des Stromverbrauchs derart gestaltet sein, dass der Verbrauch für den Nutzer klar ersichtlich ist und er motiviert wird Energie zu sparen. Gängige Strategien hierfür sind (a) entscheidungsrelevante Informationen für den Nutzer verständlicher oder emotionaler hinsichtlich ihrer Konsequenzen darzustellen (z.B. Anzeige von CO₂-Ausstoß in Form von abgestorbenen Bäumen), oder (b) durch spielerische Komponenten und Vergleiche mit anderen die intrinsische (d.h., der Aufgabenausführung selbst innewohnende) Motivation für alternatives Verhalten zu stärken.

Beim Testen persuasiver Schnittstellen liegt ein Hauptfokus darauf, zu untersuchen, ob sich die gewünschte Verhaltensänderung einstellt, und ob diese nachhaltig ist. Dies ist eigentlich nur in Langzeit-Feldtests möglich, da der Kontext und die Dauer der Nutzung das Ergebnis entscheidend beeinflussen können. Es ist jedoch möglich Teilaspekte, die kontextunabhängig sind (z.B. die Wahrnehmung oder Bewertung bestimmter Formulierungen), im Labor zu testen.

Da persuasive Schnittstellen eine Verhaltensänderung beim Nutzer bewirken sollen, sollte ausführlich bedacht werden, ob diese Verhaltensänderung auch negative Effekte für den Nutzer oder sein Umfeld haben könnte. Eine spielerische Komponente kann unter Umständen zum Selbstzweck werden, so dass die Hauptaufgabe in den Hintergrund rückt und damit verbundene Ziele manipuliert werden.

2.2.4 Dienste mit Sensoren

Sensoren können zur Erfassung von Kontext-Parametern wie Anwesenheit oder Bewegung von Personen, Temperatur oder Helligkeit sein. Basierend auf dem Kontext wird die Systemreaktion angepasst oder ausgeführt. Reagiert das System auf physikalische Parameter, indem es proaktiv die Umgebung verändert, ist eine Nutzerschnittstelle allenfalls zur Konfiguration oder Kontrolle des Systems notwendig. So bieten viele sensorbasierte Dienste die Möglichkeit, Regeln zu definieren, die die Reaktionen des Dienstes bei bestimmten Sensorwerten festlegen. Solche Schnittstellen sollten mit Nutzern getestet werden, wobei die Leitfrage ist, inwieweit die Nutzer in der Lage sind, mittels der zur Verfügung stehenden Nutzerschnittstellen den Dienst zu kontrollieren und sinnvoll einzustellen.

Werden Parameter erkannt, die mit dem Verhalten der Bewohner zusammenhängen, handelt es sich um eine implizite Interaktion, die jedoch zunächst nicht auf ihre Usability hin untersucht werden muss, da der Nutzer sich nicht auf die Schnittstelle einstellen soll. Dennoch sind Nutzertests erforderlich, um festzustellen, ob das System die erforderliche Leistung erbringt, wenn reale Nutzen in ihrem natürlichen Verhalten erfasst werden.

Da eine implizite Eingabe immer im Kontext einer anderen, natürlichen Handlung erfolgt, muss auch das Szenario das in der Evaluation verwendet wird solch einen Kontext gewährleisten. Da Smart Home-Anwendungen mit nicht immer vorhersehbaren Alltagssituationen konfrontiert werden, und insbesondere auch Ausnahmen von Regelfällen vorkommen, führen nur Nutzertests im Feld zu aussagekräftigen und realistischen Ergebnissen. Außerdem ist wegen der unerwarteten Szenarien und Ausnahmen die Steuerbarkeit, notfalls mit einem zusätzlichen Eingabegerät, ein wichtiger Aspekt bei Design und Evaluation von Sensor-basierten Systemen.

Ein weiterer Punkt der im Rahmen einer Usability-Untersuchung (und auch beim Entwurf eines Smart Home) beachtet werden muss, ist der Eingriff in die Privatsphäre der Nutzer. Ein an sich gebrauchstauglicher und nützlicher Dienst, kann dennoch durch den Nutzer abgelehnt werden, wenn sich der Nutzer in seiner Privatsphäre gestört fühlt. Ein extremes Beispiel ist die Erfassung aller Räume im Haus mit Kameras, um die Nutzer jederzeit identifizieren zu können (neben der Privatsphäre spielt hier auch der Verlust der Kontrolle eine Rolle).

Einen wichtigen Einfluss auf die Usability eines Dienstes, hat die Performanz der eingesetzten Sensoren. Dies meint hier vor allem wie zuverlässig ein Sensor eine Situation erkennt. Die Gesamtperformanz des Dienstes, und damit auch seine Usability, hängen direkt von der Performanz der eingesetzten Sensoren ab.

2.2.5 Dienste mit Aktuatoren

Aktuatoren dienen der Manipulation der Umgebung des Nutzers. Diese Einflussnahme kann entweder unmittelbar wahrgenommen werden (z. B. Jalousieantrieb oder Lichtdimmer), oder nur mittelbar oder gar nicht (z. B. Umwälzpumpe der Warmwasserversorgung). Ebenso kann ein Aktuator dafür sorgen, dass die Umwelt in bestimmten Faktoren die der Nutzer wahrnehmen kann konstant bleibt (z. B. Heizungs-/Klimaanlage, welche für eine konstante Temperatur sorgt).

Bei Diensten mit Aktuatoren, deren Wirkung vom Nutzer unmittelbar wahrgenommen wird, muss auch die mögliche Aufdringlichkeit des Dienstes mit untersucht werden, d. h., ob und in welchem Maß der Nutzer durch die Aktionen des Dienstes in seiner gerade ausgeführten Tätigkeit abgelenkt oder behindert wird.

Bei Diensten mit Aktuatoren deren Wirken sich der Wahrnehmung des Nutzers, sei es durch räumliche Trennung oder wegen einer konstanten Wirkung, entzieht, muss besonderes Augenmerk auf zusätzliche explizite Ausgaben als Feedback und zur Kontrolle für den Nutzer gelegt werden. Die konkreten Kriterien bei Evaluierung der Dienste hängt von Dienstdomäne ab (siehe Abschnitt 2.1 „Evaluationskriterien bei unterschiedlichen Domänen“).

Wichtig ist, dass der Nutzer Aktionen des Dienstes korrigieren kann, so dass bei ungewollten Anpassungen der Umgebung der Fehler behoben werden kann. Weiterhin sollte beachtet werden, ob es durch das eigenständige Agieren eines Dienstes zu Schäden kommen kann.

2.2.6 Generische Ein- und Ausgabegeräte

Auf generischen Ein- und Ausgabegeräten lassen sich Nutzerschnittstellen für unterschiedliche Dienste zur Laufzeit realisieren. So können z. B. Touchscreens (bspw. Smartphones und Tablets), die an das Heimnetzwerk angeschlossen sind, zur Bedienung verschiedener im Netzwerk verfügbarer Dienste genutzt werden. Hierbei kann die konkrete Realisierung der Schnittstelle je nach verfügbarem Gerätetyp unterschiedlich sein. Ebenso können unterschiedliche Modalitäten für Ein- und Ausgabe zur Verfügung stehen (z. B. akustisch über Sprache oder visuell via Touchscreen), welche Nutzungsbarrieren verringern.

Ist die Integration eines neuen Dienstes in eine Umgebung mit generischen Ein- und Ausgabegeräten geplant, sollte die Nutzerschnittstelle auf ihre Gebrauchstauglichkeit überprüft werden. Dabei sollten möglichst viele unterschiedliche Ein- und Ausgabegeräte berücksichtigt werden, insbesondere dann, wenn es sich um unterschiedliche Interaktionsparadigmen bzw. Modalitäten handelt. Werden verschiedene Ein- und Ausgabegeräte verwendet ist zum einen zu prüfen, ob die Systemausgaben auf allen Geräten und für alle Modalitäten geeignet sind. Jedoch sollte auch berücksichtigt werden, dass bei unterschiedlichen Eingabemodalitäten eine unterschiedliche Benutzerführung notwendig sein kann (z. B. ist bei Spracheingabe aufgrund der hohen Fehlerraten i. d. R. ein Bestätigungsdialo g zur Absicherung der Nutzereingabe erforderlich).

Damit generischen Eingabegeräten abwechselnd unterschiedliche Dienste bedient werden, ist zudem auf eine intuitive und mit anderen Diensten konsistente Bedienung zu achten, um Bedienfehler zu vermeiden. Zudem sollte der Dienst auch ohne Zuhilfenahme der generischen Ein- und Ausgabegeräte leicht in das Netzwerk integriert werden können.

2.2.7 Spracheingabe und -ausgabe

Aufgrund der Schwierigkeit Sprache fehlerfrei automatisch zu verstehen, wird die Entwicklung sprachbasierte Nutzerschnittstellen in der Regel stark durch die Möglichkeiten der Spracherkennung geprägt (vgl. Cohen et al., 2004). Fehler können sowohl bei der Spracherkennung (Umsetzung des Audiosignals in einen Text) als auch beim Sprachverstehen (Extraktion von Bedeutung aus dem Text) entstehen (vgl. Möller, 2005). Tests werden daher in der

Praxis meist zur Optimierung der Fehlerraten durchgeführt. Daneben können jedoch auch Usability-Probleme auftreten, die in Tests analysiert werden sollten.

So ist bei Spracheingabe ein wesentliches Ziel von Usability-Tests, die von den Nutzern präferierten Äußerungen und Vokabeln zu ermitteln. Zur Optimierung der Erkennungsleistung kann es sinnvoll sein, feststehende Kommandos zu definieren. In diesem Fall sollte die Intuitivität dieser Kommandos untersucht werden, da häufig nicht alle möglichen Kommandos aus der Systemausgabe ersichtlich sind. Ähnlich sollte aus den Systemausgaben ersichtlich sein, welche Aufgaben durch die Schnittstelle abgedeckt werden („In-Domain-Äußerungen“), da Out-of-Domain-Äußerungen häufig weder vom System noch vom Nutzer eindeutig als solche erkannt werden und daher zu Interaktionsproblemen führen. Da die Leistung der Spracherkennung auch stark von der Art und Positionierung des Mikrofons abhängt, ist zudem eine geeignete Form der Eingabe (z. B. Ansteckmikrofon, ins Handy sprechen o. ä.) für den jeweiligen Kontext zu klären. Ebenso sollte sichergestellt werden, dass die Systemausgaben im Nutzungskontext für den Nutzer verständlich sind.

Ferner ist bei der Gestaltung der Systemausgaben, insofern diese über Sprache erfolgen, zu beachten, dass sie sequentiell erfolgt, so dass der Nutzer zum Ende der Ausgabe nur noch auf jene Informationen aus der Sprachausgabe zurückgreifen kann, an die er sich erinnern kann (Cohen et al., 2004). Im Umkehrschluss kann bei der Evaluation die kognitive Belastung während der Interaktion gemessen werden. Die sequenzielle Darbietung der Ausgabe hat zudem zur Folge, dass das Tempo, in dem die Informationen an den Nutzer weitergegeben werden, durch das System bestimmt wird, und daher optimiert werden sollte. In einigen Fällen kann es sinnvoll sein, in Experimenten zu untersuchen, mit welchen sprachlichen Mitteln sich bestimmte Informationen besonders effizient (im Sinne von Zeit und kognitivem Aufwand) präsentieren lassen.

Handelt es sich um eine Sprachschnittstelle, die einen Dialog mit dem Nutzer führt, wird neben der Gebrauchstauglichkeit häufig die Natürlichkeit des Systems, sowie die Angenehmheit der Stimme und ggf. der Persönlichkeit der Person, die das System darstellt, gemessen (Möller, 2005).

Da die Entwicklung der Spracherkennung und des Sprachverstehens sehr aufwendig ist, werden in der Regel zunächst Wizard-of-Oz-Tests (siehe Abschnitt 3.3.2.3) durchgeführt, um die Ausdrucksweise der Nutzer zu analysieren. Soll die Thinking-Aloud-Methode (siehe Abschnitt 3.3.2.2) angewandt werden, kann man die retrospektive Methode (siehe Abschnitt 4.5.1) anwenden, wobei der Verlauf der Interaktion anhand eines Videos nachbesprochen wird. Ferner soll darauf hingewiesen werden, dass die Leistung sprachbasierter Schnittstellen durch Training auf großen Datenmengen durch manuelles oder automatisches Nachjustieren stark verbessert werden kann, wodurch auch die Gebrauchstauglichkeit erhöht wird.

2.2.8 Dienste für Senioren

Im Zuge des demografischen Wandels werden zunehmend Dienste speziell für Senioren entwickelt. Aber auch andere Dienste sollten für Senioren barrierefrei nutzbar sein, wenn sie von den Funktionen profitieren können. Viele Senioren haben Einschränkungen, die die Wahr-

nehmung, die Kognition, oder die Motorik betreffen. Diese müssen bei der Gestaltung der Nutzerschnittstelle berücksichtigt werden

Altersbedingte Einschränkungen in der Wahrnehmung betreffen alle Sinneswahrnehmungen, allerdings sind die Einschränkungen unterschiedlich ausgeprägt und nicht alle haben einen Einfluss auf das Interaktionsverhalten. Für die Interaktion mit modernen Systemen spielt heutzutage die audio-visuelle Wahrnehmung die entscheidende Rolle. Einschränkungen im Bereich des Sehens können trotz der Verwendung einer Sehhilfe nicht immer vollständig korrigiert werden (GEDA 2012, S.76).

Motorische Einschränkungen, die zu Problemen bei der Bedienung von interaktiven Systemen führen, werden häufig durch altersbedingte Krankheiten verursacht. Arthrose, die weltweit häufigste Gelenkerkrankung, die zu Einschränkungen in der Motorik führt, aber auch Arthritis, Osteoporose oder Übergewicht und Adipositas können zu Einschränkungen in der Motorik führen.

Kognitive Einschränkungen betreffen die Verarbeitungen von Informationen um das System zu bedienen, sowie die Fähigkeit, neue Interaktionsformen zu erlernen. Hinzu kommt, dass ältere Menschen häufig wenig Erfahrung mit aktuellen Technologien haben. Hanson (2010) beschrieb das Problem der fehlenden Erfahrungen wie folgt. Ältere Menschen sind meist eine Technikgeneration hinter der restlichen Bevölkerung weil sie häufig die abgelegte Technik ihrer Kinder nutzen. Ältere Menschen sind also nicht nur durch die längere Lernphase benachteiligt, sondern sie haben auch erst später Zugriff auf neuere Eingabegeräte. Daher sollte sich die Gestaltung an bekannten Interaktionskonzepten orientieren (Frantzidis, 2009; Hanson, 2010), und es sollten wenn möglich bekannte Medien, wie z. B. dem Fernseher, verwendet werden.

3 Evaluationsverfahren

3.1 Übersicht und Einführung

Mit einer Evaluation sind üblicherweise drei Hauptziele verbunden (Dix et al., 2004): Die Bewertung der Funktionalität eines Systems, die Bewertung der Wirkung der Schnittstelle auf den Nutzer und das Identifizieren von spezifischen Problemen bei der Nutzung des Systems. Generell kann man Evaluationsmethoden nach verschiedenen Gesichtspunkten unterscheiden (Preece, et al., 1994).

Je nachdem, in welchem Stadium der Systementwicklung die Evaluation stattfindet, kann zunächst zwischen folgenden allgemeinen Zielen unterschieden werden:

- **Formative Evaluation:** Die formative Evaluation ist eine prozessorientierte Evaluation, die während des Entwicklungsprozesses stattfindet und schon vor dem Vorhandensein eines Prototyps, also schon während der Designphase, erfolgen kann. Ziel ist die Verbesserung des Produktes als Teil des iterativen (d. h. wiederholten) Designprozesses. Dabei spielt das Aufdecken und Beheben von Nutzungsproblemen häufig eine vorrangige Rolle.
- **Summative Evaluation:** Die summative Evaluation ist ergebnisorientiert und somit eine zusammenfassende Beurteilung des fertigen Systems. Diese Form der Evaluation erfolgt, wenn das System schon implementiert ist. Hier sind in der Regel quantitative Aussagen zur Leistung oder Qualität des Systems interessant.

Eine weitere Unterscheidung kann nach der Fragestellung der Evaluation erfolgen:

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

Nach Dix et al. (2004) lässt sich zudem das Niveau der erhobenen Informationen unterscheiden:

- 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 kann auch die Art der Informationen unterschieden werden zwischen *quantitativen*, d. h. numerischen Daten sowie *qualitativen*, d. h. verbalen bzw. nicht-numerischen Daten. Beide Datenarten haben besondere Vor- und Nachteile. Bspw. sind quantitative Daten im Gegensatz zu qualitativen Daten statistisch auswertbar. Allerdings können numerische Messwer-

te unter Umständen weniger detailliert und inhaltsreich sein als qualitative Daten (vgl. Bortz & Döring, 2006).

Zusätzliche Unterscheidungskriterien für Evaluationsverfahren lassen sich nach Dix et al. (2004) aus der Art und Menge der benötigten Ressourcen bilden. Bedacht werden müssen Kriterien wie Zeit, Material und die benötigte Expertise.

Zwei weitere Kriterien zur Abgrenzung von Evaluationsverfahren ergeben sich aus dem Ausmaß der Einbindung der Benutzer. Zunächst kann man unterscheiden 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). Zweitens kann man unterscheiden nach dem *Grad der Nutzerbeteiligung*: Einige Verfahren, insbesondere solche zur formativen Evaluationen im Rahmen iterativer Designprozesse, können auch ohne Beteiligung von Nutzern durchgeführt werden. Daher unterscheidet man:

- *Expertenzentrierte/analytische Verfahren*: Diese werden typischerweise in frühen Entwicklungsstadien eingesetzt. Dabei werden die Elemente der Nutzerschnittstelle und die Auswirkungen der Eigenschaften der Elemente analysiert und modelliert. Größtenteils erfolgt die Evaluierung durch Usability-Experten, d. h. Nutzer und ein fertiges System werden nicht benötigt (Holzinger, 2005).
- *Nutzerzentrierten/empirischen Verfahren*: Bei diesen erfolgt die Datenerhebung über das Beobachten von Nutzern oder über das „Messen“ von Parametern. Die gemessenen Variablen repräsentieren die Performanz des Systems oder die Einstellung der Nutzer zum System. Die Implementation des Systems ist Voraussetzung (Sturm, 2005).

Die *Invasivität* des Verfahrens bezeichnet das Ausmaß, in dem der Nutzer während der Evaluation in seinem natürlichen Verhalten „gestört“ wird. Hier sei bspw. auf Störungen durch den Versuchsleiter oder Beeinträchtigung durch Verkabelung der Versuchspersonen verwiesen.

Nicht alle Verfahren sind für alle Arten von Systemen gleichermaßen geeignet. So muss anhand der Adaptivität eines Systems und den verwendeten Modalitäten die Eignung des Verfahrens geprüft werden.

In Tabelle 2 (S. 27) sind in Anlehnung an Dix et al. (2004) einige Verfahren im Überblick dargestellt, die entweder in der Usability-Praxis etabliert sind oder für die Bewertung von Diensten im Smart Home besonders geeignet sind. In **Tabelle 3** (S. 33) findet sich zusätzlich eine Tabelle, anhand derer Messinstrumente verglichen werden können. Nähere Erläuterungen zu den in den Tabellen aufgelisteten Verfahren finden sich in den folgenden Abschnitten. Praktische Hinweise zur Durchführung der vorgestellten Verfahren finden sich in Kapitel 4. Schließlich gibt **Tabelle 4** (S. 35) eine Zusammenfassung der wesentlichen Vor- und Nachteile aller im Folgenden genannten Verfahren.

In der Regel empfiehlt es sich, Testmethoden so miteinander zu kombinieren, dass die jeweiligen Vor- und Nachteile sich ergänzen. Beispielsweise ist es sinnvoll, vor einer Laborstudie

eine heuristische Evaluation durchzuführen, um leicht zu identifizierende Fehler schon vor der Studie beheben zu können und so die Aufmerksamkeit in der Studie auf die Probleme zu fokussieren, die nur mit Test-Nutzern gefunden werden können (Nielsen, 1993).

3.2 Tests in Labor und Feld

Bei Laborstudien wird der Versuchsaufbau in einem Labor installiert. Die Versuche finden also nicht im Feld, sondern in einer potentiell gut zu kontrollierenden Umgebung statt. Dies hat den Vorteil, dass der Einfluss vieler Störvariablen vermieden oder zumindest erfasst werden kann. Ein weiterer Vorteil ist, dass in einem kurzen Zeitfenster relativ viele Probanden nacheinander betreut werden können, ohne komplizierte Test-Setups immer wieder neu umbauen zu müssen. Auch sind Probandenakquise und Versuchsdurchführung bei Laborstudien normalerweise zeitlich voneinander getrennt. Deshalb können Versuchszeiten effizient geplant werden, und so wertvolle Evaluationszeit gespart werden. Nachteil bei Laborstudien ist, dass man viele Sachverhalte nicht, oder nur schwierig im Labor nachstellen kann und die Ergebnisse deswegen nicht, oder nur teilweise auf reale Situationen übertragbar sind.

Bei einer Feldstudie nutzen Probanden die zu testende Schnittstelle in realen Nutzungssituationen. Feldstudien geben Aufschluss über Dinge, die in Laborstudien und Experimenten nur schlecht oder gar nicht untersucht werden können, wie insbesondere natürliches Nutzungsverhalten, insbesondere bei impliziter Interaktion, längerfristige Lerneffekte des Nutzers und ggf. des Systems, oder realistische Mehrpersonenbedienung. Für Kausalschlüsse sind Feldstudien jedoch nur bedingt geeignet, da die Versuchsbedingungen nur mit erheblichem Aufwand kontrollierbar sind. Außerdem sind Feldstudien häufig sehr aufwendig, da der Versuchsaufbau, also der Prototyp und alle Instrumente, die zur Beobachtung der Testnutzer erforderlich sind, im Feld aufgebaut werden müssen.

Prinzipiell können im Feld die meisten Methoden angewendet werden, die auch im Labor Verwendung finden. Um jedoch möglichst realistische Nutzung zu ermöglichen, werden häufig mehrere Prototypen parallel in verschiedenen Wohnungen über einen längeren Zeitraum genutzt. Dabei muss zunächst durch den Testleiter bzw. die Aufgabenstellung sichergestellt werden, dass die Schnittstelle überhaupt von den Probanden verwendet wird. Da es sich bei den in Feldstudien getesteten Produkten zwar meistens um funktionsfähige Produkte, aber Prototypen handelt, muss außerdem sichergestellt werden, dass die Versuchsteilnehmer auf einfache Art und Weise mit dem Support in Verbindung treten können und dass die Reaktion auf eine Anfrage zeitnah erfolgt, um Systemausfälle oder Fehlfunktionen frühzeitig zu beheben. Es sollte unbedingt vermieden werden, dass einem Versuchsteilnehmer durch das fehlende Wissen über die Bedienung eines Produktes ein Nachteil entsteht, z. B. weil er nicht weiß wie ein automatisch geöffnetes Fenster beim Verlassen des Hauses geschlossen werden kann.

Die Beobachtung der Versuchspersonen im Smart Home kann häufig nur eingeschränkt erfolgen, da dies ein Eindringen in deren Privatsphäre bedeutet. Daher werden die Probanden in der Regel im Anschluss an die Testphase befragt, oder füllen selbständig regelmäßig Fragebögen oder ein Nutzungstagebuch aus. Logfiles können als Quelle für Daten über die Nutzung dienen. Usability-bezogene Maße wie die Ausführungszeit lassen sich allerdings häufig nur bei

Kenntnis des Kontextes sinnvoll interpretieren. Der Kontext wiederum lässt sich in der Regel selten ohne ein Eindringen in die Privatsphäre der Probanden erfassen.

Viele Aspekte der Interaktion mit Diensten im Smart Home lassen sich nur sinnvoll anhand von Feldstudien analysieren, da eine längerfristige Nutzung und natürliches Verhalten der Nutzer Voraussetzung für sinnvolle Rückschlüsse über das Design sind. Aus Kostengründen sollte jedoch versucht werden, Teilaspekte der Nutzung, die sich systematischer analysieren lassen, im Labor zu testen. So können z.B. bestimmte Aspekte der Gebrauchstauglichkeit (z.B. Effizienz eines Dialogs zur Konfiguration) möglicherweise im Labor gemessen werden. In anderen Fällen kann es sinnvoll sein, Hinweise auf Nutzungprobleme aus qualitativen Daten aus Feldstudien im Labor tiefergehend zu untersuchen.

Aus einer von uns durchgeführten Interviewbefragung unter führenden Smart Home Anbietern in Deutschland ging hervor, dass Feldstudien das wichtigste Instrument sind, um neben funktionalen Problemen auch Usability-Probleme zu evaluieren. Laborstudien werden jedoch immer dann eingesetzt werden, wenn neuartige, innovative oder disruptive Nutzerschnittstellen bzw. Bedienkonzepte evaluiert werden. Die Vor- und Nachteile beider Formen sollen hier kurz erläutert werden.

Tabelle 2 Übersicht und Gegenüberstellung der in Abschnitt 3.3 beschriebenen Evaluationsverfahren

Kriterium	Cognitive Walkthrough	Heuristische Evaluation	Modellbasierte Evaluation	Simulation	Card Sorting	Thinking Aloud	Wizard-of-Oz	Experiment	Layered Evaluation	Fokusgruppe	
Entwicklungsstand	Entwurf, PT	Entwurf, PT	Entwurf	PT	Entwurf	jeder	Entwurf, PT	jeder	Layer (implementiert)	Entwurf, PT	
Evaluationsziel	formativ	formativ	formativ	formativ	formativ	formativ / summativ	formativ	formativ / summativ	formativ	formativ	
Fragestellung	Bewertung	Bewertung, Konformität	Bewertung, Vergleich	Bewertung, Vergleich	Bewertung	Bewertung	Bewertung, Vergleich	Bewertung, Vergleich	alle	Vergleich, Feedback	
Level der Information	niedrig	hoch	niedrig	hoch	niedrig	hoch und niedrig	hoch und niedrig	hoch und niedrig	niedrig	hoch	
Art der Daten	qualitativ	qualitativ	quantitativ	quantitativ	qualitativ	qualitativ	quantitativ und qualitativ	quantitativ und qualitativ	quantitativ und qualitativ	qualitativ	
Nutzer-einbindung	Nein	Nein	Nein	Nein	Ja	Ja	Ja	Ja	Ja und Nein	Ja	
Invasiv	Nein	Nein	Nein	Nein	Ja	Ja	Nein	Ja	Ja	Nein	
Eignung	Adaption	niedrig	niedrig	niedrig	hoch	nein	hoch	hoch	niedrig	hoch	mittel
	GUI	hoch	hoch	hoch	mittel	hoch	hoch	niedrig	hoch	hoch	Ja
	Sprache	mittel	mittel	niedrig	hoch	hoch	niedrig	hoch	hoch	hoch	Ja
Geste	niedrig	niedrig	niedrig	niedrig	hoch	hoch	hoch	hoch	hoch	hoch	Ja
Stil	Labor	Labor	Labor	Labor	Labor	Labor/Feld	Labor/Feld	Labor	Labor	Labor	Labor
Benötigte(s)	Zeit	niedrig	niedrig	variiert	hoch	niedrig	mittel	hoch	hoch	pro Layer niedrig	mittel
	Expertise	hoch	mittel	hoch	hoch	niedrig	mittel	mittel	mittel	hoch	mittel
	Material	Beschreibung von Aufgaben, Nutzern, System, Kontext	Heuristik(en) passend zu Aufgabenbereiche, Nutzern, System, Kontext	Beschreibung des Entwurfs	Interaktionslogik des Systems, Modell des Nutzerverhaltens	zu sortierende Begriffe	mindestens (Papier-) Prototyp, Aufgaben	Wizard und Wizard-Interface, ggf. zusätzlich. Kontrollraum	Hypothese, Experimentdesign, entsprechendes Material	Layer des Systems, Ein- und Ausgabeschnittstelle zum Layer, Testfall	niedrig

3.3 Versuchsarten

3.3.1 Expertenorientierte analytische Verfahren

3.3.1.1 *Cognitive Walkthrough*

Der *Cognitive Walkthrough* (CW) ist eine aufgabenorientierte Methode (Holzinger, 2005). Beim CW analysieren Experten – üblicherweise aus den Gebieten Design, Human Factors oder Psychologie – 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 et al., 1994). Grundlagen des CWs sind Theorien des entdeckenden Lernens und Forschungsergebnisse zum Problemlösen (Wharton et al., 1994). Der Ablauf gliedert sich in fünf Schritte (Wharton et al., 1994):

1. 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
2. Einberufen der Experten
3. Analysieren der Handlungssequenzen für jede Aufgabe, indem die Gedanken der potenziellen Nutzer durch die Experten nachvollzogen werden
4. Protokollieren kritischer Informationen
5. Überarbeiten der Schnittstelle

Vorteil des CW ist, dass für diese Methode weder Nutzer noch ein voll funktionstüchtiges System benötigt werden. 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).

3.3.1.2 *Heuristische Evaluation*

Der Begriff „heuristisch“ leitet sich vom Griechischen „heureskein“ ab und bedeutet etwas zu finden bzw. entdecken (Holzinger, 2005). Die heuristische Evaluation (HE) 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 Nutzerschnittstelle. Überprüft wird, ob gängigen *Usability-Prinzipien* (Heuristiken) Rechnung getragen wird (Nielsen, 1994). In (Nielsen, 1994) wird verlangt, dass jeder Experte 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 Experten 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 für eine heuristische Evaluation 3 bis 5 Experten benötigt, die dann laut Nielsen ca. 60 bis 70 % der Probleme aufdecken können. Bei mehr als 10 Experten 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.

3.3.1.3 Modellbasierte Evaluation

Die modellbasierte Evaluation (ME) 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 Berechnung der Ausführungszeiten bei der Bedienung von Computern mit Tastatur und Maus (*Keystroke Level Model*, oder die Methode *Goal Operators Methods Selection Rules*, GOMS). Eine genauere Modellierung der kognitiven Abläufe ermöglichen die *Cognitive Complexity Theory* (CCT) von Kieras und Polson (1985), der *Adaptive Control of Thought – Rational* (ACT-R) Ansatz von Anderson und Lebiere (1998) oder das *State, Operator And Result Modell* (SOAR) (Newell, 1990). Für Evaluierer die nur wenig Erfahrung mit der Modellierung haben empfiehlt sich die Nutzung von CogTool, einer kostenlosen Software zur modellbasierten Evaluation. Bei der Verwendung von CogTool muss zunächst ein Modell des Systems aus Screenshots bzw. Skizzen erzeugt werden. Anschließend wird das so erzeugte Systemmodell zur Demonstration des geplanten Aufgabenpfades genutzt. Der so spezifizierte Pfad wird in ein ACT-R-Modell übersetzt, welches die Ausführungszeiten berechnet. In Studien konnte gezeigt werden, dass auch wenig erfahrene Evaluierer mit Hilfe des CogTool genaue Abschätzungen der Ausführungszeit von Expertennutzern treffen konnten (John et al., 2004).

Die ME kann schon sehr früh im Designprozess angewandt werden und ermöglicht durch die theoretische Fundiertheit ein tiefes Verständnis des Verhaltens 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 bei den meisten Verfahren (Holzinger, 2005). Letzteres trifft insbesondere auf die Nutzung kognitiver Architekturen wie ACT-R und SOAR zu, die daher eher im akademischen als im praktischen Bereich genutzt werden. Weiterhin können komplexere Denk- und Wahrnehmungsprozesse nur mit unverhältnismäßigem Aufwand nachgebildet werden; hierfür ist ME daher nicht zu empfehlen (Kieras, 2007). Eine weitere wichtige Einschränkung dieser Methoden besteht darin, dass die benötigten Parameter bisher nur für klassische Nutzerschnittstellen (Maus + Tastatur, Touchscreen) bestimmt wurden. Dennoch ist sie für *Usability*-Bewertungen in Be-

reichen, in denen die Nutzergruppe sehr speziell oder eine exakte Abschätzung aller möglichen Folgen einer Aktion notwendig ist (Piloten, Kraftwerke), durchaus nützlich.

3.3.1.4 Simulation

Ziel der Usability-Evaluierung mittels Simulation ist es, Daten über das Verhalten des Dienstes in sehr vielen unterschiedlichen Situationen, bzw. bei sehr vielen unterschiedlichen Interaktionsverläufen zu beobachten. Das Verfahren ist deshalb prinzipiell gut für komplexe adaptive Dienste geeignet. Wie bei der modellbasierten Evaluation wird zunächst ein Modell des Nutzers oder der Interaktion gebildet. Dieses Modell kann Ausgaben des Systems verarbeiten und im Idealfall alle Reaktionen, die von einem Nutzer zu erwarten wären, alternativ ausführen. Die Simulation wird ausgeführt, indem das Modell des Nutzers mit dem System selbst, oder einem Modell des Systems interagiert. Solche Simulationen sind bei der Evaluierung von Sprachdialogsystemen bereits sehr verbreitet, allerdings muss in der Regel ein Simulationsframework erst geschaffen werden, da die Reaktionen des Nutzermodells bisher nicht allgemein für beliebige Nutzerschnittstellen formuliert werden konnte.

3.3.2 Nutzerzentrierte empirische Verfahren

3.3.2.1 Card Sorting

Card Sorting ist eine sehr einfache, günstige und hilfreiche Methode, um Menüs logische zu strukturieren (Spencer, 2009). In einem Smart Home existieren viele durch Sensoren erfasste Größen (z. B: Licht Temperatur, Bewegungssensoren) welche genutzt werden können, um Aktionen kontextsensitiv auszulösen. Die Auswahl bzw. Definition von Aktionen kann über grafische Nutzerschnittstellen erfolgen, welche durch Menüs bzw. Navigationsstrukturen aufgerufen werden. Damit das Navigieren in der Nutzerschnittstelle möglichst fehlerfrei ist, ist es wichtig, dass die entsprechenden Menüstrukturen logisch gut gegliedert sind.

Zum Aufbau einer logischen Struktur eines Menüs sollte die Card Sorting Methode eingesetzt werden. Beim Card Sorting werden Probanden gebeten, mit Stichworten beschriebene Karten zu sinnvolle Gruppen zu ordnen. Beim *offenen Card Sorting* werden die Kategorien nicht vom Versuchsleiter vorgegeben sondern von den Probanden benannt. Beim *geschlossenen Card Sorting* werden die Kategorien vorgegeben und der Proband ordnet die Karten mit den Stichworten den entsprechenden Kategorien zu. Die Ergebnisse eines Card Sorting Tests können mit der Menüstruktur eines bestehenden Systems verglichen werden, um dieses zu evaluieren.

3.3.2.2 Protokollanalysen und Lautes Denken (Thinking Aloud)

Bei Protokollanalysen werden Nutzeräußerungen 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 beim Thinking Aloud dazu angehalten, ihr Verhalten und ihre Gedanken zu verbalisieren und laut zu äußern, während sie die Schnittstelle benutzen (Holzinger, 2005). Werden die Handlungen erst nachträglich durch den Probanden beschrieben, bezeichnet man dieses Vorgehen als retrospektives lautes Denken. Um den Probanden in seiner Erinnerung zu unterstützen, werden beim retrospektiven lauten Denken dem Nutzer (Video)-Aufzeichnungen des Tests präsentiert, welche er dann 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 dem Thinking Aloud einer Einzelperson 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. Aus- und Eingabelemente des Systems, die ungünstig sind und vom Nutzer fehl interpretiert werden, fallen schneller auf. Jedoch zeigte sich, dass *Thinking Aloud* häufig als unnatürlich und verwirrend empfunden wird (Lin et al., 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; Aus Sympathie zum Versuchsleiter werden Fehler als weniger gravierend eingestuft oder verschwiegen (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.

3.3.2.3 Wizard-of-Oz Test

Wizard-of-Oz Tests werden verwendet, um Dienste mit einer „intelligenten“ Informationsverarbeitung testen zu können, bevor eine vollständige Implementierung dieser Prozesse erfolgt. Damit lassen sich Design-Ideen, die sehr aufwendig umzusetzen wären, frühzeitig hinsichtlich ihrer Usability evaluieren. Zudem lassen sich Daten dazu sammeln, wie der Dienst in bestimmten Situationen reagieren sollte, oder welche Sprachäußerungen oder Gesten verstanden werden müssen. Diese Daten können anschließend für das weitere Design der Schnittstelle (Adaptionsstrategien, Spracherkennervokabular, Training von Gestenerkennung o.ä.) verwendet werden. Die Methode eignet sich also zum Testen von adaptiven Diensten, Diensten mit Sensoren oder Aktuatoren, Sprachschnittstellen oder Diensten mit Gestensteuerung.

Der Name der Methode kommt von dem Roman „The Wonderful Wizard of Oz“ (Baum, 1900), in dem ein großer und schrecklicher Zauberer sich als eine durch einen normalen Menschen gesteuerte Maschine entpuppt (Möller, 2005). Ähnlich mimit bei einem Wizard-of-Oz Test ein Mensch, typischerweise einer der Entwickler oder eine weitere Testperson, den zu testenden Dienst. Die Probanden interagieren während des Versuchs mit der Schnittstelle, in dem Glauben, einen automatischen Dienst zu testen.

3.3.2.4 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. 2004).

Mit Experimenten ist es möglich, Daten hoher Qualität zu erheben, da Störvariablen im hohen Maße ausgeschaltet oder kontrolliert werden können. Experimente ermöglichen Kausal-

schlüsse 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.

3.3.2.5 Layered Evaluation

Unter Layered Evaluation versteht man ein Evaluationsverfahren, bei dem nicht das System als Ganzes getestet wird (Paramythis, 2010), sondern die einzelnen „Layer“ des Systems unabhängig voneinander evaluiert werden. Aufgrund der Vorgehensweise eignet sich diese Methode besonders gut für adaptive Systeme, die viele verschiedene Systemzustände annehmen können (Paramythis, 2010). Unter einem „Layer“ versteht man einen funktionalen Teil des Systems, den man getrennt von den übrigen Teilen betrachtet. Man stelle sich beispielsweise ein komplexes adaptives System vor, dessen Systemzustand von mehreren Adaptationsprozessen abhängig ist, z. B. könnten Änderungen im Nutzermodell, dem Umgebungsmodell oder der Inferenzmaschine ganz unterschiedliche Systemzustände zur Folge haben. Will man alle möglichen Kombinationen an Zuständen dieser drei „Layer“ systematisch testen, erreicht man sehr schnell sehr viele Bedingungen.

Die „Layered Evaluation“ Methode bietet hierfür einen Lösungsansatz. Statt das Gesamtsystem in allen möglichen Systemzuständen zu testen, wird das Verhalten im jedem Layer (also z. B. Nutzermodell, Umgebungsmodell oder Inferenzmaschine) separat betrachtet und evaluiert. Aus den Ergebnissen der Einzelevaluationen wird dann eine Schätzung für das Gesamtsystem gebildet.

Ein Vorteil der Methode ist, dass Fehlfunktionen in einzelnen Layern schnell erkannt werden können. Außerdem kann der Evaluationsaufwand erheblich reduziert werden. Nicht zuletzt ist es auch möglich, eine formative Systemevaluation durchzuführen, bevor alle Systemkomponenten funktionsfähig sind.

3.3.2.6 Fokusgruppe

Die Methode der Fokusgruppe kommt eigentlich aus dem Bereich des Marketings, kann aber für Usability-Untersuchungen genutzt werden, um Lösungen für Usability-Probleme zu finden oder Designentscheidungen zu treffen (Caplan, 1990). Eine Fokusgruppe besteht aus 6 bis 9 Nutzern die gemeinsam zu einer vorgegebenen Thematik diskutieren (Nielsen, 1997). Das Thema kann z. B. die Verwendung einer Anwendung, oder die Gestaltung einer graphischen Schnittstelle sein. Der Erfolg der Methode hängt dabei maßgeblich von (1) der Expertise des Testleiters und (2) den Möglichkeiten der Teilnehmer ihre Gedanken zu verbalisieren, ab (Caplan, 1990).

Bei der Auswahl der Teilnehmer sollte nach Caplan (1990) darauf geachtet werden, dass alle Fokusgruppenteilnehmer in etwa das gleiche Vorwissen bzw. die gleichen Vorerfahrungen haben. Caplan begründet dies damit, dass Nutzer mit geringen Vorerfahrungen andere Probleme haben können als Nutzer, die über entsprechendes Vorwissen verfügen, jedoch in der Fokusgruppe alle Teilnehmer über dieselben Probleme diskutieren sollen. Da in einem Smart

Home oftmals verschiedene Nutzergruppen interagieren, ist es sinnvoll die Methode mehrmals mit verschiedenen Nutzergruppen durchzuführen.

Bei der Durchführung der Fokusgruppe ist es wichtig, dass Stühle und Tische so positioniert werden, dass der Augenkontakt zwischen allen Teilnehmer möglich ist (Caplan, 1990). Der Moderator muss in der Lage sein neutral über das diskutierte Produkt oder Problem zu sprechen, um die Teilnehmer nicht zu beeinflussen (Lydecker, 1986). Des Weiteren benötigt er eine hohe Vermittlungskompetenz, um zu vermeiden, dass sehr dominante Personen die Meinung in einer Fokusgruppe dominieren (Fern, 1982). Der Verlauf der Diskussion sollte geplant werden, damit vermieden wird zwischen den Themen zu springen (Caplan, 1990). So können z. B. erst generelle und im späteren Verlauf detaillierte Probleme besprochen werden. Die Dauer sollte zwei Stunden nicht übersteigen (Nielsen, 1997).

Tabelle 3 Übersicht und Gegenüberstellung der in Abschnitt 3.4 beschriebenen Messinstrumente

Kriterium	Logfiles	Fragebögen	Interviews	Nutzungstagebuch
Entwicklungsstand	ab PT	jeder	jeder	PT
Evaluationsziel	formativ / summativ	formativ / summativ	formativ	formativ
Fragestellung	Bewertung, Ver- gleich, Feedback	Bewertung, Vergleich	Bewertung, Feedback	Feedback
Level der Information	niedrig und hoch	hoch	hoch	niedrig und hoch
Art der Daten	quantitativ / qualitativ	quantitativ	qualitativ	qualitativ
Nutzereinbindung	ja, außer bei Simulation	Ja	Ja	Ja
Invasiv	Nein	Nein	Nein	Ja
Eignung für adaptive Systeme	Ja	Ja	Ja	Ja
GUI	Ja	Ja	Ja	Ja
Sprache	Ja	Ja	Ja	Ja
Gesten	Ja	Ja	Ja	Ja
Stil	Labor/Feld	Labor/Feld	Labor/Feld	Feld
Benötigte Ressourcen Zeit	mittel	niedrig	hoch	hoch
Benötigte Ressourcen Expertise	mittel	niedrig	mittel	hoch
Benötigte Ressourcen Material	Unterstützung durch Entwickler	Fragebögen	Interview- leitfaden, Diktiergerät	Tagebuch mit notwendigen Informationen für den Nutzer

3.4 Messinstrumente

3.4.1 Logfiles

Logfiles sind Nutzungsprotokolle, die Systeme automatisch generieren können. Dazu muss zunächst definiert werden, bei welchen Ereignissen das System welche Informationen in das Protokoll aufnehmen soll. Man kann z. B. ein Smartphone so programmieren, dass es bei jedem Anruf automatisch die angerufene Person, die Uhrzeit und die Gesprächsdauer protokolliert. Dieses Protokoll nennt man dann Logfile. In der Regel ist das Erstellen von Logfiles wenig invasiv und lenkt den Probanden deshalb nicht ab. Logfiles lassen sich auch automatisch von entfernten Endgeräten übermitteln, so dass sie leicht im Feld gesammelt werden können. Dadurch kann man mit dieser Methode ohne großen Aufwand große Datenmengen generieren, wenn man beispielsweise Software von vielen Probanden gleichzeitig ausprobieren lässt. Da Logfiles einem vorher definierten Format folgen, sind sie im Nachhinein sehr einfach auszulesen. So kann man einige Performanz-Kriterien, wie z. B. die Ausführungszeit, einfach bestimmen. Häufig lassen sich für einen Dienst auch spezifische automatisch bestimmbare Kriterien bilden, die mit der Usability zusammenhängen (z. B. die Häufigkeit, mit der bestimmte System-Zustände erreicht wurden). Nachteil bei dieser Art von Daten ist jedoch, dass oft wertvolle Kontextinformationen nicht vorhanden sind, was die Interpretation der Daten teilweise erschwert.

3.4.2 Fragebögen

Zur Erfassung subjektiver Daten wie Nutzerzufriedenheit, Überzeugung der Nutzer und Ängste der Nutzer wird auf Fragebögen zurückgegriffen (Holzinger, 2005). Mit Fragebögen können Informationen erfasst werden, die sich anderen Verfahren entziehen. Abgesehen davon kann eine Befragung verhältnismäßig schnell und günstig durchgeführt werden. Da die genaue Formulierung von Fragen und Antwortmöglichkeiten einen starken Einfluss auf das Ergebnis der Befragung haben kann, sollte nach Möglichkeit auf validierte Fragebögen zurückgegriffen werden. Beispiele für validierte Fragebögen finden sich im Anhang.

3.4.3 Interviews

Ähnlich wie Fragebögen eignen sich auch Interviews, um subjektive Daten zu generieren. Ein Interview ist ein persönliches Gespräch zwischen Interviewleiter/in und Proband/in. Das Interview ist entweder unstrukturiert (frei), basiert auf einer vorher definierten Agenda (Leitfadeninterview) oder ist strukturiert mit festgelegten Fragen. Je nachdem, wie frei das Interview durchgeführt wird, ist es möglich Informationen zu gewinnen, die von der Versuchsleitung vorher nicht bedacht worden sind. Im Vergleich zu Fragebögen sind Interviews aufwendiger zu erheben und auszuwerten. Interviews eignen sich besonders, um qualitative Daten zu generieren.

3.4.4 Nutzungstagebücher

In Nutzungstagebüchern beschreiben Nutzer typischerweise ihre Erfahrungen im Umgang mit einem Dienst oder einem Produkt, allerdings können auch Situationen oder Abläufe vom Nutzer beschrieben werden. Da die Testteilnehmer das Tagebuch ohne Beisein des Studienleiters ausfüllen, werden die Probanden nicht durch den Studienleiter beeinflusst.

Die Vorbereitung einer Studie, die Nutzertagebücher umfasst, dauert mehrere Tage bis Wochen. Wie bei jeder anderen Methode muss auch hier im Vorfeld festgelegt werden welches Ziel mit der Studie erreicht und welche Fragestellungen durch die Methode beantwortet werden sollen. Nutzertagebücher können so gestaltet werden, dass der Nutzer völlig frei ist in dem was er einträgt, sie können aber auch restriktiver sein und sogar Antwortmöglichkeiten vorgeben.

Für explorative Studien empfiehlt es sich, ein Format vorzugeben, welches dem Nutzer ein hohes Maß an Freiheit beim Ausfüllen bietet, aber dennoch die definierten Fragestellungen anspricht. Dadurch weiß der Nutzer worauf er achten muss.

Da Nutzertagebücher die alltägliche Nutzung dokumentieren sollen, muss ein Nutzertagebuch über einen längeren Zeitraum geführt werden. Als Minimum gilt hier 1 Woche (besser 2 Wochen), wobei die maximale Dauer nur in begründeten Fällen (Nutzertagebuch als Begleitung in der Schwangerschaft) über mehrere Monate gehen sollte.

Tabelle 4 Zusammenfassung von Vor- und Nachteilen der in den Abschnitten 3.3 und 3.4 beschriebenen Evaluationsverfahren und Messinstrumente

Methoden	Vorteile	Nachteile
Cognitive Walkthrough	Ressourcensparend	nur angenommene Probleme, hohe Expertise benötigt
Heuristische Evaluation	Ressourcensparend	nur angenommene Probleme, passende Heuristiken nicht immer vorhanden, hohe Expertise benötigt
Modellbasierte Evaluation	Ressourcensparend	hohe Expertise benötigt
Simulation	wenn die Simulation steht, kann man sehr viele Nutzer/Szenarien in kurzer Zeit simulieren	keine Qualitativen Daten, keine Subjektiven Daten
Card Sorting	Ressourcensparend und sehr leicht durchzuführen	beantwortet nur Fragen zur Menüstruktur
Thinking Aloud	Daten werden von „echten“ Nutzern generiert	lautes Denken fällt Nutzern oft schwer, für Sprachdialogsysteme ungeeignet, funktionierender Prototyp notwendig
Wizard-of-Oz	System kann getestet werden bevor komplexe Eingabeschnittstelle entwickelt wird; Datensammlung für datengetriebene Eingabe-Erkennung	Systemverhalten nicht immer realistisch
Experiment	generalisierbare Daten	sehr aufwendig
Layered Evaluation	gut geeignet für hochkomplexe Systeme mit vielen Systemzuständen	erfasst nur Systemkomponenten, nicht aber das gesamte System
Logfiles	strukturierte Daten über Verlauf der Interaktion und den Systemzuständen und -parametern, Automatisiert auswertbar	zu loggende Daten und Format müssen sorgfältig ausgearbeitet werden, zusätzlicher Implementierungsaufwand
Fragebögen	einfach anzuwenden	oft lange Vorbereitungszeit zur Erstellung valider Fragebögen notwendig
Interviews	freie Antworten	zeitaufwendig/schwierig auszuwerten
Nutzungstagebuch	wenig Aufwand und realitätsnah	unkontrolliert und tendenziell lückenhaft
Fokusgruppe	interessanter Einblick in die Wünsche und Gedanken von den Nutzern	nicht immer ist das, was sich Nutzer wünschen, auch das, was sie später brauchen

4 Praktische Hinweise

4.1 Methodische Hinweise

- Nutzertests sind aufwändiger als Experten-Evaluationen. Gleichzeitig ist es immer wieder überraschend, welche nicht vorausgesehenen Fehler durch Nutzer aufgedeckt werden können. Gerade für neue, ungewohnte Produkte sind Nutzertests daher sehr zu empfehlen.
- Die zum Test ausgewählten Nutzer sollten hinsichtlich ihrer Merkmale wie Alter, Geschlecht, Bildungsstand, Technikerfahrung etc. den späteren Nutzergruppen entsprechen.
- Die Qualität von *Usability*-Tests hängt stark von der richtigen Auswahl der gestellten Aufgaben ab. Die Aufgaben sollten klar formuliert werden und repräsentativ für die spätere Nutzung sein. Dabei ist darauf zu achten, dass der Aufgabentext weder unrealistische Szenarien entwirft noch implizit den Lösungsweg 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 den Dienst 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. hier 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 die spätere Veränderungen des Prototypens ausschließen sind unnötig. Stellen Nutzer fest, dass Ihre Verbesserungsvorschläge nicht umgesetzt wurden, führt dies häufig zu einer gesteigerten Ablehnung (Reaktanz).

Eine Vielzahl von *Usability*-Problemen entsteht noch immer durch schwer erkennbare Schriften/Icons oder unklare Benennungen von Funktionen. Eine erste Möglichkeit, Fehler dieser Art aufzudecken, ist, einen Kollegen („Friendly user“) unvoreingenommen den aktuellen Entwurf bewerten zu lassen. Diese „interne Evaluation“ stößt allerdings schnell an ihre Grenzen, da die Kollegen zumeist gut mit der Grundstruktur oder der internen Logik des zu bewertenden Geräts/Programms vertraut sind.

Eine wichtige Voraussetzung für das Aufdecken von Nutzungsproblemen ist, dass dem Nutzer die interne Struktur der zu testenden 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.

4.2 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?

4.3 Ermittlung der Testaufgaben mittels Fragenkatalog

Die meisten der hier vorgestellten Verfahren zur Evaluation sind zumindest zum Teil aufgabenorientiert, d. h. zu ihrer Durchführung bedarf einer vorherigen Definition von Aufgaben, für welche die Schnittstelle getestet werden soll. Der folgende Fragenkatalog dient der Ermittlung von Testaufgaben, die als Grundlage für die Evaluation mit einem der genannten Verfahren herangezogen werden können. Dabei wird unterschieden zwischen den wichtigsten und den am häufigsten verwendeten Funktionen (Hauptfunktionen), sowie den weniger wichtigen Nebenfunktionen.

4.3.1 Ermittlung der Hauptfunktionen

Die Ermittlung von Hauptfunktionen eines Dienstes kann über die Beobachtung und Befragung realer Endnutzer oder durch eigene Definition erfolgen. Zur Ermittlung von Hauptfunktionen hilft die Beantwortung der nachstehenden Fragen.

- Welche Funktionen werden am häufigsten ausgeführt? Hauptfunktionen sind Schlüsselfunktionen des Dienstes. Sie identifizieren die essentiellen Aufgaben, die ausgeführt werden.

- Welche Funktionen sind für den Funktion des Dienstes notwendig? Schritte, die vollzogen werden müssen, bevor essentielle Aufgaben ausgeführt werden 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 bzw. typisch für den Dienst? Manche Funktionen sind charakteristisch für einen Dienst und nicht unbedingt essentiell, aber sinnvoll für die Interaktion. So soll z. B. ein Eintrag im Adressbuch eines Telefons direkt angerufen werden können.

4.3.2 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.

Je nach Szenario kann eine Nebenfunktion mehr oder weniger wichtig sein. Bei der Auswahl der zu testenden Nebenfunktionen sollte besonders auf den spezifischen Benutzungskontext und die Besonderheiten von Nutzergruppen geachtet werden.

4.4 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 und ihren Daten umzugehen ist.

Da *Usability*-Tests in der Regel keine körperlichen oder psychischen Beeinträchtigungen für die Probanden mit sich bringen, sind die Auflagen 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 *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.

Im Rahmen eines *Usability*-Test sollte es selbstverständlich sein, 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 Mindestanforderungen können gelten:

- Die Probanden sollten nur so behandelt werden, wie man auch selbst in dieser Situation behandelt werden möchte.
- Die erhobenen Daten sollten nur so verwendet werden, wie man das mit eigenen Daten zulassen würde.

Konkret bedeutet dies:

- Schon im Schreiben zur Akquise (Aushang, Rundmail o. ä.) sollte sich eine knappe Schilderung des Versuchs befinden, die dann nicht wesentlich vom tatsächlichen Ablauf abweicht.
- Die Probanden müssen sich vor Versuchsbeginn explizit und schriftlich 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 auf Nachfragen kurz und verständlich erklären können, zu welchem Zweck der Versuch durchgeführt wird, wieso die erhobenen Daten erfasst werden und warum eine bestimmte Methode ausgewählt wurde.
- Alle aufgezeichneten Daten dürfen nicht für unberechtigte Dritte zugänglich sein, auch nicht für andere Versuchsteilnehmer. Deshalb: Lassen Sie keine ausgefüllten Fragebögen anderer Teilnehmer im Untersuchungsraum liegen.
- 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 können, sollten getrennt aufbewahrt werden.
- Jeder Proband wird freundlich und mit Respekt behandelt! Auch auf vermeintlich unwichtige Befürchtungen (z. B. „Angst vor Elektrosmog“) wird ernsthaft eingegangen.
- Dem Wunsch von Seiten eines Teilnehmers, den Versuch zu beenden, ist unmittelbar Folge zu leisten.
- Den Teilnehmern wird explizit erklärt, dass der Zweck die Bewertung einer Schnittstelle/eines Systems und nicht die Bewertung seiner Person ist.
- Es werden keine Vergleiche der Angaben oder Leistung eines Probanden mit anderen Teilnehmern kommuniziert.
- Bei der Versuchsdurchführung sind nicht mehr Personen anwesend als nötig.
- Die Untersuchung dauert nicht länger als vorher angegeben. Eine Verlängerung des Versuchs erfolgt nur nach Absprache mit dem Probanden und bedingt eine entsprechende zusätzliche 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. 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 Vorfeld mit Kollegen oder bei Unsicherheit mit einem Datenschutzbeauftragten durchgesprochen werden.

4.5 Besondere Nutzergruppen

Nutzerschnittstellen im Smart Home werden von unterschiedlichen Nutzern bzw. Nutzergruppen bedient. Mit der Vielfalt der Nutzer steigt die Wahrscheinlichkeit, dass einzelne Nutzer des Systems Einschränkungen haben, auf welche das System eingehen muss. Neben der Nutzergruppe der Senioren, mit den vielfältigen altersbedingten Einschränkungen, zählen Kleinkinder und Menschen mit körperlichen oder geistigen Einschränkungen zu potenziellen Nutzern eines Smart Home. Es kommt allerdings auch vor, dass Menschen Unfälle haben oder krank werden, was zu temporären Einschränkungen führen kann.

Statistisch betrachtet erleidet jeder 13. Bundesbürger einen nicht tödlichen Unfall innerhalb von 12 Monaten, welcher ärztlich behandelt werden muss (GEDA 2012, S.83). Viele dieser Unfälle führen dazu, dass die Interaktionsmöglichkeiten der betreffenden Person danach zumindest zeitweise eingeschränkt sind. Aus diesen Gründen ist es wichtig Nutzerschnittstellen im Smart Home auch mit Nutzern zu testen, die nur eingeschränkt interagieren können. An dieser Stelle sollen einige praktische Hinweise zu Nutzertests mit Senioren und Kindern gegeben werden.

4.5.1 Empirische Verfahren mit Senioren

Evaluationsstudien mit Senioren sind meist kostenintensiv, da es schwierig ist, eine repräsentative Stichprobe von Testnutzern zu rekrutieren (Dickinson et al., 2007). Die Schwierigkeit ein repräsentatives Nutzersample zu erhalten resultiert aus dem Problem, dass die Gruppe der Senioren außerordentlich heterogen ist und in der Regel gerade die kognitiv und physisch fitten älteren Menschen an den Studien teilnehmen (Hawthorn, 2000).

Für eine Evaluationsstudie mit Senioren ist es daher wichtig, Kriterien für die Nutzerauswahl zu definieren. Für die Auswahl empfiehlt sich eine Differenzierung zwischen verschiedenen Altersgruppen (60-69, 70-79 und 80-89), zwischen den Geschlechtern, hinsichtlich der kulturellen Hintergründe, den unterschiedlichen Bildungsniveaus und dem aktuellen Gesundheitszustand (Fisk, 2009).

Bei der Auswahl der Methoden gibt es keine Einschränkungen, allerdings muss beachtet werden, dass manche Methoden besondere Anforderungen an die Versuchspersonen stellen. So zum Beispiel bei der Methode „Lautes Denken“. Hier sollte man Senioren nicht durch Fragen in der Ausführung einer Aufgabe unterbrechen, da es für ältere Menschen schwieriger ist nach einer Unterbrechung die Ausführung fortzusetzen. Insbesondere ist es für ältere Menschen mit altersbedingt eingeschränkten kognitiven Ressourcen je nach Grad der Einschränkung schwierig bis unmöglich, während der Aufgabenausführung dieselbe zu verbalisieren (Fisk, 2009). Auch retrospektives lautes Denken ist hier nicht sinnvoll, da bei diesen Probanden verstärkt Erinnerungsprobleme auftreten (Hill, 2011). Nach Fisk (2009) kann man versuchen die Methode des „Lauten Denkens“ abzuwandeln, indem die Versuchspersonen eine Aufgabe zuerst ohne lautes Denken durchführt und direkt im Anschluss die Aufgabe mit lautem Denken wiederholt (Fisk, 2009 S. 37). Nach Friesdorf (2007, S. 194 ff.) eignen sich für die Einbindung von Senioren die Methoden Fokusgruppe und Experiment, wobei Experimente in späteren Phasen eingesetzt werden können um Fehler in der Benutzerschnittstelle zu finden und Fokusgruppen geeignet sind um Probleme in der Handhabung eines Produktes zu diskutieren.

Bei der Auswahl der Aufgaben ist darauf zu achten, dass die Aufgabenstellung auch für ältere Menschen sinnvoll ist. Sinnvoll meint in diesem Zusammenhang, dass die Aufgabenstellung bzw. das verwendete Szenario Funktionen umfasst, die dann später im Feld auch von älteren Menschen genutzt werden. Das reflektierte Nachdenken über die Aufgabenstellung ist wichtig, da die Art der Nutzung eines Produktes zwischen verschiedenen Altersgruppen abweichen kann, weil verschiedene Altersgruppen unterschiedliche Systemfunktionen priorisieren können (Fisk, 2009; S. 36).

Da Evaluationsstudien häufig eine nicht alltägliche Anforderung an ältere Menschen stellen, muss sichergestellt werden, dass die Teilnehmer vor dem Test entsprechend gut über die Ziele der Studie informiert werden. Besonders bei Versuchen mit älteren Teilnehmern muss klargestellt werden, dass die Studie nicht dazu dient die Leistungsfähigkeit des Probanden zu testen und dass die Probanden jederzeit den Test abbrechen können. Allgemein muss darauf geachtet werden, dass genügend Erholungspausen eingeplant und in der Zeitplanung berücksichtigt werden.

4.5.2 Empirische Verfahren mit Kindern

Kinder zählen, neben den Senioren, zu einer besonderen Nutzergruppe welche häufig im Smart Home anzutreffen ist. Wie bei der Gruppe der Senioren kann auch bei Kindern die Altersgruppe als ein Kriterium für die Auswahl von Probanden definiert werden. Bei den Altersgruppen unterscheidet man im Allgemeinen zwischen drei Altersklassen (Hanna, 1997) und zwar dem Vorschulalter (von 2 bis 5 Jahren), dem Grundschulalter (zwischen 6 und 10 Jahren) und der Pubertät (von 11 bis 14 Jahren). Wenn Kinder zu einem Laborversuch eingeladen werden, dann sollte darauf verzichtet werden eigene Kinder oder Kinder von Beschäftigten einzuladen, da diese Kinder durch ihre Eltern einen Bezug zum Produkt haben und eine missliche Lage kommen können, wenn sie das Produkt nicht mögen (Hanna, 1997).

Die Aufgabenstellung muss, wie bei den Senioren, an für Kinder typische Situationen angepasst sein, allerdings sollte speziell in der Altersklasse der Kinder im Vorschulalter die Aufgabenstellung so offen wie möglich sein, da Kinder in diesem Alter lieber „zeigen was sie können“ als fest definierte Aufgabenstellungen zu lösen (Hanna, 1997). Kinder aus den beiden anderen Altersklassen können durchaus definierte Aufgabenstellungen lösen, da sie dieses Vorgehen bereits aus der Schule kennen, allerdings sollten die Aufgabenstellungen nicht zu lang bzw. zu komplex sein. Komplexe Aufgabenstellungen sollten in kleinere Unteraufgaben unterteilt werden.

Bei der Durchführung der Studie bilden Kinder im Vorschulalter die Altersklasse, die die höchsten Anforderungen an das Studiendesign stellen (Hanna, 1997). Da die Aufmerksamkeitsspanne relativ kurz ist und die Motivation an einer Studie teilzunehmen relativ stark schwankt, muss insbesondere bei dieser Gruppe auf die Dauer des Tests geachtet werden; Als Richtlinie gilt: nicht länger als 30 Minuten für Kinder im Vorschulalter (Hanna, 1997). Wenn kleine Kinder anfangen im sitzen zu hüpfen oder zu zappeln ist es definitiv Zeit eine kleine Pause zu machen. Zwischen zwei Tests sollte ausreichend Zeit eingeplant werden, da Tests mit Kindern sehr anstrengend sein können (Hanna, 1997).

Kinder müssen, wie andere Nutzergruppen auch, darüber informiert werden, dass die Ergebnisse vertraulich behandelt werden, allerdings muss die Einverständniserklärung von den Eltern unterschrieben werden (Hanna, 1997). Für Tests mit Kindern bis zum Alter von 7 bis 8 Jahren ist es wichtig, dass der Versuchsleiter während des Versuches mit im Raum ist. Zum einen kann der Versuchsleiter die Kinder während des Tests motivieren, zum anderen kann er evtl. im Raum befindliche Eltern davon abhalten ihr Kind zu beeinflussen.

5 Anhang

5.1 Beispiele für eine Versuchsbeschreibung und Einverständniserklärung

5.1.1 Beispiel: Versuchsbeschreibung

Liebe Teilnehmerin, lieber Teilnehmer!

Vielen Dank, dass Sie sich die Zeit für diesen Versuch nehmen, der in etwa 1 Stunde dauert.

Stellen Sie sich vor, dass Sie in einem innovativen Haushalt leben, in dem Sie verschiedene Hausgeräte durch Sprache steuern können: den Fernseher und den Videorekorder, den Anrufbeantworter, die Lampen, das Rollo und den Ventilator. In diesem Versuch sollen Sie in drei Situationen mit diesen Geräten kommunizieren und bestimmte Aufgaben lösen.

Nach jeder Situation möchten wir von Ihnen erfahren, wie Sie die Qualität des Systems einschätzen. Dazu haben wir jeweils einen Fragebogen vorbereitet, den Sie bitte durch Ankreuzen ausfüllen. Beurteilen Sie selbstbewusst und bedenken Sie während des gesamten Versuchs:

Nicht Sie werden getestet, sondern Sie testen das System und bewerten es!

Uns ist Ihr spontanes Verhalten und Ihre offene Meinung über das System wichtig. Es ist aber nicht wichtig, wenn Sie einzelne Aufgaben trotz mehrerer Versuche nicht erfüllen können.

Die Dialoge die Sie mit dem System führen, werden aufgezeichnet, um das System verbessern zu können. Lassen Sie sich dadurch aber bitte nicht in Ihrem Verhalten oder in Ihrer Ausdrucksweise einschränken. Für die Auswertung des Tests benötigen wir einige Informationen über Sie, die selbstverständlich anonym bleiben. Deshalb beantworten Sie bitte zunächst die Fragen auf der folgenden Seite.

Und nun: Viel Spaß beim Versuch!

5.1.2 Beispiel: Einverständniserklärung

Hiermit gebe ich mein Einverständnis, dass von mir heute Video- und Audioaufnahmen gemacht werden. Mir ist klar, dass sowohl die Video- und Audioaufnahmen wie auch die von mir zur Verfügung gestellten Informationen streng vertraulich behandelt sowie anonym ausgewertet und nur für Forschungszwecke verwendet werden.

Mir ist zudem bekannt, dass alles was ich während der heutigen Sitzung gesehen und gehört habe vertraulich ist und erkläre, dass ich diese Informationen nicht weitergeben werde.

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

5.2 Beschreibung von Beispielfragebögen

5.2.1 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.

5.2.2 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

5.2.3 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.

5.2.4 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 Likert-Skala beantwortet werden. Mittlerweile wird eine zweifaktorielle Struktur des SUS mit den Dimensionen *Ease of Use* und *Learnability* diskutiert (Lewis & Sauro, 2009).

5.3 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>ATT</i>
lahm vs. fesselnd	AttrakDiff	<i>Joy of Use</i>	<i>HQ-S</i>
stilvoll vs. stillos	AttrakDiff	<i>Joy of Use</i>	<i>HQ-I</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>	/

5.4 Instruktion und Protokoll - Heuristische Evaluation

5.4.1 Beispiel: 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!

5.5 Fallbeispiel 1: Evaluation der Akzeptanz von technischen Unterstützungssystemen im Smart Home

Lehner M., Prieler-Woldan M. (2008) *Projekt Smart Home - Eine Akzeptanzanalyse – Report, FH OÖ Fakultät für Gesundheit und Soziales Linz, Österreich*

Funktionsbereich:	Mit dem Internet verbundenes Fernsehgerät für Bildtelefonie, Nachrichten vorlesen, Bestellung von Essen auf Rädern und sonstiger Versorgungsleistungen sowie Sicherheitsfunktionen wie Wasserstopp und Herdalarm.
Nutzungsszenario:	komplett integriertes System in einem Musterhaus
Entwicklungsphase im Usability Engineering Lifecycle:	Experiment
Eingabemodalität:	Fernbedienung und Wandschalter (haptisch)
Ausgabemodalität:	GUI, gesprochene Sprache, Töne
Erfasste Qualitätsaspekte:	Gesamteindruck, Bedienbarkeit, Kosten, Wichtigkeit/Notwendigkeit (Relevanz)
Angewendete Verfahren:	Experiment
Fragestellung	Bewertung und Akzeptanz von AAL Technologien durch die Zielgruppe.

Die Untersuchung wurde über einen Zeitraum von vier Monaten in einem Musterhaus durchgeführt. Im Rahmen der Untersuchung beurteilten 76 Frauen und 24 Männer u. a. den Gesamteindruck die Bedienbarkeit, Relevanz und Kosten verschiedener Smart Home-Anwendungen. Die Teilnehmer waren 54 bis 90 Jahre alt und wurden in für die weitere Auswertung in jeweils eine der folgenden vier Altersklassen eingeteilt (Zahlen in Klammern geben die Größe der Klasse an): unter 60 Jahre (12), 60-69 Jahre (31), 70-89 Jahre (33) und über 80 Jahre (24). Insgesamt wurden 32 einstündige Tests mit jeweils 2 bis 5 teilnehmenden Personen durchgeführt. Am Anfang eines jedem Test fand eine Einführung in die vorhandenen System statt, dann konnten die Teilnehmer frei mit den Anwendungen agieren (Testaufgaben waren scheinbar nicht vorgegeben) und im Anschluss wurde ein Fragebogen ausgefüllt als auch Kommentare der Teilnehmer gesammelt.

Die in dem Smart Home verwendete Technologie bestand aus drei Komponenten:

- *vitale Systeme* (Sicherheit, wie Herdabschaltung und Wasserstopp beim Verlassen der Wohnung)
- über den Fernseher zugängliche *unterstützende Systeme* für Unterhaltung, Information und Komfort
- *Servicefunktionen* wie Termin- und Medikamentenerinnerung oder Bestellung von „Essen auf Rädern“

Letztere sind unter Laborbedingungen nur schwer zu testen, sondern erfordern eine Feldtest (wie auch die Autoren der hier zitierten Studie feststellen). Die Beantwortung des Fragebogens in Bezug auf diese Funktionen, ist deshalb nur als Einschätzung der Teilnehmer anhand ihrer Lebenserfahrung anzusehen.

Neben Alter und Geschlecht wurden bei den Teilnehmern u. a. auch die Vorerfahrung beim Umgang mit bestimmten Technologien (Haushaltsgeräte, Fernseher, Video/DVD-Rekorder, Handy, Digitaler Fotoapparat/Videokamera, Anrufbeantworter sowie Computer/Internet), die Schulbildung und die aktuelle Wohnsituation erfasst. Der Fragebogen zur Bewertung des Gesamtsystems und der einzelnen Komponenten enthielt die Dimensionen: Gesamteindruck, Kosten, Wichtigkeit und Bedienbarkeit (jeweils für Vitalfunktionen, unterstützende Systeme und Servicefunktionen).

Die gesammelten Ergebnisse zeigen, dass die Einschätzung der Bedienbarkeit von den Funktionen innerhalb der Teilnehmergruppe stark von der Schulbildung und technischen Vorerfahrung (insbesondere mit Computer/Internet) abhängt. Die Bedienbarkeit wurde mit höherer/vorhandener Vorerfahrung besser bewertet. Der Rückschluss ist, dass für ältere Menschen mit geringer technischer Vorerfahrung die Nutzerschnittstelle Anpassungen benötigt. Konkrete Probleme die behoben werden müssen können aus den Kurzinterviews abgeleitet werden.

Aufgrund des hohen Anteils an Frauen und Älteren in der Stichprobe, sind die Ergebnisse, für die Gruppe von Menschen im Alter von 60 bis 90 Jahre nicht repräsentativ.

5.6 Fallbeispiel 2: Adaptives Exer-Serious Game

Lucht, M., Breitbarth, K., & Heimbuch, A. (2013). *Using adaptation to foster social cooperation in exercise and learning. In 3rd International Conference on Social Science and Humanity (pp. 210-216).*

Funktionsbereich:	Steigerung der Lern- und Bewegungsmotivation, Vermittlung von Domänenspezifischen Faktenwissen
Nutzungsszenario:	selbstgesteuertes Lernen, intergeneracionales Spielen
Entwicklungsphase im Usability Engineering Lifecycle:	Testing & Development
Eingabemodalität:	Eingabematte (haptisch)
Ausgabemodalität:	Bildschirm (visuell) und Lautsprecher (Signaltöne)
Erfasste Qualitätsaspekte:	Game Experience, Fehlerrate und Erlernbarkeit
Angewendete Verfahren:	Experiment, Laborstudie, Fragebogen, Interviews, basierend auf Game Experience-Questionnaire, Analyse der Verhaltensspuren
Fragestellung:	Die Playability eines existierendes Serious Game sollte durch die Implementierung von nutzerzentrierter Adaptivität gesteigert werden. Dabei sollte die Veränderung der Dimensionen der Game Experience bei unterschiedlichen Zielgruppen betrachtet werden.

Zur Untersuchung der Fragestellung wurden mit 60 Probanden Experimente im Längsschnitt durchgeführt. Die 60 Probanden teilen sich in 30 Junioren (im Alter zwischen 7 bis 11) sowie in 30 Senioren (älter als 60 Jahre) auf. Zur Ermittlung von Unterschieden wurden die Probanden weiterhin jeweils in Experimental- (Adaptive Version des Spiels) und Kontrollgruppe (Standard Version des Spiels) unterteilt. Insgesamt wurden so 4 Gruppen mit jeweils 15 Probanden gebildet.

Die Probanden wurden über drei Wochen hinweg in einem wöchentlichen Intervall gebeten die jeweilige Version des Serious Games 30 Minuten lang zu spielen. Während des Spielens wurde sowohl bei der Experimental- als auch bei der Kontrollgruppe die Spielperformance der Probanden mittels Logfiles abgebildet. Der Vergleich dieser Daten ermöglichte eine objektive Beurteilung inwieweit durch die adaptive Version Fehlerraten reduziert werden konnten, als auch die Entwicklung der Fehlerrate über die drei Spielsitzungen hinweg. Nach der Spielsitzung wurden die Probanden gebeten anhand des Game-Experience-Questionnaires (IJsselsteijn, 2008) ihre Spielerfahrung wiederzugeben. Da die Junioren Schwierigkeiten mit der Wiedergabe von emotionalen Zuständen haben wurde ein vollstandardisiertes Interview auf Basis eines speziell angepassten Kids-Game-Experience-Questionnaire (bisher unveröffentlicht) durchgeführt.

Die Ergebnisse der Fragebögen und standardisierten Interviews wurden quantifiziert und Signifikanztests unterzogen. Auch die kumulierten Fehlerraten wurden für weitere korrelative Analysen mit den Daten der Interviews und Fragebögen über Pseudonyme verknüpft. Dabei wurden hinsichtlich der objektiven Daten die Experimental- und Kontrollgruppen der Junioren sowie Senioren untereinander verglichen. Bei den subjektiven Daten konnte ein gruppenübergreifender Vergleich durch die unterschiedlichen Erhebungsinstrumente nicht umgesetzt werden.

5.7 Fallbeispiel 3: Evaluation einer Mobiltelefon-basierten Nutzschnittstelle zur Steuerung von Smart Home-Funktionen

Koskela, T., Väänänen-Vainio-Mattila, K. und Lethi, L. (2004) *Home Is Where Your Phone Is: Usability Evaluation of Mobile Phone UI for a Smart Home. MobileHCI 2004, LNCS 3160, S. 75-85. Springer Verlag Berlin.*

Funktionsbereich:	Steuerung und Monitoring von Geräten und Sensoren im Smart Home
Nutzungsszenario:	Steuerung des Smart Home mittels Mobiltelefon inner- und außerhalb der Wohnung
Entwicklungsphase im Usability Engineering Lifecycle:	Analysis, Design, Prototyping, Expert Evaluation, Empirical Testing
Eingabemodalität:	Handytastatur (haptisch)
Ausgabemodalität:	GUI (visuell, textbasiert), teilweise unmittelbar (visuell)
erfasste Qualitätsaspekte:	Effizienz, feeling of safety, feeling of belongingness
Angewendete Verfahren:	Fokusgruppen, Interview, heuristische Analyse, Experiment und Feldtest
Fragestellung:	Ermittlung der Usability und Akzeptanz eines Mobiltelefons als Nutzerschnittstelle zum Smart Home

Sich an der ISO EN 13407 *Benutzer-orientierte Gestaltung interaktiver Systeme* (seit 2011 von EN ISO 9241-210 abgelöst) orientierend wurde für eine Mobiltelefon-basierte Nutzschnittstelle zur Steuerung von Smart Home-Funktionen entworfen und implementiert. Aufgrund des Nutzer-zentrierten Entwicklungsansatzes fanden Evaluationen während der Design- und Entwicklungsphase und anschließend auch im Feld statt. In dem Smart Home konnten alle Lampen und andere elektronische Geräte sowie die Lamellenjalousien an den Fenstern gesteuert werden. Ebenso konnte der aktuelle Status all dieser Geräte abgefragt werden. Zusätzlich konnten mittels Sensoren die Temperatur, Beleuchtungsverhältnisse und Substratfeuchte der Pflanzen gemessen werden. Die Funktionen des Smart Home konnte neben dem Mobiltelefon (über Textmenüs) auch mittels PC und Fernseher (in Verbindung mit einer Fernbedienung) genutzt werden. PC und Fernseher wurden in diesem Fallbeispiel nicht tiefergehend betrachtet.

Während er Anforderungsanalyse fanden kontextuelle Interviews (in der Wohnung des jeweils Befragten) und Sitzungen mit Fokusgruppen, mit insgesamt 22 Teilnehmern, statt. Die Teilnehmergruppen setzten sich aus Erwachsene, Familien mit Kindern, Paaren mittleren Alters und älteren Erwachsenen zusammen. Mit Hilfe der kontextuellen Interviews wurden Muster und Bedürfnisse beim alltäglichen Umgang mit Technologien, im eignen Zuhause, ermittelt. Mittels der Fokusgruppen wurden bei den Teilnehmern die *Einstellungen gegenüber* und *Erwartungen an* zukünftiges Wohnen erhoben.

Die Designentwürfe wurde zum einem mittels heuristischer Analyse als auch in einem Usability Experiment (teilweise mit Wizard-of-Oz Setting) mit 5 Nutzern evaluiert. Ziel bei dem Einsatz der Evaluationsmethoden war die Ermittlung von Usability-Problemen, die Bewertung der Usability stand nicht im Vordergrund. Da an dem späteren Feldtest zunächst nur junge Erwachsene teilnehmen sollten, wurden auch hier nur Teilnehmer dieser Gruppe eingeladen.

Für den Feldtest wurde eine Wohnung vollständig mit den oben genannten Funktionen ausgerüstet. Dabei wurde darauf geachtet, die notwendige Netzwerkinfrastruktur so zu installieren, dass sie für die Nutzer nicht sichtbar war. Die Wohnung wurde für 3 Monate von einem Paar (26 und 27 Jahre) alt bewohnt, während Beide ihrem gewohnten Alltag – inklusive Arbeiten – nachgingen. Während des Feldtests wurden mit den Teilnehmern kontextuelle Interviews und Participatory Walkthroughs, in denen die mobile Nutzerschnittstelle untersucht wurde, durchgeführt.

5.8 Fallbeispiel 4: Evaluation einer multimodalen Fernbedienung

Wechsung, I. und 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)* (S. 19-22). New York: IEEE.

Funktionsbereich:	Medienaufzeichnungs- und Abspielanwendung
Nutzungsszenario:	Festinstalliertes System im privaten Kontext
Entwicklungsphase im Usability Engineering Lifecycle:	Prototyping
Eingabemodalität:	Tasten (haptisch), Spracheingabe (linguistisch), Bewegungssteuerung (Gesten (mit Telefon ausgeführt))
Ausgabemodalität:	GUI mit Second Screen (Fernseher und Smart Phone Display)
erfasste Qualitätsaspekte:	Ease of Use, Joy of Use
Angewendete Verfahren:	Heuristische Evaluation, Experiment und Feldstudie jeweils mit Fragebögen und Protokollanalysen
Fragestellung:	Für ein 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.

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 (s. Abschnitt 3.3.1.1) mit Elementen der Heuristischen Evaluation (s. Abschnitt 3.3.1.2). 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 sämtlicher 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.

5.9 Tabelle der Abkürzungen sowie englischen Termini und ihrer Übersetzung

AAL	s. Ambient-Assisted-Living
Acceptability	Akzeptanz
ACT-R	s. Adaptive Control of Thought–Rational
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
Ambient-Assisted-Living	technikunterstütztes Leben
Analysis	Analyse
Apperance	äußeres Erscheinungsbild des Systems
ATT	Attraktivität, eine Dimension im Fragebogen AttrakDiff
Avatar	Avatar, eine virtuelle Figur
Button	das Bedienelement Knopf in einer graphischen Nutzeroberfläche
card sorting	„Karten sortieren“, empirisches Verfahren zur Kategorisierung von Begriffen
Cognitive Complexity Theory (CCT)	„Kognitive Komplexitätstheorie“, Methode zur Aufgabenanalyse
Cognitive Demand	kognitive Beanspruchung
Cognitive Walkthrough	kognitiver Durchgang, Durchdenken eines Problems
Cooperativity	Kooperativität
DIN	Deutsches Institut für Normung
Discount Usability Engineering	pragmatische und kostengünstige Methode des Usability Engineering
Discoverability	„Entdeckbarkeit“, beschreibt inwiefern das System zur Vermehrung des Wissens und der Entwicklung der Fertigkeiten, eines Nutzers, beiträgt.
Ease of Use	Einfachheit der Nutzung
Effectiveness	Effektivität
Efficiency	Effizienz
Empirical Testing	empirische Untersuchung
ETSI	European Telecommunications Standards Institute (Europäisches Institut für Telekommunikationsnormen)
Expert Evaluation	Expertenevaluation
Eye-Tracking	Aufzeichnung und Analyse von Augenbewegungen, mit dem Ziel betrachtete Bildbereiche zu ermitteln.
false negative	falsch negativ

false positive	falsch positiv
Feedback	Rückmeldung
feeling of belongingness	das Gefühl des Dazugehörens
feeling of safety	das Gefühl von Sicherheit
Fitt's Law	„Fitts' Gesetz“, Modell zur Bestimmung von Dauer und Genauigkeit von Arm- und Mausbewegungen
Friendly User	„netter Nutzer“, ein Nutzer, der motiviert ist, das zu testende System zu verbessern und daher bereitwillig am Test teilnimmt
Game	Spiel
Goal Operators Methods Selection Rules (GOMS)	Methode, die unterschiedliche Modelle der Mensch-Computer-Interaktion, z. B. zur Bestimmung von Ausführungszeiten, integriert
GOMS	s. Goal Operators Methods Selection Rules
GUI	Graphical User Interface (graphische Benutzerschnittstelle)
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
HCI	Human-Computer Interaction (Mensch-Maschine Interaktion)
HE	Heuristische Evaluation
Helpfulness	„Unterstützung“, beschreibt, wie selbsterklärend die Software ist, und spezifischere Aspekte wie die Angemessenheit der Hilfe und der Dokumentation
Hierarchical Task Analysis	„Hierarchische Aufgabenanalyse“, Methode zur Aufgabenanalyse
HQ-I	Hedonische Qualität – Identität, eine Dimension im Fragebogen AttrakDiff
HQ-S	Hedonische Qualität – Stimulation, eine Dimension im Fragebogen AttrakDiff
HTML	Hypertext Markup Language, „Hypertext-Auszeichnungssprache“
Icon	Symbolbild, Piktogramm
In-Domain-Äußerung	sprachliche Äußerung, deren Informationsinhalt innerhalb („in“) einer fachlichen Domäne liegt
Intuitiveness	Intuitivität
Intuitivity	Intuitivität
IP-TV	Internet Protocol Television, Fernsehen über das Internet
ISO	Internationale Organisation für Normung
Iterative Design	iteratives Design, zyklisches Re-Design
ITU-T	International Telecommunication Union – Telecommunication Standardization Sector
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
Layer	Ebene
Layered Evaluation	Ebenenweise Evaluation
Learnability	Erlernbarkeit
Likeability	„Beliebtheit“, enthält die subjektive Meinung des Nutzers über das System und dessen Empfinden während der Interaktion
Logfile	Datei mit protokollierten Systemparametern und -ereignissen
Machine-Learning	maschinelles Lernen, eine Methode zur Modellbildung
Monitoring	beobachten, überwachen
Multi-User	Mehrbenutzer
Obtrusiveness	Aufdringlichkeit eines Systems
Out-Of-Domain-Äußerung	sprachliche Äußerung, deren Informationsinhalt außerhalb („out“) einer fachlichen Domäne liegt
Participatory	teilnehmend
Playability	Spielbarkeit
PQ	Pragmatische Qualität, eine Dimension im Fragebogen AttrakDiff
Pramatic	pragmatisch
Privacy	Privatsphäre
Prototyping	Entwicklung eines Prototypen
push-to-talk	ein Ansatz bei Sprachdialogsystemen, bei dem der Nutzer dem System mitteilt, dass er jetzt sprechen möchte (z. B. durch das Drücken einer Taste)
SASSI	Subjective Assessment of Speech System Interfaces
Second Screen	zweiter Bildschirm
Security	Sicherheit
SOAR	s. State, Operator And Result Modell
State, Operator and Result Modell (SOAR)	eine Architektur zur Modellierung der menschlichen kognitiven Prozesse
SUMI	System Usability Measurement Inventory
SUS	System Usability Scale
Tablet	Tabletcomputer, leichter Computer mit Touchscreen
Task Factors	aufgabenbezogene Qualitätsfaktoren
Testing	testen
Thinking Aloud	lautes Denken
TTS	Text-to-Speech synthesis, „Sprachsynthese“ (Erzeugung von künstlicher Sprache aus Texten)
Usability	Gebrauchstauglichkeit, manchmal auch „Benutzerfreundlichkeit“

Usability Engineering Lifecycle	Berücksichtigung der Methoden des Usability Engineering in allen Phasen der Systementwicklung zur optimalen Systemgestaltung
Usability Testing	Untersuchung der Gebrauchstauglichkeit
Usability Engineering	Methoden zur Messung und Bewertung der Gebrauchstauglichkeit interaktiver Systeme
Usefulness	Nützlichkeit
User	Nutzer
User Experience	Nutzererlebnis, Nutzungserlebnis
Wizard-of-Oz	„Zauberer von Oz“, eine Methode, um unvollständige Systeme zu testen

6 Literatur

- Anderson & Lebiere 1998 Anderson, J. R., & Lebiere, C.: The atomic components of thought. Mahwah, N. J.: Lawrence Erlbaum Associates, 1998.
- Baum 1900 Baum, L. F.: The wonderful wizard of oz. Kansas Centennial Edition (1999), University Press of Kansas, USA-Lawrence KS, 1900.
- Bevan 2009 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.
- BMBF VDE 2011 BMBF/VDE Innovationspartnerschaft AAL (Hg.): Qualitätskriterien im Umfeld von AAL. Produkte – Dienstleistungen – Systeme, VDE Verlag, 2011.
- Bortz & Döring 2006 Bortz, J., Döring, N.: Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler, 4. Auflage. Berlin: Springer, 2006.
- Brooke 1996 Brooke, J.: SUS: A “Quick and Dirty” Usability Scale. In: Jordan, P., Thomas, B., Weerdmeester, B., McClelland, I. (Hg.), Usability Evaluation in Industry. Taylor & Francis London, S. 189-194, 1996.
- Brusilovsky 1996 Brusilovsky, P.: Methods and techniques of adaptive hypermedia. In: User Modeling and User-Adapted Interaction, 6(2-3), S. 87-129, 1996.
- Caplan 1990 Caplan, S.: Using focus group methodology for ergonomic design. In: Ergonomics, 33(5), S. 527-533, 1990.
- Cohen et al. 2004 Cohen, M. H., Giangola, J. P., Balogh, J.: Voice User Interface Design. Addison-Wesley Professional, 368 S., 2004.
- Davis 1989 Davis, F.: Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. IN: MIS Quarterly, 13(3), S. 319-340, 1989.
- DEGS1 2009 Robert Koch-Institut (Hg.): DEGS – Studie zur Gesundheit Erwachsener in Deutschland. Projektbeschreibung. In: Beiträge zur Gesundheitsberichterstattung des Bundes. RKI, Berlin, S. 26-28, 2009.
- Dickinson et al. 2007 Dickinson, A., Arnott, J., & Prior, S.: Methods for human – computer interaction research with older people. In: Behaviour & Information Technology, 26(4), 343-352, 2007.
- Dietrich et al. 1993 Dieterich, H., Malinowski, U., Kühme, T., & Schneider-Hufschmidt, M.: State of the Art in Adaptive User Interfaces. In: Adaptive User Interfaces: Principle and Practice. Amsterdam: Elsevier Science & Technology, S. 13-48, 1993.
- DIN 55350-11:2008-05 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, Beuth Verlag Berlin, 2008.
- Dix et al. 2004 Dix, A., Finlay, J., Abowd, G. D., Beale, R.: Human-Computer Interaction. 3. Auflage, Pearson, Harlow (England), S. 318-364, 1993.
- Fern 1982 Fern, E. F.: The use of focus groups for idea generation: The effects of group size, acquaintanceship, and moderator on response quantity and quality. In: Journal of Marketing Research, 19, S. 1-13, 1982.
- Fisk 2009 Fisk, A. D.: Designing for older adults: Principles and creative human factors approaches (2nd). CRC Press, Boca Raton, 2009.
- Frantzidis 2009 Frantzidis, C. A., & Bamidis, P. D.: Description and Future Trends of ICT Solutions Offered Towards Independent Living: The Case of LLM Project. In: PETRA '09, Proceedings of the 2Nd International Conference on Pervasive Technologies Related to Assistive Environments. New York, ACM, S. 59:1-59:8, 2009.
- Friesdorf 2007 Friesdorf, W., Heine, A., & Mayer, D. Sentha, seniorenrechtliche Technik im häuslichen Alltag: Ein Forschungsbericht mit integriertem Roman. Berlin, New York: Springer, 2007.

- Fuchsberger 2008 Fuchsberger, V.: Ambient assisted living: elderly people's needs and how to face them. In: SAME '08 Proceedings of the 1st ACM international workshop on Semantic ambient media experiences, New York, ACM, S. 21-24, 2008.
- GEDA 2012 Daten und Fakten: Ergebnisse der Studie „Gesundheit in Deutschland aktuell 2010“. In: Beiträge zur Gesundheitsberichterstattung des Bundes. Robert Koch-Institut, Berlin, 2012.
- Hackman & Biers 1992 Hackman, G. S., Biers, D.W.: Team Usability Testing: Are Two Heads Better than One? In: Proc. Human Factors Society 36th Annual Meeting, S. 1205-1209, 1992.
- Hanna 1997 Hanna, L., Ridsen, K., & Alexander, K.: Guidelines for usability testing with children. In: interactions, 4(5), 9-14, 1997.
- Hanson 2010 Hanson, V. L.: Influencing technology adoption by older adults. In: Interacting with Computers, 22(6), S. 502-509, 2010.
- Hassenzahl 2003 Hassenzahl, M.: The thing and I: understanding the relationship between user and product. In: Blythe, M., Overbeeke, C., Monk, A. F., P. C. Wright (Hg.), Funology: From usability to enjoyment, Kluwer Dordrecht, S. 31-42, 2003.
- Hassenzahl 2006 Hassenzahl, M.: Hedonic, Emotional, and Experiential Perspectives on Product Quality. In: C. Ghaoui (Hg.), Encyclopedia of Human Computer Interaction, Idea Group, S. 266-272, 2006.
- Hassenzahl, Burmester & Koller 2003 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: J. Ziegler, G. Szwillus (Hg.): Mensch & Computer 2003. Interaktion in Bewegung. B. G. Teubner, Stuttgart, S. 187-196, 2003.
- Hawthorn 2000 Hawthorn, D.: Possible implications of aging for interface designers. In: Interacting with Computers, 12(5), S. 507-528, 2000.
- Hegner, 2003 Hegner, M.: Methoden zur Evaluation von Software. In: IZ-Arbeitsbericht Nr. 29, Hemsens, 2003.
- Hekkert 2006 Hekkert, P.: Design Aesthetics: Principles of Pleasure in Product Design. In: Psychology Science, 48(2), S. 157-172, 2006.
- Hill, 2011 Hill, R. L., Dickinson, A., Arnott, J. L., Gregor, P., & McIver, L.: Older Web Users' Eye Movements: Experience Counts. In: CHI '11, Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM, New York, S. 1151-1160, 2011.
- Holzinger 2005 Holzinger, A.: Usability Engineering Methods for Software Developers. In: Communications of ACM, 48(1), S. 71-74, 2005.
- Hone & Graham 2000 Hone, K. S., Graham, R.: Towards a Tool for the Subjective Assessment of Speech System Interfaces (SASSI). In: Natural Language Engineering, 6(3/4), S. 287-303, 2000.
- Ijsselstein 2008 Ijsselsteijn W., Poels K. und Kort, Y.: Game Experience Questionnaire. FUGA The fun of gaming: Measuring the human experience of media enjoyment. Technical Report, Deliverable 3.3., 2008.
- ISO 9241-11 1998 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.
- ISO 9241-210 2010 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, 2010.

- ITG 2011 VDE/ITG (Hg.): VDE-ITG-Richtlinie ITG 2.1-01. Messung und Bewertung der Usability von Kommunikationsendeinrichtungen. VDE Verlag Berlin, 2011.
- Jameson 2008 Jameson, A.: Adaptive Interfaces and Agents. In: Sears, A., Jacko, J. A. (Hg.): The human-computer interaction handbook: Fundamentals, evolving technologies and emerging applications (2nd ed.). Boca Raton, CRC Press, S. 433-458, 2008.
- John et al. 2004 John, B. E., Prevas, K., Salvucci, D. D. & Koedinger, K.: Predictive Human Performance Modeling Made Easy. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, ACM, New York, S. 455-462, 2004.
- Kieras & Polson 1985 Kieras, D., Polson, P.: An Approach to the Formal Analysis of User Complexity. In: Int. J. Man-Machine Stud., 22(4), S. 365-394, 1985.
- Kieras 2007 Kieras, D. E.: The control of cognition. In: Gray, W. (Hg.), Integrated models of cognitive systems. Oxford University Press, S. 327-355, 2007.
- Kirakowski & Corbett 1996 Kirakowski, J., Corbett, M.: SUMI: The Software Usability Measurement Inventory. In: British Journal of Education Technology, 24(3), S. 210-212, 1993.
- Kurth 2012 Kurth, B. M., Erste Ergebnisse der „Studie zur Gesundheit Erwachsener in Deutschland“ (DEGS). In: Bundesgesundheitsblatt 55(8), S. 980-990, 2012.
- Lederer & Prasad 1992 Lederer A. L., Prasad, J.: Nine management guidelines for better cost estimating. In: Communications of the ACM, 35(2), ACM, New York, S. 51-59, 1992.
- Lehner & Prieler-Woldan 2008 Lehner M., Prieler-Woldan M.: Projekt Smart Home – Eine Akzeptanzanalyse - Report. FH OÖ Fakultät für Gesundheit und Soziales Linz, Österreich, 2008. Online (04.12.2013): http://www.betreuungundpflegeimalter.net/index.php/component/docman/doc_download/57-download-smart-home-bericht?Itemid=62
- Lewis & Sauro 2009 Lewis, J. R., Sauro, J.: The Factor Structure of the System Usability Scale. In: Proc. HCI International 2009, San Diego CA, Springer, S. 94-103, 2009.
- Lin et al. 1997 Lin, H. X., Choong, Y., Salvendy, G.: A Proposed Index of Usability: A Method for Comparing the Relative Usability of Different Software Systems. In: Behavior and Information Technology, 16(4/5), S. 267-278, 1997.
- Lydecker 1986 Lydecker, T. H. Focus group dynamics. In: Association Management, 38(3), S. 73-78, 1986.
- Manzeschke et al. 2013 Manzeschke, A., Weber, K., Rother, E., Fangerau, H.: Ethische Fragen im Bereich Altersgerechter Assistenzsysteme, Druckerei Thiel Gruppe Ludwigsfelde, 2013. Online (16.06.2013): http://www.mtidw.de/grundsatzfragen/begleitforschung/dokumente/ethische-fragen-im-bereich-altersgerechter-assistenzsysteme-1/at_download/file
- Mohs et al. 2006 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“. In: MMI-interaktiv Journal, 11, S. 75-84, 2006.
- Möller 2005 Möller, S.: Quality of Telephone-Based Spoken Dialogue Systems, Springer, New York, 2005.
- Möller et al. 2006 Möller, S., Krebber, J., Smeele, J.: Evaluating the Speech Output Component of a Smart-Home System. In: Speech Communication 48(1), S. 1-27, 2006.
- Newell, 1990 Newell, A.: Unified Theories of Cognition. Harvard University Press, Cambridge, 1990.
- Nielsen 1993 Nielsen, J.: Usability engineering. Kaufmann, San Francisco, 1993.
- Nielsen 1994 Nielsen, J.: Heuristic Evaluation. In: Nielsen, J., Mack, R. L. (Hg.), Usability Inspection Methods, John Wiley & Sons, New York, S. 25-62, 1994.

- Nielsen 1997 Nielsen 1997 – Nielsen, J.: The use and misuse of focus groups. In: IEEE Software, 14(1), S. 94-95, 1997.
- Paramythis et al. 2010 Paramythis, A., Weibelzahl, S., & Masthoff, J.: Layered evaluation of interactive adaptive systems: framework and formative methods. In: User Modeling and User-Adapted Interaction, 20(5), 383-453, 2010.
- Preece et al. 1994 Preece, J., Rogers, Y., Sharp, H., Benyon, D., Holland, S., Carey, T.: Human-Computer Interaction. Addison-Wesley, Wokingham (UK), 1994.
- Schleicher & Trösterer 2009 Schleicher, R., Trösterer, S.: The ‘Joy-of-Use’-Button: Recording Pleasant Moments While Using a PC. In: Human-Computer Interaction – INTERACT 2009, Springer, Heidelberg, S. 630-633, 2009.
- Schmidt 2000 Schmidt, A.: Implicit human computer interaction through context. In: Personal Technologies 4 (2-3), Springer, S. 191-199, 2000.
- Sonntag et al. 2009 Sonntag, D., Jacobs, O., Weihrach, C.: Usability Guidelines for Use Case Applications. In: Theseus Report CTC WP4, Task 4.1, MS3, April 30, Deutsches Forschungsinstitut für Künstliche Intelligenz, Saarbrücken, 2009.
- Specht 1998 Specht, M.: Adaptive Methoden in computerbasierten Lehr/Lernsystemen. Dissertation: Universität Trier, 149 S., 1998.
- Spencer 2009 Spencer, D., & Garrett, J. J.: Card sorting: Designing usable categories. Rosenfeld Media, Brooklyn, New York, 2009.
- Sturm 2005 Sturm, J. A.: On the Usability of Multimodal Interaction for Mobile Access to Information Services. Dissertation: PrintPartners Ipskamp, Nijmegen, 2005.
- Van Velsen et al. 2008 Van Velsen, L., Van Der Geest, T., Klaassen, R., & Steehouder, M.: User-centered evaluation of adaptive and adaptable systems: a literature review. In: The Knowledge Engineering Review, 23(3), S. 261-281, 2008.
- Wechsung 2014 Wechsung, I.: An Evaluation Framework for Multimodal Interaction. Dissertation: Technische Universität Berlin, 191 S., Springer International Publishing Schweiz, 2014.
- Wharton et al. 1994 Wharton, C., Riean, J., Lewis, C., Polson, P.: The Cognitive Walkthrough Method: A Practitioner’s Guide. In: Nielsen, J. and Mack, R. L. (Hg.), Usability Inspection Methods, John Wiley & Sons, New York, S. 105-140, 1994.
- Wilson 2007 Wilson, A. D.: Sensor- and Recognition-Based Input for Interaction. In: Sears A. und Jacko J. (Hg.), The Human Computer Interaction Handbook, Lawrence Erlbaum Associates, S. 177-199, 2007.