UNIVERSITÄT DES SAARLANDES

FR. 5.6 INFORMATIONSWISSENSCHAFT

ERSTGUTACHTER Prof. Dr. Ilse Harms

ZWEITGUTACHTER PD Dr. Christoph Igel

ZUM NUTZUNGSKONTEXT MOBILER ENDGERÄTE

KOGNITIVE ASPEKTE UND IMPLIKATIONEN FÜR DIE USABILITY

MAGISTERARBEIT

VORGELEGT IM

März 2010

VORGELEGT VON

Daniel Barth
Allerstraße 17
12049 Berlin

Mobil 0176 – 600 210 18
Mail contact@danielbarth.net

Matrikelnummer 2023932

EIDESSTATTLICHE VERSICHERUNG

Ich versichere hiermit an Eides statt, dass ich die vorliegende Arbeit selbständig verfasst, weder in dieser noch in ähnlicher Form bisher einer Prüfungskommission vorgelegt und keine anderen als die angegebenen Hilfsmittel verwendet habe.

Sämtliche Stellen der Arbeit, die wörtlich oder sinngemäß aus Veröffentlichungen entnommen sind, habe ich durch Quellenangaben kenntlich gemacht.

Dies gilt auch für Zeichnungen, Skizzen, bildliche Darstellungen und dergleichen sowie für Quellen aus dem Internet.

DATUM UND ORT UNTERSCHRIFT

INHALTSVERZEICHNIS

1 Einführung....5
1.1 Thematik....5
1.2 These....6
1.3 Vorgehen....6
2 Mobile Endgeräte....8
2.1 Begriffsbestimmung....8
2.1.1 Eingrenzung und Definition....8
2.1.2 Ergänzungen zur aktuellen Situation....10
2.1.3 Zusammenfassung....13
2.2 Status Quo....13
2.3 Entwicklungspotential....14
2.4 Folgerungen....16
3 Grundlagen der Usability....19
3.1 Hintergründe....19
3.2 Ziele....20
3.3 Gestaltungsgrundsätze....21
3.4 Usability Engineering und Evaluation....27
3.4.1 Usability Engineering....27
3.4.2 Usability Evaluation....29
3.5 Methodenüberblick....31
3.5.1 expertenzentrierte Methoden....32
3.5.2 nutzerzentrierte Methoden....36
4 Besonderheiten mobiler Endgeräte....45
4.1 Technische Besonderheiten....46
4.2 Besonderheiten der Zielgruppe....48
4.3 Besonderheiten der Nutzungsweise....48
4.4 Nutzungskontext....49
4.4.1 externe und interne Validität....54
5 Betrachtung kognitiver Aspekte....59
5.1 Gedächtnis und Aufmerksamkeit....60
5.2 Automatisierung....63
5.3 Dual-Task Paradigma....64

5.4 Aufgabenwechsel....65
5.5 Unterbrechungen....66
6 Implikationen....67
6.1 Implikationen für das Usability Engineering....67
6.1.1 Automatisierung unterstützen....67
6.1.2 Multimodalität anbieten....69
6.1.3 Interaktionsmöglichkeiten nutzen....71
6.1.4 Context-Awareness erwägen....73
6.1.5 Wiederaufnahme erleichtern....75
6.2 Implikationen für die Usability Evaluation....76
7 Fazit....81

8 Anhang....82
8.1 Übersichtskarten zur Abdeckung UMTS/HSDPA....82
9 Quellenverzeichnis....83
10 Danksagungen....89

1 EINFÜHRUNG

Bei der Analyse der wichtigsten Trends zukünftiger technologischer Entwicklung werden mobile Anwendungen und das mobile Internet an vorderster Stelle genannt. Hinsichtlich strategischer Dimensionen wie der Zahl der Geräte mit Internetzugang, der Anzahl verfügbarer Applikationen, der Größe der Entwickler-Community und nicht zuletzt der Zahl der Nutzer übertreffen mobile Geräte die Personal Computer bereits – oder stehen kurz davor, es zu tun [Gens 2009, S. 7]. Im Ausblick auf die Entwicklungen des Jahres 2010 gelangt das Marktanalyse-Unternehmen IDC zur Überzeugung: „It will be a watershed year in the ascension of mobile devices as strategic platforms for commercial and enterprise developers as over 1 billion access the Internet, iPhone apps triple, Android apps quintuple, and Apple's "iPad" arrives.“ [ebenda, S. 1]. Trotz dieses starken Einflusses mobiler Technologien auf den Alltag von Milliarden Nutzern weltweit, werden die Besonderheiten dieser Geräte in etablierten Usability-Verfahren nicht selten unzureichend berücksichtigt. Dieser Umstand scheint insbesondere hinsichtlich der Usability Evaluation relevant, da die Aussagekraft erhobener Daten für die Nutzungspraxis in Zweifel gezogen werden kann.

1.1 THEMATIK

Ausgangspunkt der Betrachtung sind die spezifischen Eigenschaften des Nutzungskontextes, durch die mobile Endgeräte sich signifikant von stationären Desktop-Systemen unterscheiden. Hiervon ausgehend stehen zwei Fragestellungen im Zentrum: 1. Welche Implikationen ergeben sich aus diesen Unterschieden hinsichtlich Usability Engineering und Evaluation? 2. Welchen Beitrag kann die Betrachtung kognitiver Anforderungen des mobilen Kontextes zur Optimierung der Usability leisten? Eine besondere Relevanz erhält die Thematik aufgrund der weiten Verbreitung, der hohen Nutzungsintensität und des prognostizierten dynamischen Wachstums des Sektors mobiler Endgeräte. Jede erzielbare Optimierung im Usability-Prozess mobiler Endgeräte ist somit von besonderer Tragweite, da ein deutlich positiver Impact für den Alltag weiter Bevölkerungskreise zu erwarten ist.

1.2 THESE

Die Entwicklung des Sektors mobiler Endgeräte stellt Usability Engineering und Evaluation vor Herausforderungen. Der hohen Motivation zur Optimierung stehen spezifische Probleme bei der Umsetzung entgegen. Mobile Geräte werden in anderer Weise, an anderen Orten und in anderen Kontexten genutzt als stationäre PCs. Eine einfache Übernahme der im Desktop-Bereich etablierten und bewährten Methoden greift somit möglicherweise zu kurz und wird der spezifischen Situation mobiler Nutzung unzureichend gerecht. Kontrollierte statische Laborversuche werfen zudem Fragen hinsichtlich der externen Validität der Testergebnisse auf. Feldstudien in realen Nutzungssituationen sind hingegen aufwendig in der Durchführung, insbesondere im mobilen Kontext. Zudem sind die resultierenden Daten für die quantitative Auswertung häufig wenig geeignet.

Ein besseres Verständnis des Nutzungskontexts mobiler Endgeräte und der resultierenden spezifischen kognitiven Anforderungen an den Nutzer kann hilfreiche Impulse für Usability Engineering und Usability Evaluation bieten. Eine Anpassung mobiler Endgeräte an die kognitiven Ressourcen des Nutzers kann zu Entlastungen bei der Interaktion beitragen. Die Ableitung konkreter Gestaltungsempfehlungen, die bestehende Kenntnisse sinnvoll ergänzen und in den Designprozess einfließen können, scheint möglich. Sofern kognitive Belastungen realer Nutzungsszenarien bekannt sind, können psychologische Techniken zudem eventuell genutzt werden, um durch Simulation die Realitätsnähe von Testszenarien im Labor zu erhöhen.

1.3 VORGEHEN

Im nächsten Abschnitt werden zunächst relevante Begriffe geklärt und der Betrachtungsgegenstand eingeschränkt. Hieran schließt sich eine Betrachtung der gegenwärtigen Situation und prognostizierten Entwicklung des Bereichs mobiler Endgeräte an, welche die Relevanz der Thematik illustriert. Eine kompakte Darstellung der Grundlagen der Usability und ein Überblick der wichtigsten etablierten Methoden bildet den Ausgangspunkt für weitere Betrachtungen. Im Folgenden werden die Beson-

derheiten mobiler Endgeräte im Vergleich zu Desktop-PCs dargestellt und deren Auswirkungen auf die Usability skizziert. Einen zentralen Stellenwert nehmen hierbei der Nutzungskontext und die sich daraus ableitenden Implikationen für die Validität von Laborstudien ein. Es schließt sich eine Auseinandersetzung mit jenen ausgewählten kognitiven Aspekten an, die sich aus den Anforderungen des mobilen Nutzungskontexts ergeben. Abschließend werden die Implikationen dieser kognitiven Aspekte auf Usability Engineering und Evaluation diskutiert und Arbeiten zu ausgewählten Problemstellungen vorgestellt.

2 MOBILE ENDGERÄTE

2.1 BEGRIFFSBESTIMMUNG

In erster Näherung umfasst der Begriff „mobile Endgeräte“ eine unüberschaubare und im höchsten Maße heterogene Menge von Produkten. Allein im Bereich *consumer electronics* finden sich beispielsweise so unterschiedliche Geräte wie Handys, mobile Computer, Digitalkameras, Camcorder, MP3-Player, Navigationsgeräte, Taschenrechner oder tragbare Spielkonsolen, die alle im weitesten Sinne tragbar sind und mobil genutzt werden können.

2.1.1 EINGRENZUNG UND DEFINITION

Aufgrund dieser Diversität ist zunächst eine Eingrenzung des Betrachtungsgegenstands notwendig, um im weiteren Verlauf sinnvolle Aussagen treffen zu können. Eine Kategorisierung der Gesamtheit möglicher Geräte auf Basis gemeinsamer Merkmale bildet hierbei den ersten Schritt. Auf Grundlage der für Nutzungskontext und -zweck relevanten Merkmale entstehen so Kategorien, die funktionale Ähnlichkeiten abbilden. Im Rahmen dieser Arbeit sind insbesondere jene Geräte von Interesse, die selbst als Schnittstelle dienen und den Zugang zu Inhalten und Anwendungen ermöglichen. Aspekte der Usability dieser Geräte erlangen besondere Bedeutung, da sie die Qualität der Interaktion mit allen zur Verfügung gestellten Inhalten und Anwendungen maßgeblich beeinflussen.

Scott Weiss nimmt eine Kategorisierung mobiler Geräte auf Basis sinnvoller Merkmale vor. Die Arbeitsdefinition mobiler Endgeräte erfolgt daher in enger Anlehnung an seine Systematik. Die im englischen Sprachraum als „handheld“ bezeichneten Geräte lassen sich beschreiben als „extremely portable, self-contained information

management and communication devices“ [Weiss 2002, S. 2]. Die betrachteten Geräte müssen somit drei zentralen Kriterien genügen:

- Sie müssen in der Hand gehalten verwendbar sein.
- Sie müssen ohne Kabelverbindungen verwendbar sein.
- Sie müssen das Hinzufügen zusätzlicher Applikationen gestatten und / oder über einen Internetzugang verfügen.

Mobil wird im Kontext dieser Arbeit somit nicht im Sinne von „portabel“ verwendet, Laptops, Notebooks und auch die besonders leichten und kompakten Netbooks fallen daher nicht in die betrachtete Kategorie. Im Zentrum stehen vielmehr Geräte, die mobil *genutzt* werden, also unterwegs, im öffentlichen Raum und häufig auch in Bewegung. Der Begriff „handheld“ beschreibt dieses zentrale Abgrenzungskriterium anschaulich: die betrachteten Geräte werden während der Nutzung in der Hand gehalten und ein- oder beidhändig bedient. Dieses Merkmal ist hinsichtlich der Usability von wesentlicher Bedeutung, da sich hieraus im Vergleich zu Desktop-Computern[1] grundsätzliche Unterschiede bezüglich Nutzungskontext, Nutzungsweise und tendenziell auch Nutzungszweck ergeben. Ein Laptop stellt eine kompakte und integrierte Version eines Desktop-Computers dar und folgt hinsichtlich stationärer *Nutzung* sowie der Interaktion über GUI[2] mittels Tastatur und Zeigegerät dem gleichen Interaktionsparadigma. Die im folgenden betrachteten Geräte weichen hingegen deutlich vom etablierten WIMP-Paradigma[3] ab. Das zusätzliche Kriterium Internetfähigkeit und/oder Erweiterbarkeit durch die Installation neuer Programme dient darüber hinaus zur Abgrenzung von tragbaren elektronischen Geräten wie Kameras, MP3-Playern oder Navigationsgeräten. Die Forderung nach einem kabellosen

1 Als Desktop-Computer werden im Kontext dieser Arbeit alle Systeme bezeichnet, die kabelgebunden (mindestens zur Energieversorgung) und stationär am Arbeitsplatz eingesetzt werden. Sie verfügen über eine Zentraleinheit, Bildschirm, Tastatur und ein Zeigegerät (Maus, Trackball, Touchpad, etc.). Diese Komponenten können hierbei teilweise in ein gemeinsames Gehäuse integriert sein (siehe z.B. Apple iMac). Die Interaktion erfolgt über eine grafische Benutzeroberfläche, die per Tastatur und Zeigegerät bedient wird.

2 „Graphical User Interface“, grafische Benutzeroberfläche

3 WIMP GUIs sind grafische Benutzeroberflächen, die auf Basis des „point-and-click“-Prinzips bedient werden. Die Bezeichnung WIMP steht als Akronym für die zentralen Elemente, auf denen sie basieren: „**w**indows, **i**cons, **m**enus and **p**ointing device“. Eingeführt wurde WIMP mit dem *Xerox PARC*, populär durch den *Macintosh* ab 1984 und von Microsoft mit *Windows* übernommen – und noch heute bestimmt WIMP die Interaktion mit Desktop-Rechnern [vgl. van Dam 1997, S. 63f.].

Betrieb unterstreicht zusätzlich die Bedeutung von Bewegung und Bewegungsfreiheit im vertretenen Konzept mobiler Nutzung.

2.1.2 ERGÄNZUNGEN ZUR AKTUELLEN SITUATION

Weiss identifiziert drei Geräteklassen, welche die definierten Kriterien erfüllen: Mobiltelefone, Personal Digital Assistants (PDAs) und Pager [Weiss 2002, S. 4]. Diese Auswahl bedarf im Kontext der vorliegenden Arbeit in zweierlei Hinsicht einer Relativierung. Zum einen konnten Pager sich in Deutschland nicht durchsetzen[4] und spielten hierzulande daher nie eine dem US-Markt vergleichbare Rolle. Zum anderen spiegelt die Auswahl die Situation zum Zeitpunkt ihrer Formulierung im Jahre 2002 wider. Gerade die betroffenen Geräteklassen unterliegen einem dynamischen Entwicklungsprozess, so dass in den seither vergangenen Jahren relevante Neuerungen und Verschiebungen zu verzeichnen sind.

Im Teilbereich der Mobiltelefone hat sich mit den „Smartphones“ eine neue Geräteklasse etabliert, deren Verbreitung derzeit dynamische Wachstumsraten verzeichnet. So hat sich der Absatz in Deutschland in den ersten drei Quartalen 2009 laut Branchenverband BITKOM auf 1,6 Millionen Geräte mehr als verdoppelt, für 2010 wird eine Absatzmenge von über fünf Millionen prognostiziert[5]. Smartphones verfügen gegenüber normalen Mobiltelefonen über größere Displays, häufig in Kombination mit Touchscreens, was die Darstellung und Verwendung von Applikationen und Webinhalten unterstützt. Leistungsstärkere Prozessoren, umfangreiche verfügbare Anwendungen und zusätzliche Funktionen wie GPS und WLAN-Unterstützung heben sie häufig weiter von Standardmodellen ab.

Aufgrund dieses Funktionsspektrums kann de facto von einer Fusion von Mobiltelefon und PDA gesprochen werden, klassische PDAs werden somit durch die auch als „converged mobile devices“ [Gens 2009, S. 7] bezeichneten Smartphones substitu-

4 siehe z.B. FOCUS Magazin Nr. 26 von 1998, Meldung online verfügbar unter: http://www.focus.de/finanzen/news/telekommunikation-pager-haben-ausgequixt_aid_170722.html

5 Quelle: BITKOM Pressemitteilung vom 23.10.2009 http://www.bitkom.org/de/presse/62013_61548.aspx

iert[6]. Diese Entwicklung wird anschaulich illustriert durch die Tatsache, dass die ehedem im Bereich von Pagern bzw. PDAs führenden Unternehmen *Research In Motion* (RIM) und *Palm* mittlerweile ausschließlich Smartphones anbieten[7]. Insbesondere im Hinblick auf ihre Sonderstellung als multifunktionale integrierte Geräte sollten Smartphones aus heutiger Sicht eine zentrale Rolle im Rahmen der betrachteten mobilen Endgeräte spielen.

Neben den Smartphones sind weitere technische Neuerungen in der Peripherie des Betrachtungsgegenstands zu beobachten, die mittelfristig zu mobilen Endgeräten im Sinne der dargelegten Definition reifen können oder diese bereits erfüllen. Hierzu zählen beispielsweise Geräte wie der *iPod Touch*[8] von Apple, der sich einer einfachen Kategorisierung entzieht. Aufgrund des Funktionsumfangs, der auch Email-Client und Webbrowser umfasst, kann er nicht erschöpfend als Media-Player charakterisiert werden. Aufgrund der verfügbaren Applikationen und der Bedienung in der Hand erfüllt er zudem bereits die formulierten Kriterien, darüber hinaus kann per WLAN eine Internetverbindung hergestellt werden. Es ist zu erwarten, dass mittelfristig weitere vergleichbare Geräte angeboten werden.

Weitere Kandidaten stellen eBook-Reader wie beispielsweise der *Kindle*[9] von amazon dar. Dieses bisher nur in den USA verfügbare Gerät[10] verfügt in seiner aktuellen Version bereits über einen integrierten 3G[11]-Internetzugang, auch wenn dieser derzeit nur zum mobilen Herunterladen neuer Buchtitel aus amazons Kindle Store dient. Aufgrund kompakter Abmessungen, relativ geringen Gewichts und moderater Einstiegspreise könnten sich vergleichbare Geräte als weitverbreitete und ständige Begleiter etablieren. Ohnehin bereits für die Bedienung in der Hand konzipiert, qualifizieren sie

6 Bangeman, Eric: *Smartphones are leaving PDAs in the dust*. in: arstechnica.com, 10.10.2006 http://arstechnica.com/gadgets/news/2006/10/7944.ars

7 vgl. http://www.palm.com/de/de/ sowie http://www.rim.com/products/ bzw. http://de.blackberry.com/

8 Das Gerät ist bis auf die fehlende Mobilfunkunterstützung (GSM bzw. 3G) weitgehend mit dem iPhone identisch, verwendet das selbe Betriebssystem und verfügt über Anbindung an Appels App-Store und somit die Möglichkeit der Installation von Software von Drittanbietern. (vgl. http://www.apple.com/de/ipodtouch/features/)

9 vgl. http://www.amazon.com/Kindle-Wireless-Reading-Display-Generation/dp/B0015T963C/

10 Die Lieferung nach Deutschland ist möglich, derzeit bietet der Kindle Store jedoch nur englischsprachige Inhalte an.

11 Kurzbezeichnung für den Mobilfunkstandard der dritten Generation, der auch als UMTS bekannt ist. Beide Bezeichnungen werden in dieser Arbeit synonym verwendet. (vgl. http://de.wikipedia.org/wiki/Universal_Mobile_Telecommunications_System)

sich im Rahmen der vorgenommenen Definition, sobald der technisch bereits vorhandene Internetzugang offen – also zum Surfen im WWW und/oder Abrufen von Mails, etc. – genutzt werden kann.

Eine dritte Gerätekategorie, welche die Kriterien potentiell erfüllt, sind die sogenannten Tablet-PCs. Diese haben kürzlich durch die lang erwartete und von umfangreichen Spekulationen begleitete Einführung des Apple *iPad*[12] erhöhte Aufmerksamkeit erfahren. Das iPad selbst verfügt standardmäßig über WLAN-Funktionalität und ist mit optionaler 3G-Unterstützung verfügbar, mobiler Internetzugang ist somit gegeben. Es lässt sich kabellos verwenden, die Interaktion erfolgt über einen Touchscreen, die Installation neuer Applikationen ist einfach möglich. Aufgrund der Abmessungen von 24 x 19 cm empfiehlt es sich jedoch weniger als ständiger Begleiter für unterwegs, eine dauerhafte Nutzung aus der Hand scheint ebenfalls eher problematisch. Kompaktere internetfähige Tablets wie beispielsweise die Modelle von Archos[13] erfüllen hingegen bereits heute die formulierten Kriterien, weitere relevante Modelle sowie die zunehmende Verbreitung von 3G-Unterstützung sind zu erwarten.

Als letzte Kategorie, die im Rahmen der Definition Beachtung finden sollte, sind tragbare Spielkonsolen wie beispielsweise Nintendos *DSi* oder Sonys *PSP* (PlayStation Portable) zu nennen. Beide Geräte verfügen über integrierte Browser sowie Internetzugang via WLAN[14], werden in der Hand gehalten bedient und besitzen auch aufgrund der massenhaften Verbreitung eine gewisse Relevanz. Da in Analogie zu AppStore (Apple) und Kindle Store (amazon) auch für diese Geräte der Kauf neuer Inhalte per Download über eine eigene Plattform möglich ist, erscheint die Integration einer 3G-Unterstützung – wie in Kindle und iPhone bereits vorhanden – mittelfristig auch hier als sinnvolle Option. Dies würde den Bewegungsradius mit Internetzugang über verfügbare WLAN-Zellen hinaus erweitern und eine eigenständige Option für mobile Internetnutzung schaffen.

12 vgl. http://www.apple.com/de/ipad/specs/

13 vgl. http://www.archos.com/products/imt/index.html

14 vgl. http://www.opera.com/devices/ sowie http://en.wikipedia.org/wiki/PlayStation_Portable

2.1.3 ZUSAMMENFASSUNG

Abschließend kann festgehalten werden, dass die heutige Situation sich weniger klar strukturiert darstellt, als die 2002 von Weiss vorgenommene Dreiteilung in Mobiltelefone, Pager und PDAs. Durch neue Geräteklassen in der Peripherie der Definition kann nicht immer mit letzter Trennschärfe eine Zuordnung vorgenommen werden. In diesem Bereich entscheidet nicht zuletzt die Art der Nutzung über die Anwendbarkeit der im Folgenden dargestellten Betrachtungen. Dennoch dienen die drei zentralen Kriterien von Weiss weiterhin als sinnvolle definitorische Grundlage.

Aufgrund von Verbreitung und Nutzungsintensität und der daraus resultierenden Relevanz für große Teile der Weltbevölkerung ist nach wie vor der Bereich der Mobiltelefone von zentralem Interesse. Dies gilt umso mehr, als hier mit den Smartphones eine Geräteklasse von erheblichem Wachstumspotential vertreten ist, die durch ihren Funktionsreichtum bisherige Geräte wie PDAs substituiert bzw. integriert und sich als Gateway zu mobilen Applikationen, Web- und Medieninhalten zu etablieren scheint.

2.2 STATUS QUO

Laut einer EITO-Studie[15] überschritt die Zahl der Mobilfunknutzer weltweit im Jahr 2009 erstmals die Viermilliardengrenze. Die stärksten Wachstumsraten sind erwartungsgemäß in den Schwellen- und Entwicklungsländern zu verzeichnen. In der EU war 2009 ein moderates Wachstum (+4%) zu beobachten, mit aktuell rund 641 Millionen Mobilfunkverträgen wird die Zahl der Einwohner deutlich übertroffen.

Während sich die Industriestaaten somit langsam einer Sättigungsgrenze nähern, die das Gesamtwachstum verlangsamt, findet sich hier ein deutlich ausgeprägter Trend hin zu leistungsstärkeren Mobiltelefonen mit Internetfähigkeit und multimedialen Funktionen. Die Zahl der Geräte, die mit UMTS über einen breitbandigen mobilen

15 Das *European Information Technology Observatory* (EITO) ist ein Ableger des Branchenverbands BITKOM. Die Ergebnisse der Studie finden sich in zusammengefasster Form in einer Pressemitteilung vom 7.8.2009 (siehe http://www.eito.com/pressinformation_20090807.htm).

Internetzugang[16] verfügen, stieg in der EU 2009 um 36% auf etwa 172 Millionen, in den USA gar um 74% auf 108 Millionen Geräte. Der Anteil von GSM-Geräten, die noch über deutlich geringere Bandbreiten verfügen, ist entsprechend rückläufig. Der Anteil der Smartphones am Gesamtabsatz lag in Westeuropa im 3. Quartal 2009 bei 23% [MorganStanley 2009, S. 40]. Laut BITKOM-Präsident Scheer ist „mit der steigenden Verbreitung von Smartphones (...) die mobile Internetnutzung im geschäftlichen und privaten Umfeld zum Massenphänomen geworden."[17]

2.3 ENTWICKLUNGSPOTENTIAL

Ausgehend von der aktuellen Situation wird für die nahe Zukunft ein dynamisches Wachstum für den gesamten mobilen Telekommunikationsmarkt, insbesondere für die Verbreitung breitbandiger mobiler Datendienste, prognostiziert. Das Bankhaus Morgan Stanley gelangt beispielsweise im umfangreichen *Mobile Internet Report* im Dezember 2009 zu der Einschätzung: „Regarding pace of change, we believe more users will likely connect to the Internet via mobile devices than desktop PCs within 5 years" [MorganStanley 2009, S. 22]. Diese Erwartung spiegelt sich auch im Whitepaper zur Vorhersage des globalen Internettraffic des Netzwerkausstatters Cisco wider: „Globally, mobile data traffic will double every year through 2013, increasing 66x between 2008 and 2013" [Cisco 2009, S. 2].

Das Gesamtwachstum wird dabei maßgeblich von der zunehmenden Verbreitung von Smartphones getragen. In diesem Segment stiegen die Verkäufe weltweit allein im zweiten Quartal 2009 um 150% und setzten den Trend somit trotz widrigen Marktumfelds fort (der Gesamtbereich Telekommunikation verzeichnet für den gleichen Zeitraum einen Umsatzrückgang von 10,6%)[18]. In Japan – traditionell Vorreiter in der

16 Standardmäßig wird eine maximale Datenübertragungsraten von 384 kbit/s unterstützt, diese kann durch Verwendung von HSPDA auf bis zu 7,2 Mbit/s (downstream) gesteigert werden. (vgl. http://de.wikipedia.org/wiki/Universal_Mobile_Telecommunications_System)

17 BITKOM Pressemitteilung vom 13.10.2010: IT- und Telekommunikations-Trends 2010 http://www.bitkom.org/62121_62116.aspx

18 GfK Pressemitteilung zum GfK TEMAX Westeuropa zeigt im zweiten Quartal 2009, http://www.gfk.com/group/press_information/press_releases/004548/index.de.html

Nutzung mobiler Datendienste[19] – waren im dritten Quartal 2009 bereits 52% der neu ausgelieferten Mobiltelefone Smartphones (+4% zu Q3 2008), in Nordamerika und Westeuropa liegt die Quote mit 25% respektive 23% noch deutlich darunter, hier sind jedoch entsprechend höhere Zuwachsraten (+19% bzw. +15%) zu verzeichnen [MorganStanley 2009, S. 40].

Der Anteil von Geräten mit UMTS-Unterstützung (bzw. schnelleren Nachfolgern) wird deutlich zunehmen. Diese Entwicklung wird durch die Verbreitung der Smartphones begünstigt, doch gehört UMTS mittlerweile – von Einstiegsmodellen abgesehen – zur Standardausstattung moderner Mobiltelefone. Der weltweite Anteil wird sich den Prognosen zufolge von derzeit unter 20% bis 2014 auf etwa 43% erhöhen und dann gut 2,7 Milliarden Nutzer umfassen. Bei der globalen Verteilung bleiben erhebliche regionale Unterschiede bestehen, mit deutlich höherer Dichte in den Industrienationen (EU 2014: 92%, USA 2014: 74%). Japan spielt auch hier eine Vorreiterrolle, da bereits 2009 ein Anteil von über 90% erreicht wurde und binnen weniger Jahre eine nahezu vollständige Abdeckung zu erwarten ist [MorganStanley 2009, S. 25 f.].

Abschließend ist ein weiterer dynamischer Entwicklungstrend zu beobachten: Smartphones gestatten die Installation von Software von Drittanbietern (Third-Party Software), mit der sich der Funktionsumfang nach Belieben erweitern lässt. Aufgrund der massenhaften Verbreitung entwickeln sich die Betriebssysteme der Smartphones für Softwareentwickler zu Absatzmärkten von wirtschaftlicher und strategischer Bedeutung. Allein die Lager der etablierten Systeme – allen voran Symbian (Nokia, etc.),

19 Der mobile Datendienst i-mode, der in Japan von NTT docomo seit 1999 angeboten wird, überschritt bereits kaum ein Jahr nach Einführung die Grenze von 10 Millionen Nutzern und verfügte im Dezember 2009 über 49,7 Millionen Kunden, was etwa 39% der Gesamtbevölkerung entspricht. Damit gilt i-Mode unangefochten als „world's success story for mobile data communication“ [Weiss 2002, S. 25]. Die in Europa und den USA verfügbaren konkurrierenden WAP-Angebote konnten sich hingegen nie in der Breite am Markt etablieren – während in Japan bereits 2001 rund zweidrittel der Mobilfunkkunden auch mobile Datendienste nutzte, blieb deren Anteil im Rest der Welt weiterhin gering [vgl. Sigurdson 2001, S. 10f.].

Ursachen hierfür werden neben der Kostenstruktur vor allem auch in Defiziten hinsichtlich der Usability gesehen: „WAP has failed to serve effectively the market for which it was created, because of marketing, reliability and usability problems“ [Weiss 2002, S. 26]. Im Herbst 2000 wurde von der Nielsen Norman Group eine Feldstudie zur WAP-Nutzung durchgeführt. Nach dem Testzeitraum von einer Woche beantworteten 70% der Probanden die Frage, ob sie wahrscheinlich binnen eines Jahres selbst WAP nutzen werden, mit nein. Die Tester gelangen zu dem Fazit: „our basic conclusion is that WAP usability fails miserably; accomplishing even the simplest of tasks takes much too long to provide any user satisfaction“ [Ramsay Nielsen 2000, S. 6].

Quelle der i-Mode Kundenzahlen: NTT docomo (http://www.nttdocomo.com/about/operating/)

RIM (BlackBerry), Mac OS X (iPhone) und Windows Mobile (HTC, etc.) – sowie der gewichtige Neueinsteiger Android von google werden 2010 voraussichtlich um weitere über 200 Millionen neue Geräte wachsen [Gens 2009, S.7]. Für das kürzlich gestartete Android System stehen bereits 10.000 Anwendungen – im Smartphone-Kontext kurz *Apps* genannt – zum Download zur Verfügung [ebenda, S.7]. In Apples App Store für iPhone und iPod Touch stehen im Januar 2010 nach eigenen Angaben über 140.000 Apps[20] bereit. Gemäß einer Pressemitteilung des Unternehmens vom 5. Januar 2010 belief sich die kumulierte Zahl der Downloads in weniger als 18 Monaten seit Eröffnung des App Store auf über drei Milliarden[21].

Auf diese Weise etablieren Smartphones sich als universelle Plattform für fremde Anwendungen. Dieser Aspekt ist aus Usability-Perspektive von besonderem Interesse, da das mobile Endgerät zum Gateway zu allen auf ihm lauffähigen Applikationen wird. Anders als im Desktopbereich, wo Ein- und Ausgabegeräte gegen ergonomischere Modelle ausgetauscht, einem spezifischen Anwendungszweck angepasst oder auf individuelle Nutzerbedürfnisse zugeschnitten werden können, beeinflusst im mobilen Bereich die Usability des Geräts in erheblichem Maße die erreichbare Usability aller verwendeten Anwendungen.

2.4 FOLGERUNGEN

Bereits heute ist die Verwendung mobiler Endgeräte für viele Menschen selbstverständlicher Alltag. In Zukunft wird sich die Bedeutung dieser Geräteklasse zudem in mindestens zweierlei Hinsicht weiter vergrößern. Einerseits steigt die Zahl der Betroffenen durch die zahlenmäßige Zunahme der Geräte und ihre geografische Verbreitung hinein in bisher wenig erschlossene Gebiete, insbesondere Schwellen- und Entwicklungsländer.

Für die Nutzer in den Industriestaaten wird insbesondere eine zweite Entwicklungstendenz zu spürbaren Veränderungen in der Nutzung mobiler Endgeräte führen.

20 vgl. http://www.apple.com/de/ipad/app-store/

21 siehe http://www.apple.com/de/pr/library/2010/01/05appstore.html

Durch den wachsenden Funktionsumfang etablieren sich diese Geräte zunehmend als universelle multifunktionale Begleiter. Smartphones sind de facto bereits kompakte Computer, die ein breites Nutzungsspektrum eröffnen. Bisher primär an Desktop-PCs ausgeführte Aktivitäten werden sich zunehmend auf mobile Geräte verlagern. Dies erhöht zugleich die Nutzungsintensität der mobilen Endgeräte, da sie häufiger und über längere Zeiträume als bisher genutzt werden.

Diese Umstände verleihen der Thematik der Usability mobiler Endgeräte eine besondere Relevanz, sowohl aus Sicht der Nutzer als auch aus Perspektive der Unternehmen als Systemanbieter. Da die Zahl der Nutzer weiterhin wächst, sind stetig mehr Menschen betroffen. Diese werden zudem mehr Zeit in Interaktion mit mobilen Geräten verbringen, umfangreichere Aufgabenstellungen bearbeiten und komplexere Herausforderungen meistern müssen. Jede Erleichterung und Optimierung, die sich erzielen lässt, hat in globaler Perspektive immense Zeiteinsparungen, gesteigerte Effizienz und entsprechend reduzierte Kosten zur Folge. Auf privater Ebene wird der Alltag erleichtert, Zeit- und Kostenersparnis wird erzielt und nicht zuletzt wird Frustration vermindert oder vermieden, was sich in erhöhter Zufriedenheit mit dem verwendeten Produkt widerspiegeln sollte.

Aus Unternehmenssicht eröffnet gerade letzterer Einflussfaktor wirksame Möglichkeiten, das eigene Produktportfolio positiv von Mitbewerbern abzusetzen. Jakob Nielsen, einer der führenden Usability-Experten, betont dies, indem er auf die Effekte einzelner Produkte auf die Wahrnehmung ganzer Produkfamilien und letztlich des Unternehmens selbst hinweist: „For a company (...), the usability of each product will contribute to the company's general reputation as a quality supplier, and just a single product with poor usability can cause severe damage to the sales of the entire product family“ [Nielsen 1993, S. 72]. In dem Maße, in dem Geräte markenübergreifend vergleichbare Eigenschaften besitzen und das Hinzufügen neuer Funktionen kaum mehr eine sinnvolle Option darstellt, in der Gunst der Kunden zu punkten, können die Vorzüge eines Produkts hinsichtlich der Usability noch stärker als bisher zum entscheidenden Faktor für den Markterfolg avancieren.

Die Entwicklung des Sektors mobiler Endgeräte stellt Usability Engineering und Evaluation vor Herausforderungen. Aufgrund der weiterhin wachsenden Bedeutung

mobiler Endgeräte für eine immer größere Zahl von Menschen sollte der Optimierung der Usability eine hohe Priorität zugewiesen werden. Dies gilt umso mehr, als die Ergebnisse entsprechender Bemühungen allen Beteiligten – Systemanbietern wie Systemnutzern – zugute kommt.

3 GRUNDLAGEN DER USABILITY

Der Begriff Usability wird weit verbreitet und häufig genutzt, nicht selten als Schlagwort in Marketingkontexten. Der Bedeutungsinhalt wird dabei häufig mit „benutzerfreundlich“ oder auch „einfach“ und „intuitiv“ zu bedienen angegeben. Obgleich diese Aspekte ihre Berechtigung haben, werden sie dem Konzept der Usability unzureichend gerecht. Als deutsche Bezeichnung hat sich der Begriff „Gebrauchstauglichkeit“ etabliert [Dahm 2006, S. 133], der zugleich eine angemessenere Übersetzung aus dem Englischen darstellt. Im folgenden werden Usability und Gebrauchstauglichkeit synonym verwendet.

3.1 HINTERGRÜNDE

Den Entstehungshintergrund der Usability bildet die Ergonomie, die „Lehre von der menschlichen Arbeit, ihrer Beschreibung, Modellierung und Verbesserung“ [Dahm 2006, S. 28]. Im Mittelpunkt der Betrachtung steht dabei der Mensch, seine Aufgaben und Bedürfnisse. Laut DIN EN ISO 6385 ist Ergonomie definiert als „wissenschaftliche Disziplin, die sich mit dem Verständnis der Wechselwirkungen zwischen menschlichen und anderen Elementen eines Systems befasst (…) mit dem Ziel, das Wohlbefinden des Menschen und die Leistung des Gesamtsystems zu optimieren“ [zitiert nach Sarodnick 2006, S. 16]. Aus ergonomischer Perspektive wird ein komplexes Gesamtsystem betrachtet, welches alle relevanten Elemente der (Arbeits-)Umwelt berücksichtigt (Mensch, Werkzeug, Aufgabe, Umgebung) [vgl. ebenda, S. 17]. Entsprechend weit gefasst ist der Systembegriff.

Im Rahmen der Usability fokussiert sich die Betrachtung auf die Mensch-System-Interaktion, wobei der Systembegriff in diesem Kontext auf technische Systeme eingeschränkt wird. Darüber hinausgehende Faktoren können als Nutzungskontext Berücksichtigung finden. Der Betrachtungsgegenstand kann somit als Mensch-Maschine-Interaktion beschrieben werden. Der vielleicht populärere Begriff der Mensch-Computer-Interaktion bildet hier einen Schwerpunkt, die Grundsätze der

Usability sind jedoch für ein breiteres Spektrum anwendbar und können in weiten Teilen allgemein für die Optimierungen technischer Geräte herangezogen werden.

Verbindliche Vorgaben für die Usability sind in der internationalen Norm DIN EN ISO 9241 festgehalten. Diese ursprünglich unter dem Titel „Ergonomische Anforderungen für Bürotätigkeiten mit Bildschirmgeräten“ veröffentlichte Norm trägt heute die Bezeichnung „Ergonomie der Mensch-System-Interaktion“, was Universalität und Breite des Anwendungsspektrums der definierten Anforderungen unterstreicht [vgl. Sarodnick 2006, S. 31f. sowie Deutsches Institut für Normung (DIN), Normenausschuss Ergonomie (NAErg)[22]].

3.2 ZIELE

Allgemein gesprochen wird das Ziel verfolgt, die Mensch-System-Interaktion zu optimieren und höchstmögliche Gebrauchstauglichkeit technischer Geräte in konkreten Nutzungsszenarien herzustellen. Zur systematischen Erreichung dieser Ziele werden Leitkriterien, Heuristiken und spezifische Methoden herangezogen, die im Folgenden in kompakter Form dargestellt werden.

Die übergeordneten Leitkriterien der Usability sind *Effektivität*, *Effizienz* und *Zufriedenheit*. Diese sind in Teil 11 der ISO 9241 verbindlich festgelegt. Effektivität beschreibt hierbei, wie vollständig und genau ein bestimmtes Ziel erreicht wird. Effizienz setzt die erreichte Genauigkeit und Vollständigkeit ins Verhältnis zu dem dafür erforderlichen Aufwand. Zufriedenheit berücksichtigt die subjektive Beurteilung des Systems durch den Benutzer [Dahm 2006, S. 132]. Es ist hervorzuheben, dass Usability keine Produkteigenschaft darstellt, sondern sich im konkreten Nutzungskontext erschließt. Sie wird insofern verstanden als das Ausmaß, „in dem ein technisches System durch bestimmte Benutzer in einem bestimmten Nutzungskontext verwendet werden kann, um bestimmte Ziele (...) zu erreichen“ [Sarodnick 2006, S. 17]. Größtmögliche Gebrauchstauglichkeit ist somit nur erreichbar, wenn Nutzer, Aufgabe und System aufeinander abgestimmt sind.

22 http://www.naerg.din.de/

3.3 GESTALTUNGSGRUNDSÄTZE

Die definierten Leitkriterien geben die angestrebten Ziele vor, sind jedoch abstrakter Natur und nur bedingt operationalisierbar. Konkretere Anforderungen, die handlungsleitend herangezogen werden können, sind beispielsweise in Teil 110 der ISO 9241 als „Grundsätze der Dialoggestaltung“ zusammengefasst [vormals Teil 10, vgl. Sarodnick 2006, S. 32 sowie DIN, NAErg]. Der Dialog findet hierbei zwischen Nutzer und System statt, es handelt sich somit um Gestaltungsgrundsätze für die Benutzerschnittstelle der Mensch-Maschine-Interaktion. Diese sind allgemeingültig und unabhängig von der spezifischen Dialogtechnik. Für jeden Anwendungsfall werden konkrete Umsetzungen der allgemeinen Grundsätze abgeleitet. Sieben Faktoren werden aufgeführt, welche als maßgeblich für die Qualität der Interaktion angesehen werden. Diesen werden jeweils Empfehlungen sowie eine Veranschaulichung anhand von Beispielen zugeordnet. Die Faktoren lauten im einzelnen [vgl. Dahm 2006, S. 133ff. Sowie Sarodnick 2006, S. 31ff.]:

- *Aufgabenangemessenheit* | Der Dialog soll den Nutzer dabei unterstützen, seine Aufgabenstellung effektiv und effizient zu erledigen. So befreit die Dialoggestaltung den Nutzer idealer Weise von allen überflüssigen Informationen und Handlungsschritten. Es ist anzustreben, die Zahl erforderlicher Schritte zur Erreichung eines Ziels zu minimieren, beispielsweise durch die Eliminierung unnötiger Handlungsschritte, die Vorgabe sinnvoller Standardwerte und die weitgehende Automatisierung wiederkehrender Handlungssequenzen. Die Informationsdarstellung soll den Kontext der Aufgabe berücksichtigen und somit im Dialog beispielsweise nur relevante Informationen anzeigen, die im Zusammenhang mit der aktuellen Aufgabenstellung stehen.

- *Selbstbeschreibungsfähigkeit* | Der Dialog soll relevante Rückmeldungen auf Dialogschritte zur Verfügung stellen und den Nutzer über den Systemzustand informieren. Die Meldungen sollen für den Nutzer verständlich sein, sich also beispielsweise einer vertrauten Terminologie bedienen, natürlichsprachig und prägnant formuliert sein (z.B. Vermeidung kryptischer Fehlercodes) und sich idealer Weise dem Kenntnisstand des jeweiligen Nutzers anpassen. Der

Aspekt der Selbstbeschreibungsfähigkeit umfasst beispielsweise Hinweise des Dialogsystems dazu, in welcher Form Eingaben erwartet werden, Warnungen vor schwerwiegenden Folgen einer Handlung oder Ankündigungen zum Systemverhalten, z.B. lange Wartezeiten beim Laden komplexer Daten, sowie Information über den derzeitigen Systemzustand, z.B. zum aktuellen Ladefortschritt während eines Ladevorgangs.

- *Erwartungskonformität* | Die Dialoggestaltung soll den Erwartungen der Nutzer gerecht werden. Dies umfasst die Orientierung an allgemein anerkannten Standards und Konventionen, die Verwendung von Bezeichnungen und Terminologien, die gebräuchlich und dem Nutzer vertraut sind sowie die Konsistenz innerhalb des Dialogsystems selbst. (Quasi-)Standards ergeben sich beispielsweise aus Vorschriften, Konventionen des zugrundeliegenden Betriebssystems oder vertrauten ähnlichen Anwendungen. Bezeichnungen sollten sich an den Nutzererwartungen orientieren, indem beispielsweise die im Unternehmen gebräuchlichen Begriffe auch im Dialogsystem Verwendung finden. Konsistenz innerhalb des Systems wird unter anderem durch den einheitlichen Aufbau und gleichartiges Verhalten von Dialogen hergestellt. Ähnliche Arbeitsabläufe sollen zudem durch ähnliche Dialoge und Abläufe bewältigt werden.

- *Steuerbarkeit* | Der Dialog sollte jederzeit durch den Nutzer kontrolliert werden können. Der Nutzer initiiert den Dialogablauf und entscheidet über Richtung und Geschwindigkeit. Der Ablauf soll jederzeit durch den Nutzer unterbrochen oder komplett abgebrochen werden können, jeder Dialogschritt soll zudem rückgängig zu machen sein. Sofern alternative Eingabeformen unterstützt werden, soll die Auswahl dem Nutzer obliegen.

- *Individualisierbarkeit* | Der Dialog soll sich den Erfordernissen der konkreten Arbeitsaufgabe ebenso wie den individuellen Fähigkeiten und Bedürfnissen des Nutzers anpassen lassen. Dies wird beispielsweise erreicht durch konfigurierbare Menüs, individuelle Vorlagen oder Makros für wiederkehrende Aufgaben oder alternative Eingabemöglichkeiten für fortgeschrittene Anwender, die

z.B. Funktionen zeitsparend über Tastaturbefehle (Short-Cuts) aufrufen können. Die Wählbarkeit von Schriftgrößen oder Farbschemata für die Darstellung stellen weitere Möglichkeiten der Individualisierung eines Systems dar.

- *Lernförderlichkeit* | Ein Dialogsystem soll den Nutzer beim Erlernen seiner Funktionsweise unterstützen und Zeit und Aufwand des Lernprozesses minimieren. Idealer Weise unterstützt das System hierzu verschiedene Herangehensweisen des Nutzers wie exploratives oder verständnisorientiertes Lernen. Auf Anforderung des Nutzers sollen zu jedem Schritt detaillierte und verständliche Hilfestellungen zur Verfügung gestellt werden, die auch Schritt-für-Schritt-Anleitungen auf dem Weg zu angestrebten Zielen beinhalten. Auch im regulären Gebrauch können Hinweise auf Optimierungspotential erfolgen, die dem Nutzer fortgeschrittene Interaktionsformen (z.B. Verwendung von Tastaturkürzeln, Makros, etc.) vermitteln und ihn somit beim Erlernen des Systems weiter unterstützen.

- *Fehlertoleranz* | Das Dialogsystem soll trotz erkennbar fehlerhafter Eingaben des Nutzers zum beabsichtigten Ergebnis führen. Der hierfür erforderliche Korrekturaufwand seitens des Nutzers soll minimiert werden und erfolgt im Idealfall automatisch durch das System. So können Daten beispielsweise direkt bei Eingabe auf Plausibilität und Gültigkeit geprüft werden. Korrekturen sollten stets mit minimalem Aufwand möglich sein, z.B. durch Auswahl aus Korrekturvorschlägen. Sofern Korrekturen automatisch vorgenommen werden, sollte der Nutzer hierauf hingewiesen werden und die Änderungen gegebenenfalls leicht modifizieren können. Fehlerhaft ausgeführte Aktionen sollen ferner jederzeit rückgängig gemacht werden können.

Diese zentralen Anforderungen an die Gebrauchstauglichkeit stehen hierbei nicht unverbunden nebeneinander, sondern bedingen und beeinflussen sich wechselseitig. Exploratives Lernen wird beispielsweise durch eine hohe Fehlertoleranz unterstützt, weil probeweise durchgeführte Handlungsschritte bei Bedarf jederzeit rückgängig gemacht werden können. Da Rückmeldungen auf Fehleingaben idealer Weise auch

Hinweise auf mögliche Ursachen des Fehlers und Vorschläge zu Korrektur und Vermeidung enthalten, fördern sie das Verständnis des Systems und seiner Funktionsweise und spielen insofern auch hinsichtlich der Lernförderlichkeit eine Rolle.

Aus den Anforderungen können sich jedoch auch widersprüchliche Implikationen ergeben, so dass im Einzelfall Abwägungen vorzunehmen sind. So folgt aus der Forderung nach Aufgabenangemessenheit eine auf das Wesentliche und Relevante reduzierte Informationsdarstellung. Im Hinblick auf die Lernförderlichkeit sind jedoch Hilfestellungen und Hinweise auf alternative Interaktionsformen wünschenswert. Werden diese dargestellt, so muss der Nutzer eine größere Informationsmenge verarbeiten und Relevantes von Überflüssigem trennen. Wird auf die Darstellung verzichtet, so bleiben effizientere Formen der Bedienung dem Nutzer eventuell verborgen. Eine ähnliche Situation kann aus der Forderung nach Individualisierbarkeit und Erwartungskonformität entstehen. Ein im hohen Maße individualisierter Arbeitsplatz wird in vielerlei Hinsicht von vertrauten Standards abweichen. Für andere Nutzer ist ein solcher Arbeitsplatz nur eingeschränkt verwendbar, dem eigenen Nutzer kann hingegen durch Gewöhnung an die individuelle Konfiguration der Umgang mit alternativen, standardkonformeren Systemen erschwert werden.

Generell handelt es sich bei den vorgestellten Anforderungen um allgemeine Vorgaben dazu, *was* ein Dialogsystem leisten soll. Das *wie* der konkreten Umsetzung muss aus diesen wünschenswerten Zielzuständen abgeleitet und dem jeweiligen Anwendungsfall sowie der eingesetzten Dialogtechnik angepasst werden. Die ISO 9241 stellt somit allgemeingültige Forderungen und Empfehlungen zur Verfügung, beinhalten jedoch keine verbindlichen Vorgaben für konkrete Umsetzungen.

Neben den Gestaltungsgrundsätzen der ISO Norm finden sich umfangreiche Vorgaben und Empfehlungen in Form von Kriterienkatalogen, die als *Heuristiken* bezeichnet werden. Heuristiken beinhalten einen Satz „von Usability-Prinzipien (...), die auf bestimmte Problemkategorien bei der Gestaltung von Dialogsystemen hinweisen“ [Sarodnick 2006, S. 135]. Aus diesen Prinzipien leiten sich Richtlinien für die gebrauchstaugliche Gestaltung von Benutzerschnittstellen ab. Idealer Weise basieren Heuristiken auf empirischen Erkenntnissen und werden von Usability-Experten

systematisch zusammengestellt. Häufig finden darüber hinaus auch Erfahrungswerte aus der Praxis dieser Experten Eingang in Heuristiken.

Die Vielzahl verfügbarer Heuristiken und die nicht immer nachvollziehbare Herkunft und Validität der enthaltenen Prinzipien veranlasste Jakob Nielsen, einen der führenden Experten im Bereich der Usability, bereits 1994 zu einer Untersuchung der Erklärungskraft verschiedener Heuristiken. Als Resultat synthetisiert er aus den vorhandenen Heuristiken eine Liste der einflussreichsten Faktoren, die er in den folgenden sieben Kategorien strukturiert [Nielsen 1994, S. 153]:

1. *Visibility of system status* – enthält vor allem Forderungen nach Feedback des Systems (beispielsweise auf Benutzereingaben, Informationen zum Systemzustand und Rückmeldungen zum Bearbeitungszustand) und allgemeiner Sichtbarkeit von Objekten, Features und Handlungsfolgen. Es finden sich hier deutliche Schnittmengen zur *Selbstbeschreibungsfähigkeit* nach ISO 9241.

2. *Match between system and real world* – fordert den Gebrauch natürlicher Sprache und vertrauter Terminologien, den Einsatz von Metaphern und Konventionen aus der realen Welt, die Verwendung vertrauter konzeptioneller Modelle und ein Eingehen auf das vorhandene Hintergrundwissen der Nutzer. Diese Forderungen lassen sich zu großen Teilen in ähnlicher Form aus dem Grundsatz der *Erwartungskonformität* ableiten.

3. *User control and freedom* – umfasst Forderungen nach Initiation und Kontrolle von Handlungen durch den Nutzer, der Möglichkeit, Handlungen rückgängig machen zu können und der deutlichen Darstellung von Handlungsoptionen. Es sind somit starke Parallelen zum Grundsatz der *Steuerbarkeit* erkennbar.

4. *Consistency and standards* – fordert Konsistenz innerhalb des Systems und die Beachtung von Plattformkonventionen hinsichtlich der Interaktion. Auch diese Forderungen lassen ähnlich wie Punkt zwei in weiten Teilen aus dem Grundsatz der *Erwartungskonformität* ableiten.

5. *Error prevention* – fordert, bei der Systemgestaltung mögliche Fehler des Nutzers zu antizipieren und das Auftreten dieser Fehler durch entsprechende Vorkehrungen zu vermeiden bzw. automatisierte Korrekturen anzubieten. Es besteht eine Nähe zum Kriterium der *Fehlertoleranz*, doch fokussiert Nielsen wesentlich stärker auf die Fehlervermeidung.

6. *Recognition rather than recall* – repräsentiert die grundsätzliche Forderung, die kognitiven Ressourcen des Nutzers soweit möglich zu schonen. Es fällt Nutzern beispielsweise leichter, korrekte Eingabemöglichkeiten zu erkennen und ggf. zu modifizieren, als diese selbst aus der Erinnerung zu rekonstruieren. Die Dialoggestaltung sollte entsprechend automatisches Erkennen fördern (sinnvolle Vorgaben, Beispiele, etc.) und aufwendige Erinnerungsleistungen somit überflüssig machen. Alle für einen Handlungsschritt relevanten Informationen sollten ferner im jeweiligen Dialog ersichtlich sein, um nicht im Kurzzeitgedächtnis bereitgehalten werden zu müssen. Eine Minimierung der Gedächtnislast kann zudem durch eine auf das Wesentliche reduzierte Gestaltung erreicht werden. Hier ergeben sich Anknüpfungspunkte an den Grundsatz der *Aufgabenangemessenheit*.

7. *Flexibility and efficiency of use* – fordert die Anpassungsfähigkeit des Dialogsystems an den individuellen Nutzer und deckt sich in weiten Teilen mit dem Gestaltungsgrundsatz der *Individualisierbarkeit*.

Darüber hinaus diagnostiziert Nielsen drei weitere nachrangige Faktoren, die ebenfalls Einfluss auf die Usability haben und als weitere Heuristiken berücksichtigt werden. Er bezeichnet sie als 8. *aesthetic and minimalist design*, 9. *well-structured features that are easy to discriminate* und 10. *use of default values*.

Insgesamt spielen Heuristiken im Rahmen der Usability eine bedeutende Rolle, da sie in strukturierter Form validierte Gestaltungsrichtlinien (Guidelines) zur Verfügung stellen. Im Unterschied zu den *Grundsätzen der Dialoggestaltung* der ISO Norm bieten Heuristiken zudem häufig sehr detaillierte Anweisungen bis hin zu Vorgaben zu Schriftarten und -größen, Farben, Kontrasten, Textauszeichnungen oder der Verwendung illustrierender Grafiken. Dieser Detailgrad und die einhergehende erhöhte Spe-

zifität schränken den Anwendungsbereich von Heuristiken jedoch ein, so dass sie häufig nur für konkrete Anwendungsfälle Gültigkeit besitzen, beispielsweise für Desktop-Anwendungen oder Webseiten.

3.4 USABILITY ENGINEERING UND EVALUATION

3.4.1 USABILITY ENGINEERING

Usability Engineering kann als Prozess beschrieben werden, der unter systematischer Anwendung geeigneter Methoden zur Entwicklung gebrauchstauglicher Produkte führt. Idealer Weise begleitet das Usability Engineering hierbei den gesamten Lebenszyklus eines Produkts. Insbesondere in den Entwicklungsprozess sollte das Usability Engineering frühzeitig integriert werden, um bereits in der Analyse- und Konzeptionsphase – also vor der Gestaltung der eigentlichen Benutzerschnittstelle – Usability-Gesichtspunkte gezielt zu berücksichtigen [vgl. Sarodnick 2006, S. 81 sowie Nielsen 1993, S. 71]. Insbesondere in diesen frühen Phasen können Erkenntnisse des Usability Engineering einfach und mit geringem Kostenaufwand in die Produktgestaltung einfließen, da Fehlentwicklungen bereits im Vorfeld vermieden werden können und aufwendige Neugestaltungen und Änderungen somit entfallen. Die Bedeutung dieser zeitlichen Abfolge mit umfassender Analyse *vor* Aufnahme der Designaktivitäten wird in den Stufen des „usability engineering lifecycle models“ nach Nielsen betont, welches die folgenden Phasen unterscheidet [ebd., S. 72]:

1 Know the user
2 Competetive analysis
3 Setting usability goals
4 Parallel design
5 Participatory design
6 Coordinated design of the total interface
7 Apply guidelines and heuristic analysis
8 Prototyping
9 Empirical testing
10 Iterative design
11 Collect feedback from fielduse

Dieses Modell wird in der Praxis aufgrund zeitlicher und finanzieller Restriktionen nicht immer mit allen Phasen umgesetzt, dient aber dennoch als sinnvolle Gliederung eines idealtypischen Prozesses. Hervorzuheben ist die Bedeutung umfangreicher Analysephasen. Nielsen bezeichnet die Empfehlung *know the user* als „the most basic of all usability guidelines“ [Nielsen 1993, S. 73].

Wissen über den Nutzer wird in diesem Kontext als komplexes Konstrukt verstanden und erstreckt sich über mehrere Dimensionen. Individuelle Charakteristiken der Nutzer finden hier Beachtung. Darüber hinaus erfolgt eine detaillierte Analyse der übergeordneten Aufgaben der Nutzer, der derzeitigen Herangehensweise an die Aufgabenbewältigung und möglicher Verbesserungen dieses Ablaufs. Eine zusätzliche Betrachtung der Nutzerhandlungen auf rein funktionaler Ebene verschiebt den Fokus weg von der Frage *wie* ein Nutzer eine bestimmte Handlung ausführt hin zu den zugrundeliegenden Zielen und Motivationen dieser Handlung. Die Trennung in beobachtbare Oberflächenaktivität und funktionale Grundlage kann hierbei Anregungen für alternative, optimierte Handlungsfolgen bieten. Abschließend findet auch die Veränderung der Nutzer über die Zeit und die wechselseitige Beeinflussung von Nutzer, System und Nutzungsweise Berücksichtigung [vgl. ebd., S. 73ff.].

Das Ausmaß, in dem konkrete Informationen über den Nutzer erschlossen werden können, ist hierbei neben finanziellen und organisatorischen Rahmenbedingungen in hohem Maße davon abhängig, wie klar definiert, abgrenzbar und homogen zusammengesetzt die intendierte Zielgruppe des betrachteten Systems ist. Generell gilt jedoch, dass ein möglichst umfassendes Verständnis von Nutzer und Nutzungsweise eine wesentliche Grundlage des Usability-Engineering-Prozesses ist.

Auch mit detailliertem Wissen über den Nutzer ist das Ziel des Usability Engineering – die Entwicklung gebrauchstauglicher Produkte – jedoch nicht allein durch die Befolgung von Guidelines und Berücksichtigung angemessener Heuristiken erreichbar. Aufgrund der Komplexität der Systeme und individueller Eigenschaften konkreter Nutzer und Nutzungskontexte werden auch erfahrene Usability Experten häufig davon überrascht, wie weit antizipierte und tatsächliche Nutzeraktionen auseinanderliegen. Nielsen betont: „it is impossible to design an optimal user interface just by giving it your best try. Users have infinite potential for making unexpected misinter-

pretations of interface elements and for performing their job in a different way than you imagine“ [Nielsen 1993, S. 10]. Ferner unterscheiden sich die im Entwicklungsprozess beteiligten Experten durch ihre Fachkenntnis und ihr produktspezifisches Wissen deutlich von repräsentativen Nutzern. Hierdurch wird ihre Fähigkeit getrübt, sich in die Situation dieser weniger versierten Nutzer hineinzuversetzen und deren mögliche Schwierigkeiten vorherzusehen [vgl. ebenda, S. 13f.].

3.4.2 USABILITY EVALUATION

Im Entwicklungsprozess ist daher eine systematische Überprüfung der Design-Entscheidungen notwendig, die einen Gradmesser der Zielerreichung darstellt und als Grundlage für Anpassungen und Optimierungen dient. Diese Funktion erfüllt die *Usability Evaluation* und stellt damit einen zentralen Bestandteil des Usability Engineering dar. Voraussetzung für die Beurteilung eines Systems ist die Definition von „Evaluationszielen und -kriterien oder Fragestellungen (…), an denen die Systeme gemessen oder die Probleme interpretiert werden können“ [Sarodnick 2006, S. 115]. In erster Näherung lassen sich drei grundsätzliche Fragestellungen unterscheiden.
1. “Warum schlecht?“ zielt ab auf eine systematische Ermittlung von Schwachstellen, die im weiteren Entwicklungsverlauf zu beheben sind. 2. „Wie gut?“ hat die Bestimmung der Ausprägung von Systemeigenschaften zum Ziel. 3. „Which is better?“ stellt Vergleiche zwischen mehreren Systemen oder Systemvarianten auf Basis definierter Kriterien an und ermöglicht Entscheidungen auf Basis konkreter Daten der Evaluationsergebnisse [ebenda].

Auch die Evaluation begleitet den gesamten Entwicklungsprozess und erfolgt auf Basis eines breiten Spektrums von Methoden zur Gewinnung quantitativer und/oder qualitativer Daten (vgl. Abschnitt 3.5). Unter Rückgriff auf Nielsens *usability engineering lifecycle model* könnte beispielsweise die Entscheidung zwischen alternativen Varianten des *parallel design* (Phase 4) sinnvoller Weise auf Evaluationsergebnissen basieren. Prototypen (Phase 8) ermöglichen es, frühzeitig im Entwicklungsprozess Tests mit realen Nutzern durchführen zu können [vgl. Nielsen 1993, S. 93f.] und schaffen somit die Voraussetzung für das *empirical testing* (Phase 9).

Sarodnick und Brau betonen in ihrem Prozessmodell des Usability Engineerings noch deutlicher die zentrale Rolle der Usability Evaluation. Als „entwicklungsbegleitende Evaluation“ [Sarodnick 2006, S. 82] zieht sie sich durch alle vier unterschiedene Phasen des Modells – Analyse-, Konzept-, Entwicklungs- und Einführungsphase. Zwischen diesen gibt es „eine enge Rückkopplung, die durch die prozessbegleitende Evaluation gesteuert wird, d.h. in fast allen Schritten des Usability-Engineerings werden Evaluationsmaßnahmen durchgeführt“ [Sarodnick 2006, S. 86]. Die Evaluation wird damit – noch stärker als bei Nielsen – nicht als eine Phase des Usability Engineerings angesehen, sondern als integrierter und begleitender Prozess betrachtet.

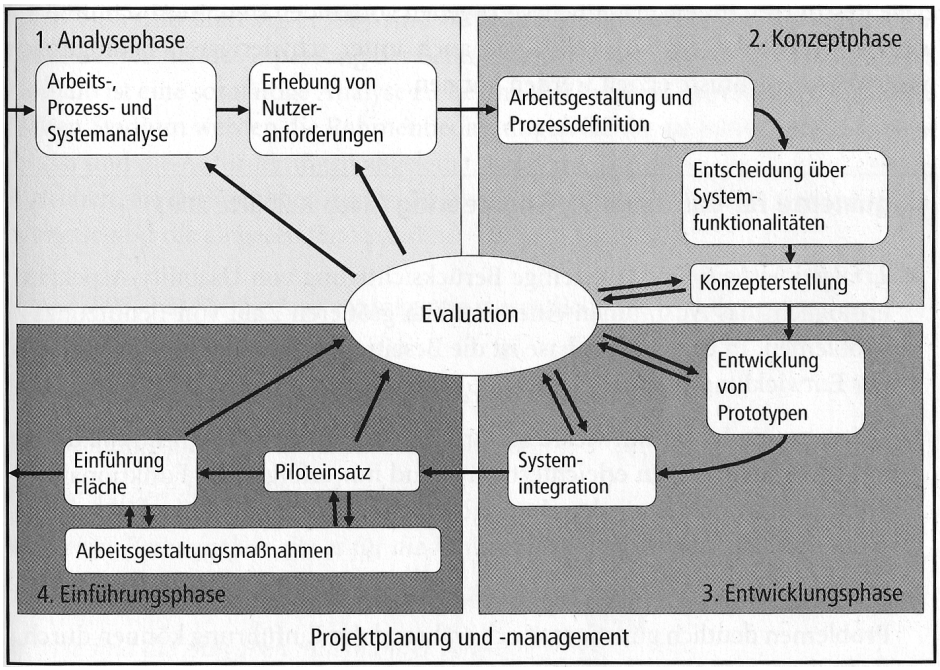

Abbildung 1: Prozessmodell "Usability-Engineering" von Sarodnick und Brau, entnommen aus [Sarodnick 2006, S. 85]

Beide Prozessmodelle weisen deutliche Übereinstimmungen in zentralen Aspekten auf, insbesondere hinsichtlich der Bedeutung umfassender Analyse, frühzeitiger Integration von Aktivitäten des Usability Engineerings in den Entwicklungsprozess und der Möglichkeit von Feedback-Schleifen auf Basis von Evaluationsergebnissen. Aufgrund des letztgenannten Aspekts ist für wirkungsvolles Usability Engineering die Einbettung in einen *iterativen* Entwicklungsprozesses Voraussetzung. Im Unterschied zum klassischen „Wasserfallmodell“, in dem definierte Phasen streng sequentiell durchlaufen werden, gestattet ein iteratives Vorgehen die Rückkehr zu früheren Entwicklungsphasen, um dort Anpassungen und Optimierungen vorzunehmen.

Für eine möglichst reibungsfreie Integration des Usability Engineering eignen sich hierbei in besonderer Weise *Prototyping* und *Spiralmodell*. Beim Prototyping werden die definierten Anforderungen zunächst in Prototypen umgesetzt und durch Evaluation abgesichert. Über Rückkopplungsschleifen werden die Anforderungen soweit nötig modifiziert. Auf diese Weise gelangt man zu einer verbindlichen Systemspezifikation, die im Folgenden Implementiert wird. Begleitend können weitere Evaluationen durchgeführt werden, deren Erkenntnisse in den Implementierungsprozess einfließen. Im Spiralmodell kommt ebenfalls ein Prototyp zum Einsatz. Jedoch werden hier die Phasen – Analyse, Design, Implementierung, Test – zyklisch immer wieder durchlaufen, so dass der Prototyp sich kontinuierlich dem Zielzustand annähert. Auf diese Weise schraubt das System sich quasi in einer Spiralbewegung zu immer höherer Komplexität empor [vgl. Sarodnick 2006, S. 98ff.].

3.5 METHODENÜBERBLICK

Die Usability Evaluation verfügt über eine Vielzahl etablierter Methoden, die nach verschiedenen Kriterien kategorisiert werden können. Von zentraler Bedeutung ist die Unterscheidung in *quantitative* und *qualitative* Methoden[23]. Quantitative Methoden haben die Gewinnung von Ergebnissen als Zahlenwerte zum Ziel, die Vergleiche ermöglichen und mittels statistischer Verfahren auswertbar sind. Hierzu zählen beispielsweise Messungen von Fehlerraten, Bearbeitungszeiten oder Erfolgsquoten. Auch subjektive Faktoren wie Zufriedenheit oder wahrgenommene Anstrengung können jedoch in Zahlenwerte überführt werden, indem sie beispielsweise über Rating-Skalen in Fragebögen erhoben werden [vgl. Dahm 2006, S. 319]. Da quantitative Methoden Daten durch die Beobachtung oder Befragung realer Nutzer gewinnen [vgl. Sarodnick 2006, S. 113], werden sie auch als *nutzerzentrierte* Methoden bezeichnet. Dem gegenüber stehen die *expertenzentrierten* oder *qualitativen* Methoden, bei denen eine Analyse und Beurteilung durch Usability-Experten erfolgt. Im Folgenden werden die wichtigsten Methoden und ihre Charakteristiken in kompakter Form dargestellt.

23 Alternativ werden auch die Bezeichnungen *empirische* und *analytische* Methoden verwendet [vgl. Sarodnick 2006, S. 113], wobei die quantitative der empirischen und die qualitative der analytischen Methode entspricht.

3.5.1 EXPERTENZENTRIERTE METHODEN

1. *Heuristische Evaluation* – Die bereits eingeführten Heuristiken werden hier zu Evaluationszwecken herangezogen. Usability-Experten explorieren das zu untersuchende System und überprüfen dabei systematisch die Einhaltung der in den Heuristiken formulierten Forderungen. Verstöße dienen als Hinweise auf mögliche Usability Probleme. Bei ihrer Auseinandersetzung mit dem System bemühen die Evaluatoren sich in der Regel darum, die Perspektive zukünftiger realer Nutzer einzunehmen. Um dies zu erreichen ist neben der Usability-Expertise häufig auch Fachwissen aus der jeweiligen Anwendungsdomäne sowie umfassende Kenntnisse über die Zielgruppe erforderlich [vgl. Sarodnick 2006, S. 135f.]. Ergebnis der heuristischen Evaluation ist eine Liste erkannter Probleme mit Hinweisen auf die Gestaltungsprinzipien, gegen die jeweils verstoßen wurde. Konkrete Lösungsvorschläge werden nicht unterbreitet, doch können aus den betroffenen Heuristiken häufig die erforderlichen Änderungen abgeleitet werden [vgl. Nielsen 1993, S. 159].

2. *Walkthrough-Verfahren* – Walkthrough Verfahren bilden gemeinsam mit der Heuristischen Evaluation die Gruppe der Inspektionsmethoden, deren Schwerpunkt auf „Voraussage von potenziellen Usability-Problemen (…) durch Usability-Experten“ [Sarodnick 2006, S. 134] liegt. Es gibt eine Reihe verschiedener Walkthrough-Verfahren, die unterschiedliche Fragestellungen verfolgen. Allen gemeinsam ist, dass einer Gruppe von Evaluatoren ein Entwurf für eine mögliche Gestaltung des Systems zur Bewertung anhand zuvor definierter Kriterien vorgelegt wird. Der Fokus des *Cognitive Walkthrough* liegt beispielsweise in der Überprüfung, wie einfach das Funktionsprinzip eines Systems sich unerfahrenen Nutzern erschließt. Hierzu wird von Usability-Experten eine ideale Handlungsabfolge entwickelt. Diese Abfolge wird anschließend von Experten evaluiert, um Punkte aufzuzeigen, an denen eine Abweichung der Nutzer vom idealen Lösungsweg erwartet werden muss. Für die erkannten Problematiken werden im Anschluss alternative Lösungen erarbeitet. Im Unterschied zu der umfassenden Heuristischen Evaluation fokussiert der Cognitive Walkthrough somit auf Aspekte der Erlernbarkeit [Sarodnick 2006, S. 144f.].

Der *Pluralistic Usability Walkthrough* stellt eine Variante dar, die auf Partizipation im Evaluationsprozess setzt. Neben Usability-Experten werden auch repräsentative Nutzer und Produktentwickler beteiligt. Allen Evaluatoren werden auch hier Entwürfe zu einem Handlungsweg im System zur Beurteilung vorgelegt. Nach individueller Bewertung jedes Teilschrittes erfolgt jeweils eine Diskussion in der Gruppe. Die Vorzüge dieser Methode bestehen in der Möglichkeit, sehr frühzeitig reale Nutzer in den Evaluationsprozess einzubinden und dem vergleichsweise geringen Aufwand der Durchführung. Ferner können in dieser Konstellation Defizite der Usability-Experten auf Seiten des spezifischen Domänenwissens kompensiert werden. Die direkte Interaktion mit realen Nutzern und ihren Fragen kann zudem hilfreiche Impulse für den weiteren Entwicklungsprozess liefern [vgl. Sarodnick 2006, S. 150ff. sowie Nielsen 1993, S. 162].

3. *Guidelines* – Ähnlich wie bei der heuristischen Evaluation stellen Guidelines Listen von Gestaltungsrichtlinien zur Verfügung, deren Einhaltung und Umsetzung von den Evaluatoren überprüft wird. Das Vorgehen ist hierbei meist frei und explorativ, nur Sonderfälle wie Expertenleitfaden und Checkliste machen strikte Vorgaben zur Vorgehensweise und erfordern ein systematisches Durcharbeiten. Auch die Form der Problemdokumentation ist meist freigestellt, eine sinnvolle Festlegung muss daher häufig im Rahmen des jeweiligen Evaluationsvorhabens getroffen werden. Die Durchführung dieser Methode ist vergleichsweise einfach und mit geringem Aufwand verbunden. Zugleich gibt es jedoch einige Kritik an diesem Ansatz, insbesondere in Hinblick auf Aussagekraft der Ergebnisse sowie Verlässlichkeit und Qualität vieler verfügbaren Guidelines [vgl. Sarodnick 2006, S. 117ff.].

Expertenzentrierte Methoden können grundsätzlich im gesamten Entwicklungsprozess eingesetzt werden, ein genereller Vorzug besteht jedoch in ihrer frühzeitigen Anwendbarkeit. Bereits erste Konzepte, Skizzen und Entwürfe können der kritischen Prüfung durch Experten unterzogen werden, noch bevor die Implementierung oder die Entwicklung von Prototypen begonnen wird. Erkenntnisse aus dem Evaluationsprozess stehen auf diese Weise bereits bei der Umsetzung von Prototypen zur Verfügung und dienen zu deren Optimierung. Die Ergebnisse expertenzentrierter

Methoden müssen zudem nicht aufwendig analysiert und interpretiert werden. Die Experten selbst liefern detaillierte Listen mit entdeckten Mängeln, die in der Regel auch bereits kategorisiert und hinsichtlich ihres Schweregrads bewertet sind. Durch das Aufzeigen verletzter Heuristiken werden ferner Ansätze für verbesserte Lösungen zur Verfügung gestellt. Mitunter, wie beispielsweise im Rahmen des Cognitive Walkthroughs, werden gar konkrete Lösungsempfehlungen von den Experten formuliert.

Diesen Vorzügen stehen jedoch auch Nachteile gegenüber. Nutzungskontext, Charakteristika realer Nutzer und die tatsächliche Form der Nutzerinteraktion mit dem betrachteten System finden in den expertenzentrierten Methoden keine bzw. nur geringe Beachtung. Relativ starre Vorgaben von Checklisten, Guidelines und Heuristiken erlauben nicht immer eine optimale Anpassung an die Bedürfnisse des konkreten Evaluationsvorhabens. Die Fähigkeit von Usability-Experten, sich in die Situation repräsentativer Nutzer zu versetzen und deren mentale Modelle nachzuvollziehen, muss ferner als eingeschränkt betrachtet werden. Vorhandene Fachkenntnis, Wissen um das konkrete System und umfangreiche Erfahrungen können kaum wirkungsvoll ausgeblendet werden. Eine Beeinflussung und somit Abweichung von repräsentativen Nutzern ist somit zu erwarten.

Expertenzentrierte Methoden werfen zudem Fragen hinsichtlich interner Validität und Reliabilität auf. Interne Validität bezeichnet hierbei allgemein das Ausmaß der Eindeutigkeit, mit der eine Veränderung der abhängigen Variable auf Variationen der unabhängigen Variable zurückgeführt werden kann und Alternativerklärungen somit unwahrscheinlich werden. Übertragen auf expertenzentrierte Methoden kann interne Validität als das Ausmaß der Unabhängigkeit des Evaluationsergebnisses vom jeweiligen Evaluator verstanden werden. Sie ist bei expertenzentrierten Methoden generell gering ausgeprägt, dies bedeutet: „unterschiedliche Evaluatoren decken mit derselben Methode teilweise deutlich abweichende Usability-Probleme im selben System auf und bewerten diese unterschiedlich“ [Sarodnick 2006, S. 185]. Daraus folgt, dass die Art und Anzahl der entdeckten Usability-Probleme – und somit die Güte der Evaluationsergebnisse und die Leistungsfähigkeit der Methode – in starkem Maße von den beteiligten Evaluatoren abhängig sind.

Aus geringer interner Validität folgt zugleich eine geringe Reliabilität. Als Reliabilität wird das Ausmaß der Übereinstimmung der Ergebnisse wiederholter Evaluationen desselben Systems mit derselben Methode bezeichnet [vgl. Sarodnick 2006, S. 186]. Aufgrund der Abhängigkeit der Ergebnisse von den Evaluatoren sind bei geringer interner Validität bei neuen Evaluatoren abweichende Ergebnisse zu erwarten, auch wenn System und Methode konstant gehalten werden. Methoden mit Gruppenevaluation und Konsensbildung zwischen den Evaluatoren vermindern die Problematik geringer interner Validität naturgemäß, da interindividuelle Unterschiede zwischen den Evaluatoren tendenziell ausgeglichen werden. Bei Methoden ohne dieses ausgleichende Element, insbesondere der Heuristischen Evaluation, kann der geringen internen Validität mit einer größeren Anzahl von Evaluatoren begegnet werden. Auf diese Weise wird die Tatsache, dass unterschiedliche Experten unterschiedliche Usability-Probleme entdecken, sogar zum Vorteil. In der aggregierten Summe der Entdeckungen wird ein breites Spektrum von Usability-Problemen diagnostiziert und interindividuelle Unterschiede relativieren sich bzw. werden ausgeglichen. Nielsen betont: „it is possible to achieve substantially better performance by aggregating the evaluations from several evaluators“ [Nielsen 1993, S. 156]. Wichtig hierbei ist, dass die Evaluationen zunächst individuell erfolgen, um gegenseitige Beeinflussung zu vermeiden. Die Ergebnisse werden erst im nächsten Schritt zusammengeführt.

Zusammenfassend lässt sich festhalten, dass die Vorteile expertenzentrierter Methoden – frühe Anwendbarkeit, geringer Aufwand, qualitatives Feedback inklusive Hinweise auf Ursachen von Usability-Problemen und deren mögliche Behebung – sie zu einem unverzichtbaren Bestandteil des Evaluationsprozesses machen. Insbesondere das frühzeitig verfügbare Feedback ermöglicht Korrekturen bereits vor der Entwicklungsphase, die besonders einfach, effizient und kostengünstig umsetzbar sind. Zudem profitieren alle weiteren Entwicklungs- und Evaluationsschritte von den auf diese Weise bereits erreichten Optimierungen. Da die Nutzerperspektive in diesen Methoden jedoch deutlich unterrepräsentiert ist, sollte sich das Usability-Engineering keinesfalls allein auf die Ergebnisse der expertenzentrierten Methoden stützen. Sobald ausreichende Grundlagen geschaffen sind (z.B. verwendbarer Prototyp) sollten vielmehr umgehend nutzerzentrierte Methoden zur Anwendung kommen, um die bereits vorliegenden Erkenntnisse abzusichern und Erfahrungen mit konkreten Vertretern der Nutzergruppe zu sammeln, da letztlich „der Nutzerkontext und die Ziel-

gruppe über die tatsächliche Qualität der Schnittstelle entscheiden (...)“ [Sarodnick 2006, S. 123].

3.5.2 NUTZERZENTRIERTE METHODEN

1. *Fragebögen* – Fragebögen bestehen aus einer Anzahl von Fragen und/oder Aussagen, die als *Items* bezeichnet werden. Die Teilnehmern der Evaluationsstudie rekrutieren sich idealer Weise aus der intendierten Zielgruppe und stellen somit repräsentative Nutzer dar. Eine Fragebogenstudie erfasst grundsätzlich die subjektive Beurteilung der Nutzer. Die Subjektivität der Aussagen relativiert sich jedoch tendenziell bei der Auswertung, sofern entsprechende Teilnehmerzahlen erreicht werden. Nach einer Auseinandersetzung mit dem zu evaluierenden System bewerten die Nutzer das System auf Grundlage der dargebotenen Items. In stark standardisierten Fragebögen ist die Formulierung der Items, die Reihenfolge der Bearbeitung und das Format der Antwort festgelegt. Es werden ausschließlich numerische Daten gewonnen, primär über Ratingskalen, die vorgegebenen Antworten Zahlenwerte zuordnen und damit beispielsweise die Ausprägung von Merkmalen abbilden[24]. Solch standardisierte Fragebögen stellen ein wichtiges Instrument zur Erhebung quantitativer Daten dar, die eine statistische Auswertung gestatten und den systematischen Vergleich zwischen Systemen oder Systemvarianten bzw. -versionen ermöglichen. Sie variieren hierbei hinsichtlich Art und Anzahl der gemessenen Konstrukte sowie dem Fokus der Betrachtung, der beispielsweise auf den Aspekten System, Interface oder Aufgabe liegt. Auch Kombinationen mehrerer Schwerpunkte in einem Fragebogen sind möglich [vgl. Vuolle 2008, S. 2] Weniger standardisierte Fragebögen gestatten zusätzliche Antwortmöglichkeiten wie Multiple-Choice oder auch Freitexteingaben und ermöglichen somit auch die Erfassung qualitativer Daten. Eine statistische

24 Ratingskalen ermöglichen eine differenzierte Beurteilung eines Merkmals bei gleichzeitiger Übersetzung in numerische Werte. Häufig werden beispielsweise fünf bis sieben Abstufungen angeboten, aus denen der Teilnehmer den am zutreffendsten erscheinenden Wert auswählt. Eine Bewertung findet hierbei meist zwischen zwei Polen statt (z.B. verwirrend – klar, stimme absolut zu – stimme überhaupt nicht zu, ist mir sehr wichtig – ist mir sehr unwichtig, etc.). Grundsätzlich sind auch Skalen mit geringerer Auflösung möglich, die beispielsweise nur drei Abstufungen (Zustimmung, Ablehnung, unentschlossen) gestatten [vgl. Sarodnick 2006, S. 175ff.].

Auswertung dieser Daten ist naturgemäß nicht möglich. Besonders frei formulierte Texte erfordern zudem einen hohen Aufwand bei der Auswertung.

Die Konstruktion eines Fragebogens, der hohen Ansprüchen an Validität. Reliabilität und Objektivität genügt, ist aufwendig und bedarf spezieller Fachkenntnis. Da in Entwicklungsprojekten selten die Ressourcen vorhanden sind, maßgeschneiderte Fragebögen für das eigene Evaluationsvorhaben zu erstellen, wird in der Praxis häufig auf standardisierte und wissenschaftlich abgesicherte Fragebögen zurückgegriffen, die für unterschiedliche Systeme und Fragestellungen öffentlich verfügbar sind[25] [Sarodnick 2006, S. 170].

Fragebögen erheben primär das subjektive Erleben der Nutzer, Eindrücke und Gefühle während der Interaktion und sind somit geeignet, die subjektive Zufriedenheit der Nutzer zu erfassen. Damit schließen sie eine Lücke, da sich die Zufriedenheit der Nutzer mit einem System weder aus Beurteilungen der Experten noch aus der Messung von Performance-Daten sicher ableiten lässt. Dies wiegt umso schwerer, als Zufriedenheit neben Effektivität und Effizienz eines der drei zentralen Ziele der Usability darstellt. Hohe Effektivität und Effizienz eines Systems können generell als gute Indikatoren für die Nutzerzufriedenheit gelten, doch ist der Kausalzusammenhang nicht immer zwingend und eindeutig. Nielsen stellt fest: „Users often do not know what is good for them“ [Nielsen 1993, S. 11]. An gleicher Stelle schildert er das Beispiel einer Studie, in der Telefonhörer unterschiedlichen Gewichts miteinander verglichen wurden. Fragte man die Testnutzer, ob sie sich leichtere Hörer wünschten, so erhielt man überwiegend eine ablehnende Antwort, die Nutzer waren glücklich und zufrieden mit den Geräten, die sie hatten. Bei einem Test mit identisch aussehenden Hörern unterschiedlichen Gewichts präferierten die meisten Nutzer entgegen ihrer Aussage jedoch ein Modell, das um die Hälfte leichter als das vorhandene Gerät war. In diesem Fall äußerten die Nutzer somit ihre Zufriedenheit mit einer objektiv nachteiligen Lösung. Ebenso ist denkbar, dass ein objektiv verbessertes System (beispielsweise hinsichtlich Fehlerrate, erforderlicher Bearbeitungszeit, etc.) hinsichtlich der Nutzerzufriedenheit schlechter

25 Beispiele sind der *Questionnaire for User Interface Satisfaction* (QUIS), das *Software Usability Measurement Inventory* (SUMI) oder *AttrakDiff*. Eine kompakte Darstellung dieser und weiterer Fragebögen finden sich bei Sarodnick und Brau [Sarodnick 2006, S. 175 – 181].

bewertet wird. Bei der Bewertung der Nutzer spielen neben objektiven Kriterien nicht zuletzt Gewohnheit und Erwartungshaltung eine Rolle. Fragebögen können hier dazu beitragen, Diskrepanzen zwischen Systemleistung und subjektiver Beurteilung aufzudecken.

2. *Usability Test* – Usability Tests, auch als Nutzertests bezeichnet, werden ebenfalls mit repräsentativen Nutzern durchgeführt. Anders als bei Fragebögen stehen jedoch nicht subjektive Bewertungen im Vordergrund. Statt dessen sollen durch Beobachtung von Nutzern in Interaktion mit dem zu testenden System all jene Schwachstellen aufgezeigt werden, die zu Problemen oder Irritationen führen. Vielen Usability-Experten gilt der Nutzertest als wichtigste Evaluationsmethode, da nur hier Daten über reale Nutzer bei der Bewältigung konkreter, realitätsnaher Aufgabenstellungen gewonnen werden. Nielsen hält hierzu fest: „User testing with real users is the most fundamental usability method and is in some sense irreplaceable, since it provides direct information about how people use computers and what their exact problems are with the concrete interface being tested“ [Nielsen 1993, S. 165]. Um dieses Potential auszuschöpfen bedarf es einer sorgfältigen Auswahl der Testpersonen, damit diese repräsentative Vertreter der Produktzielgruppe darstellen. Lässt sich die Zielgruppe nicht sinnvoll eingrenzen, so sollte eine heterogene Zusammensetzung gewählt werden, die der Gesamtpopulation bzw. relevanten Teilpopulationen entspricht. Hierzu sind insbesondere die Faktoren Alter, Geschlecht, Ausbildung, Beruf sowie Vorerfahrung mit der entsprechenden Produktgattung zu berücksichtigen. Ferner sollten die Probanden idealer Weise keine Erfahrung mit dem zu testenden System haben [vgl. Sarodnick 2006, S. 159f.]. Die Auswahl der Testpersonen hat direkte Auswirkungen auf die Validität der Ergebnisse. Geeignete Probanden sind somit eine wichtige Voraussetzung für aussagekräftige Usability Tests [vgl. Nielsen 1993, S. 169].

   Neben der Entscheidungen hinsichtlich der Testnutzer sind für einen Usability Test weitere umfangreiche Vorbereitungen erforderlich. Es werden zunächst die grundsätzlichen Ziele definiert und die Form der Umsetzung entschieden. Ferner müssen beispielsweise repräsentative Aufgaben erstellt werden, die von den Probenden zu bearbeiten sind, es muss geklärt werden, unter wel-

chen Bedingungen eine Aufgabe als erfolgreich abgeschlossen gelten soll oder inwieweit ein Experimentalleiter bei Bedarf Hilfestellung geben darf [vgl. Nielsen 1993, S. 170f.].

Der Test selbst besteht in der Beobachtung der Probanden, während diese vorgegebene Aufgabenstellungen am zu evaluierenden System[26] (bzw. mit einem für die Aufgabenbewältigung ausreichenden Prototypen) bearbeiten. Die Aufgaben sollen so gewählt werden, dass sie möglichst repräsentativ für die tatsächliche Nutzung des fertigen Systems sind und eine sinnvolle Bandbreite der wichtigsten Interface-Komponenten nutzen [vgl. ebenda, S. 185]. Die Beobachtung findet in der Regel in einem Labor statt, da die Rahmenbedingungen dort konstant gehalten werden können und Störvariablen (z.B. Lärm, Ablenkung, schwankende Systemperformance, etc.) dadurch weitgehend eliminiert werden. Labore besitzen neben dem Testraum, in dem der Proband arbeitet, häufig auch einen Kontrollraum, von dem aus ein Evaluator den Probanden per Monitor oder Einwegspiegel im Blick hat [vgl. Sarodnick 2006, S. 160f.]. Dem Probanden steht meist ein weiterer Evaluator zur Seite, der durch den Test führt und auf die Einhaltung der Testbedingungen achtet. Abbildung 3 bietet eine exemplarische Darstellung eines Labors.

26 Der Standardfall dieser Methode ist die Bearbeitung von Aufgaben an einem Desktop-PC, meist zur Evaluierung von Software oder Webseiten.

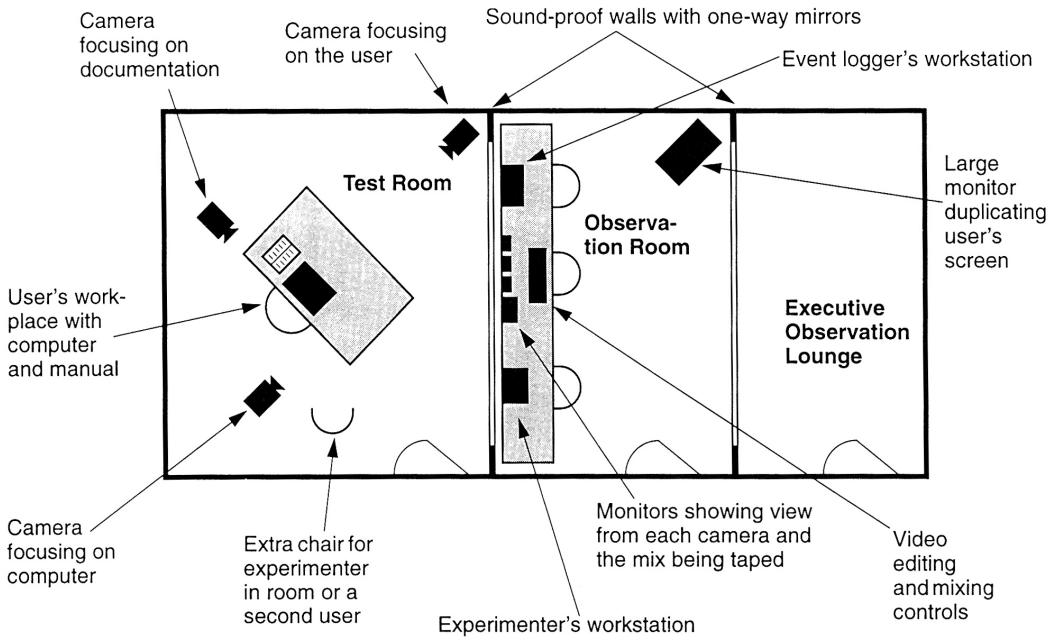

Abbildung 2: Schematische Darstellung eines hypothetischen, jedoch typischen, Usability-Labors nach Nielsen, entnommen aus [Nielsen 1993, S. 201]

Für eine spätere ausführliche Analyse wird üblicher Weise der gesamte Testverlauf auf Video aufgezeichnet, meist aus mehreren Blickwinkeln. Insbesondere der Gesichtsausdruck des Probanden sowie ggf. die Arme und Hände während der Interaktion sind von besonderem Interesse. Ergänzend werden meist mittels spezieller Hard- oder Software Aufzeichnungen vom Bildschirminhalt des Testsystems inklusive aller Eingaben und Mausbewegungen angefertigt. Auch Audioaufzeichnungen kommen häufig zum Einsatz. Häufig werden ferner wichtige Einblicke in Denkweise und mentale Modelle der Nutzer gewonnen, indem man den Nutzer während der Aufgabenbearbeitung seine Handlungen kommentieren und seine Gedanken verbalisieren lässt (Lautes Denken oder *Thinking Aloud*). Nielsen bezeichnet dieses Vorgehen als „single most valuable usability engineering method“ und führt dazu aus: „By verbalizing their thoughts, the test users enable us to understand how they view the computer system, and this again makes it easy to identify the users' major misconceptions (…), because the thinking-aloud method shows how users interpret each individual interface item“ [Nielsen 1993, S. 195]. Eine weitere Informationsquelle im Usability Test bietet die Messung von Performance-Daten [ebenda, S. 193] (z.B. Bearbeitungsdauer, Fehlerzahl, Abbrüche, etc.).

Hinsichtlich der Gütekriterien sind in diesem Kontext insbesondere Reliabilität und externe Validität genauer zu betrachten. Die Reliabilität von Usability Tests ist hierbei nicht ganz unproblematisch, da erhebliche interindividuelle Unterschiede zwischen Probanden bestehen: „It is not uncommon, to find that the best user is 10 times as fast as the slowest user“ [Nielsen 1993, S. 166]. Entsprechend hoch sind die auftretenden Standardabweichungen. Auf Grundlage einer vergleichenden Studie gibt Nielsen die mittlere Standardabweichung mit 33% für expert-user und gar 46% für novice-user an. Insbesondere bei Vergleichen zwischen Systemen bzw. Varianten sollte auf eine hohe Reliabilität geachtet werden, da unterschiedliche Ergebnisse andernfalls womöglich nicht auf tatsächliche Unterschiede in den Systemen zurückzuführen sind sondern zufällige Schwankungen darstellen. Grundsätzlich gilt, dass die Reliabilität der Ergebnisse mit zunehmender Teilnehmerzahl steigt, bei wenigen Teilnehmern müssen hingegen große Konfidenzintervalle in Kauf genommen werden [vgl. Nielsen 1993, S. 167ff.]. Im Rahmen einer Studie wurde für unterschiedliche Gruppengrößen untersucht, welchen Anteil der im System bekannten Usability-Probleme die verschiedenen Gruppen jeweils fanden. Kleine Gruppen von nur fünf Personen weichen hierbei hinsichtlich des jeweils gefundenen Anteils erwartungsgemäß stark voneinander ab, was sich in der Bandbreite zwischen minimal und durchschnittlich aufgedeckten Usability-Problemen niederschlägt. Mit zunehmender Gruppengröße verringert sich diese Diskrepanz zwischen Minimum und Durchschnitt jedoch zusehends. In der Analyse wurden folgende Zusammenhänge ermittelt: bei 5 Nutzern lag das Spektrum zwischen Minimum und Durchschnitt bei 55% vs. 85,55%, bei 10 Nutzern nahm es ab auf 82% vs. 94,69%, bei 20 Nutzern beträgt es noch 95% vs. 98,4% und bei 40 Nutzern sind mit 98% vs. 99,6% kaum noch Unterschiede vorhanden [Faulkner 2003, S. 381]. Faulkner hebt die Bedeutung dieser Zusammenhänge für die Praxis hervor. Obgleich bei einer Gruppengröße von fünf Personen auf Grundlage ihrer Daten durchschnittlich 85% der Probleme erkannt werden, gilt: „any given set of 5 users may reveal only 55% of the usability problems“ [ebenda].

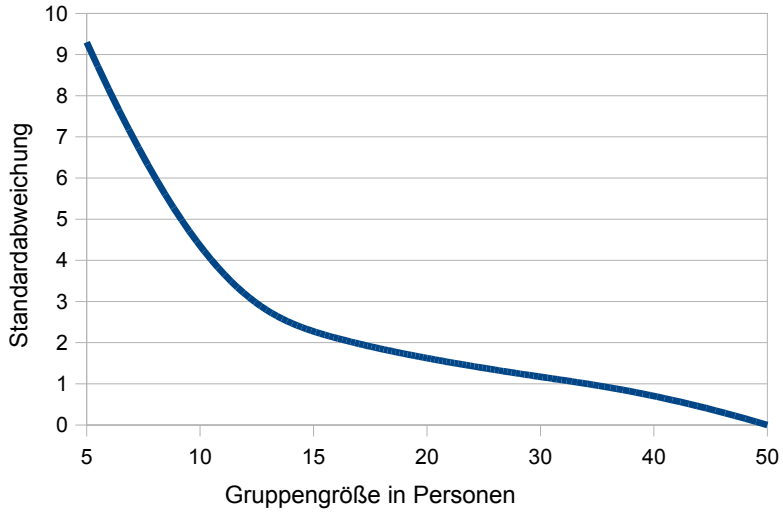

Abbildung 3: Standardabweichung in Abhängigkeit von der Gruppengröße, Datenreihe nach Faulkner [Faulkner 2003, S. 381]

In Abbildung 3 sind ist die Standardabweichung in Abhängigkeit von der Gruppengröße visualisiert. Da die Kurve abflacht, nimmt der qualitative Zugewinn durch das Hinzufügen weiterer Testnutzer mit jedem neuen Probanden ab. Aufgrund dieser Tatsache und der meist begrenzten Ressourcen werden Nutzertests häufig mit eher kleinen Gruppen durchgeführt.

Als zweites Gütekriterium dieser Evaluationsmethode soll die externe Validität betrachtet werden. Diese ist dann gegeben, wenn die ermittelten Ergebnisse über die im Labor getestete Stichprobe hinaus verallgemeinerbar sind und somit auch Aussagen und Prognosen hinsichtlich realer Personen in realen Kontexten gestatten. Es lässt sich somit sagen: „Validity is a question of whether the usability test in fact measures something of relevance to usability of real products in real use outside the laboratory“ [Nielsen 1993, S. 169]. Dieses Kriterium lässt sich nicht durch statistische Verfahren oder eine Erweiterung der Stichprobengröße optimieren. Von zentraler Bedeutung für die Validität sind neben der erwähnten Auswahl repräsentativer Testnutzer die Konstruktion angemessener Aufgaben und die Herbeiführung realistischer Testumstände (z.B. zeitliche Beschränkungen, soziale Einflüsse, etc.) [ebenda]. Entscheidend ist insgesamt, dass die Testsituation der realen Nutzungssituation in allen relevanten Aspekten möglichst vergleichbar gestaltet ist [vgl. Sarodnick 2006, S. 157].

Nutzerzentrierte Methoden bedeuten in der Umsetzung meist deutlich höheren Aufwand als expertenzentrierte Methoden. Dies gilt bereits, da die sorgfältige Auswahl und die Rekrutierung geeigneter Testpersonen sich häufig zeit- und ggf. auch kostenintensiv gestaltet [vgl. Nielsen 1993, S. 175f.]. Insbesondere bei Usability Tests kommen umfangreiche Vorbereitungen hinzu, die beispielsweise die Erstellung eines Testplans, die Instruktion der Experimentalleiter und die Erarbeitung repräsentativer Aufgabenstellungen für die Testnutzer umfassen. Die aus nutzerzentrierten Methoden resultierenden Daten müssen im Unterschied zu den expertenzentrierten Methoden ferner aufwendig analysiert, interpretiert und ausgewertet werden. Dennoch stellen nutzerzentrierte Methoden eine unverzichtbare Komponente des Evaluationsprozesses dar. Nur sie berücksichtigen die Nutzerperspektive aus erster Hand und stellen belastbare Daten über subjektive Zufriedenheit bzw. konkrete Usability-Probleme realer Nutzer zur Verfügung. Rückmeldungen von Nutzern sollten im Usability Engineering zudem generell besondere Beachtung finden, da der gesamte Prozess letztlich darauf ausgerichtet ist, maximale Effektivität, Effizienz und Zufriedenheit *für die Nutzer* zu erreichen.

Nutzerzentrierte Methoden können grundsätzlich ebenfalls in allen Entwicklungsphasen durchgeführt werden – als entwicklungsbegleitende *formative* Evaluation, die primär Schwachstellen aufdecken und Optimierungsmöglichkeiten aufzeigen soll ebenso wie als *summative* Evaluation, die Systeme in ihrer Leistungsfähigkeit beurteilt und somit Vergleiche zwischen Systemen oder Erfolgskontrolle von Optimierungsmaßnahmen ermöglicht. Generell erfordern nutzerzentrierte Methoden jedoch weiter fortgeschrittene Entwicklungsergebnisse als die expertenzentrierten Methoden und können somit erst später im Entwicklungsverlauf zur Anwendung kommen. Dies gilt insbesondere im Fall des Usability Tests, für den typischer Weise ein (für die Testaufgaben ausreichend) funktionaler Prototyp erforderlich ist.

In der Praxis empfiehlt sich eine durchdachte Kombination verschiedener Methoden, die zum jeweils geeigneten Zeitpunkt über den gesamten Entwicklungsverlauf hinweg zum Einsatz kommen. Mitunter lassen sich Methoden ferner sinnvoll kombinieren. Da sowohl Nutzertests als auch Fragebögen eine Interaktion von Testnutzern mit einem System erfordern, bietet es sich beispielsweise an, Probanden im Anschluss an den Test Fragebögen ausfüllen zu lassen oder *Interviews* zu führen, um ergän-

zend subjektive Beurteilungen zu erheben. Ferner gibt es neben den vorgestellten Methoden eine Reihe weiterer Techniken, die im Evaluationsprozess zur Anwendung gebracht werden können. Da auch eine kompakte Übersicht verfügbarer Methoden den Rahmen dieser Arbeit sprengen würden, sei an dieser Stelle nur beispielhaft auf *Formal-analytische Verfahren* (vgl. GOMS), *Feldstudien* oder *Fokusgruppen* verwiesen.

4 BESONDERHEITEN MOBILER ENDGERÄTE

> *„Established concepts, methodologies, and approaches in human–computer interaction are being challenged by the increasing focus on systems for wearable, handheld, and mobile computing devices. This move beyond office, home, and other stationary use settings has created a need for new approaches to design and evaluate useful and usable systems“* [Kjeldskov 2004, S. 600]

Die grundlegenden Prinzipien der Usability haben einen weiten Geltungsanspruch, die Verwirklichung der zentralen Ziele – Effektivität, Effizienz und Zufriedenheit – wird unabhängig von der Gerätegattung und der verwendeten Dialogtechnik angestrebt. Mit jeder Stufe der Konkretisierung in der Anwendung dieser Prinzipien muss jedoch eine gewisse Einschränkung und Spezialisierung einhergehen, die den jeweiligen besonderen Charakteristiken der betroffenen Systeme Rechnung trägt.

Die ISO Norm 9241 wurde beispielsweise ursprünglich für den Kontext von Bürotätigkeiten mit Bildschirmgeräten entwickelt. Auch nach Erweiterung des Anwendungsbereich und Umbenennung in der Norm in „Ergonomie der Mensch-System-Interaktion“ bleibt diese Herkunft ersichtlich. Trotz generell breiter Anwendbarkeit der in Teil 110 definierten Grundsätze der Dialoggestaltung, erscheinen diese für die Entwicklung computerbasierter Systeme oder Anwendungen angemessener. Dies liegt nicht zuletzt an der implizit vorausgesetzten Komplexität und Informationsbandbreite der Dialogschnittstelle. Forderungen nach umfassender Selbstbeschreibungsfähigkeit, Individualisierbarkeit oder Lernförderlichkeit des Dialoges stoßen bei den technischen Möglichkeiten von Waschmaschinen, Mikrowellen und Videorekordern beispielsweise rasch an Grenzen.

Noch deutlicher wird die Spezialisierung auf der Ebene von Guidelines, Kriterienkataglogen und speziellen Heuristiken, die detaillierte Vorgaben für konkrete Anwendungsfelder enthalten (z.B. Webdesign bin hin zu spezifischen Teilbereichen wie e-Commerce-Webseiten). Dies hat auch Auswirkungen auf die Usability Evaluation. In direkter Weise durch Verwendung entsprechender Guidelines und Heuristiken zu Evaluationszwecken, allgemeiner in der Tatsache, dass bestimmte Evaluationsme-

thoden im Kontext konkreter Nutzungsszenarien entwickelt wurden. Spezifische Eigenschaften von Gerätegattungen sind somit aus Usability-Sicht zu berücksichtigen. Dies gilt im Besonderen, wenn diese Eigenschaften Auswirkungen auf die Validität von Evaluationsergebnissen erwarten lassen.

4.1 TECHNISCHE BESONDERHEITEN

Die wichtigsten technischen Eigenschaften mobiler Endgeräte sind auf den ersten Blick zu erfassen: zentral ist eine kompakte Bauweise, mit der zugleich Einschränkungen hinsichtlich der Datenausgabe (Displaygröße) und Dateneingabe (Größe und Umfang der Tastenfläche) einhergehen. Auch die verwendeten Prozessoren orientiert sich primär an den Kriterien Minimierung von Abmessung und Energieaufnahme und bleiben daher hinsichtlich der Leistungsfähigkeit weit hinter Desktop-Systemen zurück. Die aus diesen Beschränkungen resultierenden Herausforderungen an die Usability dieser Gerätegattung sind jedoch an anderer Stelle umfassend thematisiert und nicht Gegenstand der weiteren Betrachtung.

Weitere Merkmale sind die hochintegrierte Bauweise, die kaum austauschbare Komponenten besitzt und somit keine oder nur sehr geringe Erweiterungsmöglichkeiten bietet. Eine Anpassung von Funktionsumfang und Leistungsfähigkeit der Hardware an den individuellen Bedarf kann somit quasi nur über die Auswahl eines geeigneten Geräts erfolgen. Selbiges gilt für das Betriebssystem, da dieses nicht wie im Desktop-Bereich gegen Alternativen ausgetauscht werden kann. Dieser Faktor wiegt umso schwerer, als im mobilen Bereich kein Quasi-Standard wie das Betriebssystem Windows existiert. Statt dessen gibt es eine große Vielzahl überwiegend proprietärer Systeme der jeweiligen Hersteller, die sich in Struktur und Bedienweise häufig erheblich voneinander unterscheiden. Selbst verschiedene Modellreihen eines Herstellers weichen mitunter hinsichtlich der Bedienung deutlich voneinander ab. Hinzu kommt, dass im Bereich mobiler Geräte vielfältige grundsätzlich verschiedene Interaktionsformen nebeneinander bestehen, z.B: Nummernfeld-Tastatur, QWERTZ-Tastatur, Touchscreen mit Bedienung per Stylus oder per Finger. Während Desktops über klare Standards und Konventionen verfügen (z.B. WIMP-Paradigma, standardisiertes

Tastaturlayout, etc.), fehlen diese verbindlichen Orientierungspunkte im Sektor der Mobilgeräte bisher. Die resultierende Vielfalt kann Nutzer leicht verwirren, bedeutet für Software-Entwickler zusätzlichen Portierungs- oder Entwicklungsaufwand, um alle relevanten Plattformen und Systeme abzudecken und erschwert nicht zuletzt den objektiven Vergleich zwischen Geräten.

Eine weiterer Aspekt dieser Vielfalt betrifft die Usability direkt – und zwar auf Seiten der Inhalte, insbesondere der Web-Usability. Während die Evaluation einer Webseite unabhängig vom konkreten Desktop-System erfolgen kann – die Webseite also auf jedem PC, der gewisse Mindestanforderungen erfüllt – in vergleichbarer Weise genutzt werden kann, gilt dies nicht für den Bereich mobiler Geräte. Dort kommt es zu einer Vermischung der Einflüsse von Hardware und Content, so dass eine isolierte Betrachtung erschwert wird. Nielsen stellt in seiner Alertbox vom 20. Juli 2009 eine Usability-Studie zur Nutzung von Webseiten mittels Mobiltelefon vor. Hierbei erweisen sich erwartungsgemäß sowohl Eigenschaften der getesteten Webseite als auch der zum Testen verwendeten Hardware als starke Einflussfaktoren. Die Erfolgsquote auf Seiten, die für mobile Nutzung konzipiert sind, lag mit 64% deutlich höher als bei der mobilen Nutzung regulärer Seiten (53%). Betrachtet man die Erfolgsquote in Abhängigkeit von der Hardware, so finden sich noch stärkere Abweichungen. Für die hier unterschiedenen Geräte *Feature phone* (kleines Display, numerische Tastatur), *Smartphone* (mittleres Display, QWERTZ-Tastatur) und *Touch phone* (vollflächiges Display, fortschrittliches GUI) ergaben sich folgende Erfolgsquoten [vgl. Nielsen 2009]:

Feature phones	38%
Smartphones	55%
Touch phones	75%

Die Verquickung relevanter Faktoren bei mobilen Endgeräte wird auch von Sarodnick und Brau angeführt: es „vermischen sich in der Evaluation die Probleme des Gerätes, der Software und auch des Providers“ [Sarodnick 2006, S. 29].

Die Gerätevielfalt hat noch eine weitere Implikation von potentieller Relevanz für die Usability Evaluation. So reicht es für mobile Endgeräte eventuell nicht aus, konkrete Vorerfahrungen mit dem zu testenden Gerät zu erfassen. Aufgrund der Vielzahl ver-

fügbarer Systeme mit abweichenden Bedienkonzepten und des zu erwartenden hohen Maßes von Gewohnheit und Automatisierung aufgrund der intensiven Nutzung scheint es geboten, auch die Nutzung alternativer Systeme zu kontrollieren. Andernfalls können auftretende Probleme im Nutzertest womöglich nicht eindeutig auf das getestete Gerät zurückgeführt werden. Als alternative Erklärung – und somit konfundierende Variable – kämen Konflikte zwischen den Gewohnheiten der Probanden und der Funktionsweise des getesteten Systems in Betracht.

4.2 BESONDERHEITEN DER ZIELGRUPPE

Die Zielgruppe der Anwender mobiler Endgeräte ist nicht sinnvoll einzugrenzen, insbesondere die Nutzer von Mobiltelefonen decken das gesamte Spektrum in jeder Dimension ab (Alter, Geschlecht, Ausbildung, Vorerfahrung, allgemeines technisches Verständnis, etc.). Diese Tatsache gilt grundsätzlich für viele Consumer-Produkte, doch im Vergleich mit MP3-Playern oder Digitalkameras sind moderne Mobiltelefone außergewöhnlich komplexe Geräte, die in kompakter Form eine große Vielzahl von Funktionalitäten zur Verfügung stellen. Entsprechend komplex fällt die Benutzerschnittstelle aus, die dem Funktionsumfang und der Fülle von Optionen Rechnung trägt, was Aspekten der Usability wiederum besonderes Gewicht verleiht. Dies gilt umso mehr, als die Nutzungsintensität von Mobiltelefonen die anderer Consumer-Produkte übertrifft. Das Mobiltelefon ist in der Regel immer an, wird stets mitgeführt und von vielen Nutzern mehrfach täglich verwendet. Für die Usability Evaluation folgt aus der breiten und heterogenen Zielgruppe ein höherer Bedarf an Probanden, um repräsentative Vertreter aller relevanten Personengruppen berücksichtigen zu können.

4.3 BESONDERHEITEN DER NUTZUNGSWEISE

Sowohl Desktop-Rechner als auch mobile Endgeräte dienen grundsätzlich dem Informationsmanagement und der Kommunikation [vgl. Weiss 2002, S. 6]. Dennoch hebt

Weiss hervor, dass beide Gerätegattungen verschiedene Funktionen erfüllen und sich hinsichtlich der Nutzungsweise entsprechend unterscheiden. Desktop-Rechner werden seiner Position folgend genutzt, um umfangreiche Datenmengen zu erstellen und/oder umfassend zu bearbeiten (Textdokumente, komplexe Tabellen oder Grafiken, etc.). Mobile Geräte dienen hingegen eher der raschen Anzeige und ggf. eingeschränkten Modifikationen bestehender Dokumente oder der Erfassung kompakter Datensätze (z.B. Adressen, Kontaktdaten, etc.) unterwegs. Hieraus lässt sich ableiten, dass die Nutzung von Desktop-Systemen zeitlich ausgedehnt erfolgt, während mobile Geräte eher kurz, episodisch und spontan nach Bedarf genutzt werden. In ähnlicher Weise unterscheidet sich auch die Nutzung von Webinhalten, die Weiss prägnant als als „Surf vs. hunt“ [Weiss 2002, S. 16] charakterisiert. Während am Desktop-PC ausgedehntes Surfen praktiziert wird, das häufig auch einen Unterhaltungsaspekt beinhaltet, besteht die mobile Nutzung eher in dem gezielten Aufsuchen gerade benötigter Information.

4.4 NUTZUNGSKONTEXT

Die zentrale Besonderheit mobiler Endgeräte im Rahmen dieser Arbeit bildet der Kontext, in dem die Nutzung dieser Geräte erfolgt. Dieser Nutzungskontext unterscheidet sich in mehrerer Hinsicht grundsätzlich von demjenigen stationärer Desktop-Systeme. Bevor eine Darstellung dieser Unterschiede und ihrer Implikationen für die Usability erfolgen kann, bedarf es zunächst einer näheren Betrachtung des Begriffs „Kontext“.

Eine allgemeinverbindliche Definition von Kontext im Rahmen der Usability ist nicht existent. Entsprechend fehlt auch eine universelle Festlegung der Faktoren, die einen Kontext begründen oder ihm zu eigen sind, wie Bristow et al. beklagen: „Unfortunately, there remains a distinct lack of agreement as to what constitutes ‘context’“ [Bristow 2004, S. 798]. In erster Näherung lässt sich der Kontext der Nutzung über folgende Fragen erfassen [ebenda, S. 798f.]:

- Wer nutzt das System?
- Wofür wird das System genutzt?
- Wo wird das System genutzt?
- Wann wird das System genutzt?
- Warum wird das System genutzt?

Eine Vielzahl weiterer Faktoren wird verwendet, um Kontext zu beschreiben. So beispielsweise Charakteristiken der Umgebung des Nutzers wie die herrschende Geräuschkulisse. In weiteren Fällen werden physische oder physiologische Veränderung oder die vom Nutzer durchgeführte Aufgabe als Kontext herangezogen [vgl. Bristow 2004, S. 800]. Andere Autoren bedienen sich Aufzählungen relevanter Faktoren wie „location, identities of nearby people and objects and changes to those objects“ und berücksichtigen mitunter auch Uhrzeit, Jahreszeiten oder herrschende Temperatur [vgl. Dey 2001, S. 140]. Dey et al. setzen sich im Rahmen ihrer Arbeit zum Thema „Context-Aware Applications“ intensiv mit der Definition von Kontext auseinander. Sie stellen dabei fest: „Definitions of context by previous authors fall in two categories: enumerations of examples of context information and categorizations“ [ebenda] und weisen beide Varianten als unzureichend zurück. Letztlich gelangen sie zu folgender Definition:

> „Context: any information that can be used to characterize the situation of entities (i.e., whether a person, place, or object) that are considered relevant to the interaction between a user and an application, including the user and the application themselves. Context is typically the location, identity, and state of people, groups, and computational and physical objects“ [Dey 2001, S. 106].

Dieser Ansatz mag breit gefasst sein, bietet in seiner Allgemeingültigkeit jedoch wenig Orientierung, da sich letztlich wiederum fast beliebige Merkmale und Eigenschaften unter dem Begriff „Kontext“ subsumieren lassen. In der Praxis werden somit meist schlicht jene Faktoren als Kontext definiert, die für die jeweilige Anwendung, das jeweilige System oder die jeweils leitende Fragestellung Relevanz besitzen. Dabei gestattet die Offenheit des definitorischen Ansatzes positiv formuliert die flexi-

ble Berücksichtigung auch neuer relevanter Faktoren. Diese Vorgehensweise erweist sich als pragmatisch, wenn auch offen bleibt, inwiefern stets die geeignetsten Faktoren zur jeweiligen Definition von Kontext verwendet werden [vgl. Bristow 2004, S. 800]. Ferner folgt für Bristow et al. aus der ständigen Neudefinition auf Grundlage der Intuition der jeweiligen Forscher eine grundsätzliche Problematik: „While such a 'craft-based' approach is proving fruitful in many applications, it tends to lead to researchers rediscovering (or at least redefining) the whole concept of context. This, in turn, is producing a degree of circularity into both the literature and the research community, i.e. 'context' is in danger in being defined simply in terms of whatever features a 'context-aware' device is able to measure" [ebenda].

Diese Kritik scheint jedoch etwas einseitig. Eine universelle, aber detaillierte und verbindliche Definition von Kontext – so man sich innerhalb einer Disziplin auf eine solche einigen könnte – wäre ohnehin kaum praktikabel handhabbar. Die resultierende Komplexität würde eindeutige Aussagen, Abhängigkeiten, Schlussfolgerungen und Prognosen in hohem Maße erschweren. Eine Universaldefinition scheint jedoch auch nicht notwendig, da kein konkretes Modell alle potentiell verfügbaren Aspekte eines Kontextes gleichzeitig berücksichtigt. Es ist somit – wie auch Bristow et al. bestätigen – pragmatisch, wenn nicht gar unumgänglich, dass es abhängig vom jeweiligen konkreten Systems oder der jeweils leitenden Fragestellung unterschiedliche Faktoren sind, die den relevanten Kontext konstituieren. Solange die Auswahl der Faktoren nachvollziehbar begründet und deutlich deklariert wird, ist gegen diese Vorgehensweise kein grundsätzlicher Einwand vorzubringen.

Für das weitere Vorgehen liegt der Fokus auf Kontext als Menge relevanter Umgebungsfaktoren, denen die Nutzung mobiler Geräte generell unterworfen ist. Kontext im Sinne dieser Arbeit wird somit durch jene Faktoren konstituiert, welche die Nutzung allgemein beeinflussen – relativ unabhängig von der aktuellen Aufgabenstellung, dem konkret verwendeten System oder den individuellen Eigenschaften des Nutzers. Obgleich letztgenannte Aspekte einen Einfluss auf das konkrete Nutzungserlebnis haben und die Auswirkung der Umgebungsfaktoren moderieren können, bleiben diese Umgebungsfaktoren doch grundsätzlich wirksam. Es werden vier Dimensionen unterschieden:

1. *Umweltfaktoren* – hierzu zählen die jeweiligen Eigenschaften des (offenen oder geschlossenen) umgebenden Raumes, in dem die Nutzung stattfindet, beispielsweise Geräuschkulisse, Lichtverhältnisse oder Erschütterungen bei Nutzung in Fahrzeugen.

2. *Physische Aktivität* – umfasst die Bewegungsaktivität des Nutzers, primär ob die Benutzung im Sitzen, Stehen, Gehen oder Laufen erfolgt.

3. *Kognitive Faktoren* – berücksichtigen die kognitiven Belastungen, denen der Nutzer während der Interaktion mit dem System ausgesetzt ist. Solche Belastungen folgen zum einen aus der Nutzung in Bewegung, die eine gleichzeitige Orientierung im Raum und Koordination der Bewegung im Bezug auf die Umwelt (Passanten, Fahrzeuge, etc.) erfordert. Zum anderen resultieren kognitive Belastungen aus gleichzeitig ausgeführten Tätigkeiten (z.B. parallele Besprechung oder Verarbeitungen von Informationen der Umwelt). Diese parallelen Aktivitäten konkurrieren um kognitive Ressourcen und haben unter anderem Unterbrechungen, erhöhten Koordinationsaufwand beim Wechsel zwischen Tätigkeiten sowie erhöhte Anforderungen an Aufmerksamkeit und Konzentration zur Folge.

4. *Technische Faktoren* – hier werden jene Faktoren berücksichtigt, die von allgemeiner Relevanz für die Nutzung mobiler Endgeräte sind, insbesondere die Verfügbarkeit, Bandbreite und Stabilität der Infrastruktur (Mobilfunknetz, Internetzugang, etc.). Obgleich die technische Entwicklung vergangener Jahre die Unterschiede zwischen mobiler und Desktop-Nutzung reduziert hat (z.B. durch neue Standards wie UMTS, HSDPA, etc.), bestehen sie tendenziell fort. Insbesondere moderiert über die Variable geografische Position, da die Verfügbarkeit und ggf. Verlässlichkeit breitbandiger Verbindungen nicht flächendeckend gewährleistet ist (vgl. Übersichtskarten im Anhang, S. 82). Technische Faktoren werden im Kontext allgemeiner Betrachtungen vor allem berücksichtigt, weil sie – durch Verzögerungen, Abbrüche und Ausfälle – Auswirkungen auf die kognitiven Ressourcen des Nutzers haben können.

Faktoren von spezifischer Relevanz nur für konkrete Anwendungen, Systeme oder Nutzergruppen werden hingegen ausgeklammert. Im weiteren Verlauf werden die oben aufgeführten Kontextfaktoren berücksichtigt, welche die Nutzung mobiler Endgeräte grundsätzlich beeinflussen. Die daraus resultierenden Erkenntnisse können ihrerseits in weiterführende Betrachtungen zu spezifischeren Fragestellungen einfließen.

Der Nutzungskontext mobiler Endgeräte hat für den Nutzer entsprechend folgende grundsätzlichen Konsequenzen:

1. Die Benutzung mobiler Systeme erfolgt häufig in Bewegung oder schließt einen Wechsel von Ruhe und Bewegung ein.

2. Während der Interaktion mit dem mobilen System ist der Nutzer häufig mehreren kognitiven Belastungen gleichzeitig ausgesetzt. Diese können aus der Nutzung in Bewegung resultieren, die parallel kognitive Ressourcen für die Orientierung im Raum und die Bewegungskoordination erfordert. Häufig ist der Nutzer zudem mit weiteren Aufgaben in der Umwelt beschäftigt, die mitunter höhere Priorität als die Systeminteraktion haben [vgl. York 2004, S. 775]. Ablenkende Reize der Umgebung können zusätzliche Beeinträchtigungen der kognitiven Fähigkeiten des Nutzers zur Folge haben [vgl. ebenda, S. 776].

3. Der Nutzungskontext ist dynamisch und kann sich somit während der Nutzung jederzeit verändern. Dies ist einerseits durch Veränderung des Kontextes selbst möglich, beispielsweise durch veränderte Beleuchtungssituation, Umgebungsgeräusche oder erhöhte kognitive Belastung, z.B. in Folge dichteren Gedränges von Passanten auf einem Streckenabschnitt, welches zu erhöhtem Aufwand zur Koordination der Bewegung führt. Darüber hinaus können Kontextwechsel aus der Bewegung des Nutzers durch verschiedene Kontexte resultieren (z.B. Wechsel zwischen abgegrenzten Räumen, Innen- und Außenbereich, etc.). In ersterem Fall hat hierbei der Nutzer keinerlei Kontrolle über den Kontextwechsel, im zweiten Falle kann er die Qualität des neuen Kontextes in der Regel nicht bestimmen.

Die Nutzung von Desktop-PCs erfolgt statisch, der Nutzer sitzt in Ruhe vor dem verwendeten System. Üblicherweise ist der Systeminteraktion hierbei die volle Aufmerksamkeit und Konzentration gewidmet, da parallel selten weitere Tätigkeiten ausgeführt werden. Die zusätzlichen Belastungen des Nutzers durch Orientierung und Koordination im Raum entfallen grundsätzlich. Die Umweltfaktoren am Arbeitsplatz – egal ob Büro oder privater Arbeitsbereich – sind relativ konstant und können zudem in weiten Teilen durch den Nutzer beeinflusst werden, beispielsweise durch die Betätigung von Jalousien, die Änderung der Beleuchtungsintensität oder die Begrenzung akustischer Reize. Die Faktoren dieses typischen stationären Nutzungskontextes lassen sich ferner im Labor in einem Maße nachbilden, dass realer Nutzungskontext und Laborsituation des Usability Test sich hinsichtlich wesentlicher Kriterien hinreichend ähneln, um valide Ergebnisse erwarten zu lassen.

Im Unterschied zur Situation der Desktop-PCs gilt: „Mobile systems are typically used in highly dynamic contexts“ [Kjeldskov 2004, S. 600]. Mobile Endgeräte unterscheiden sich hinsichtlich des Nutzungskontextes deutlich von Desktop-PCs. Realistische Nutzungsszenarien mobiler Endgeräte sollten somit Elemente von Bewegung und Dynamik, parallel zu bewältigende Aufgaben und Konfrontation mit veränderlichen Umweltfaktoren beinhalten. Die Bedeutung realitätsnaher Bedingungen im Usability Test für die Validität wurde bereits dargestellt. Wie Vuolle et al. hervorheben, ist die Bedeutung des Kontextes bereits in der zentralen ISO Norm fest verankert: „According to ISO 9241 standard, usability is an extent to which the intended users of a product achieve specified goals in an effective, efficient and satisfactory manner *within a specified context of use*“ [Vuolle 2008, S. 2, Hervorhebung hinzugefügt]. Auf Basis dieser Ausgangslage stünde zu erwarten, dass der spezifische Nutzungskontext mobiler Endgeräte bei der Usability Evaluation eine zentrale Rolle spielt, insbesondere für die nutzerzentrierten Methoden.

4.4.1 EXTERNE UND INTERNE VALIDITÄT

Statt dessen stellt man bei Betrachtung der gängigen Praxis fest, dass die Evaluation mobiler Endgeräte weitgehend dem im Desktop-Bereich etablierten und bewährten

Vorgehen folgt. Tests werden im Labor durchgeführt, wo die Experimentalbedingungen in idealer Weise kontrolliert werden können. Die Beobachtung der Probanden und ihrer Interaktion mit dem System sowie die Aufzeichnung relevanter Informationen ist in diesem Umfeld effektiv möglich. So beschreibt Scott Weiss, einer der Pioniere auf dem Gebiet mobiler Usability, seinen üblichen Experimentalaufbau im Jahr 2002 folgendermaßen: das zu evaluierende System steht an einem kleinen Stativ befestigt auf einem Tisch, während eine stationär montierte Kamera Display und Eingabefläche filmt. Abbildungen 4 und 5 illustrieren dieses Vorgehen.

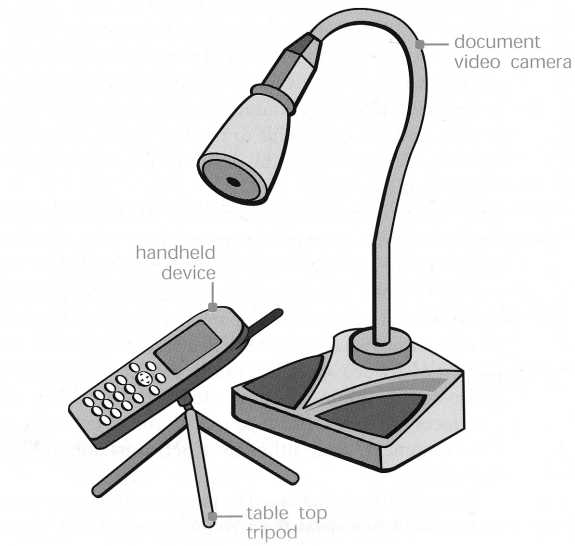

Abbildung 4: Video Setup für eine Usability-Studie, Abbildung entnommen aus [Weiss 2002, S. 174]

Abbildung 5: Das Setup im Einsatz, Abbildung entnommen aus [Weiss 2002, S. 176]

Das Testgerät kann bei diesem Aufbau also nicht wie üblich in der Hand gehalten werden, was wie in Abbildung 5 deutlich zu erkennen ist ferner zu einer ungewöhnlichen Haltung der bedienenden Hand führt, die der realen Nutzungsweise ebenfalls nicht entspricht.

Kjeldskov et al. berichten von einer Literaturstudie, die für den betrachteten Zeitraum von 2000 bis 2002 für Usability-Studien mit mobilen Endgeräten feststellt: „71% of these evaluations are conducted by means of laboratory experiments and very few of these laboratory evaluations involve special techniques being applied to meet the challenges of evaluating a *mobile* system“ [Kjeldskov 2004, S. 600, Hervorhebung im Original]. In einer eigenen vergleichenden Untersuchung, die den Zeitraum von 1996 bis 2002 umfasst, machten nur 6 von 114 berücksichtigten Studien Gebrauch von

„new techniques for usability evaluations in order to reflect or recreate a mobile use situation“ [ebenda, S. 602]. An dieser Situation hat sich in der Praxis bis heute wenig geändert. In einem aktuellen Projekt am *mobile media and communication lab* (m2c-lab) der FH Aachen wurde kürzlich eine Evaluationsstudie mit Smartphones unter vergleichbaren Voraussetzungen durchgeführt: „Testperson und Testleiter sitzen sich an einem Tisch gegenüber. Die Testperson wird gebeten das zu testende Gerät während der Bedienung unter eine Dokumentenkamera zu halten.[27]“

Die Einbeziehung des Nutzungskontextes bei Evaluationsmaßnahmen mobiler Endgeräte bildet in der Praxis weiterhin eine Ausnahme. Testbedingungen und realer Nutzungskontext weichen somit hinsichtlich relevanter Dimensionen erheblich voneinander ab. Als Folge resultiert die grundsätzliche Problematik geringer externer Validität der Evaluationsergebnisse, da unklar ist, inwieweit Laborergebnisse auf reale Nutzungssituationen übertragbar sind. Zhang et al. führen dazu aus: „A mobile application tested in a real environment may not work as well as it does in a controlled laboratory setting due to the changing and unpredictable (...) environmental factors. In a laboratory, participants may not experience the potential adverse effects of those contextual factors“ [Zhang 2005, S. 298]. Insbesondere die Ergebnisse des Nutzertests werden durch geringe externe Validität hinsichtlich ihrer Aussagekraft in Zweifel gezogen, was in Anbetracht der Tatsache, dass Nielsen ihn als „most fundamental usability method“ und „the cornerstone of most recommended usability practice“ [Nielsen 1993, S. 165 bzw. S. 207] bezeichnet, besondere Aufmerksamkeit verdient. Die Abhängigkeit vom Nutzungskontext betrifft letztlich jedoch alle drei zentralen Ziele der Usability, was für die Interpretation resultierender Evaluationsergebnisse berücksichtigt werden sollte und zu einer entsprechenden Komplexität der Datenlage führt. Obgleich tendenziell positive Korrelationen zwischen Effektivität, Effizienz und Zufriedenheit erwartet werden, ist laut Vuolle et al. zu beachten „that correlations between the three aspects of usability are dependent on the application domain, the user’s experience, *and the use context*, and that no simple correlation between the three can be expected“ [Vuolle 2008, S. 3, Hervorhebung hinzugefügt].

27 Das Projekt ist online dokumentiert unter:
http://www.m2c-lab.fh-aachen.de/smartphoneusability/content/einleitung.php

Die Ursache der häufig geringen Realitätsnähe bei der Evaluation mobiler Geräte ist primär in den vielfältigen Herausforderungen zu sehen, welche eine Berücksichtigung des mobilen Nutzungskontextes im Evaluationsprozess mit sich bringt. Im Bereich der Desktop-PCs entspricht das Laborsetting weitgehend dem realen Nutzungskontext. Auf dieser Grundlage lassen sich bei Bedarf auch zusätzliche Faktoren kontrolliert in den Experimentalablauf integrieren, um die Realitätsnähe weiter zu erhöhen. Soll beispielsweise der Einfluss von Unterbrechungen durch Anrufe während der Aufgabenbewältigung erfasst werden, um der Situation am realen Arbeitsplatz näher zu entsprechen, so lassen diese Anrufe sich in standardisierter Weise in den Testablauf integrieren. Die Schaffung identischer Situationen für jeden Probanden gewährleistet dabei weiterhin ein hohes Maß von Vergleichbarkeit der Ergebnisse. Den Anforderungen eines mobilen Nutzungskontextes ist in einem kontrollierten Laborsetting hingegen kaum gerecht zu werden. Allein die Aspekte Bewegungsfreiheit und Koordination in einem realistischen Umfeld werfen Grundsätzliche Fragen hinsichtlich der Umsetzbarkeit auf. Auch ist „Realitätsnähe“ bei mobilen Endgeräten nicht klar definierbar. Die unüberschaubare Vielfalt möglicher Nutzungskontexte kann nicht erschöpfend berücksichtigt werden. Eine Auswahl repräsentativer Kontexte, die eine breite Basis realer Situationen abdeckt, erscheint hier jedoch als angemessene Vorgehensweise. Auf Grundlage der hier vertretenen Definition von Kontext lassen sich zudem vielfältige spezifische Situationen auf Basis von Gemeinsamkeiten hinsichtlich der Umweltfaktoren, physischen Aktivität, kognitiven Faktoren und/oder technische Faktoren zusammenfassen.

Eine naheliegende Lösung zur Berücksichtigung des Nutzungskontextes scheint der Einsatz von Feldstudien zur Evaluation mobiler Endgeräte anzubieten, da hier grundsätzlich repräsentative Nutzer in realen bzw. möglichst realitätsnahen Situationen untersucht werden. Beobachtung und Messung gestalten sich hierbei jedoch schwierig und aufwendig [vgl. Zhang 2005, S. 300]. Zudem leidet die Realitätsnähe gegebenenfalls unter der Begleitung des Probanden durch einen beobachtenden Usability-Experten und/oder den Einsatz zusätzlicher Technik zu Aufzeichnungszwecken, beispielsweise am mobilen Gerät angebrachte Miniaturkameras, welche die Nutzerinteraktion aufzeichnen. Da Vorbereitung, Durchführung und Auswertung zudem zeit-, kosten- und ressourcenintensiv sind, sprengen umfassende Feldstudien häufig den Rahmen des in der Praxis vieler Usability-Projekte Realisierbaren. Darüber hinaus

muss bei Feldstudien für eine erhöhte externe Validität zugleich eine geringere interne Validität in Kauf genommen werden. Aufgrund der Durchführung im öffentlichen Raum können viele Faktoren nicht konstant gehalten werden. Jeder Durchlauf variiert somit stark von den übrigen, Störvariablen sind kaum zu kontrollieren. Die resultierenden Ergebnisse sind daher nicht zwingend auf das getestete System zurückzuführen, da die Beeinflussung durch vielfältige konfundierende Variablen nicht ausgeschlossen werden kann.

Ein alternativer Ansatz, der eine Verknüpfung der Vorzüge von Kontrollierbarkeit und Realitätsnähe anstrebt, scheint auf Basis der der bereits angestellten Betrachtungen und der im folgenden dargestellten kognitiven Aspekte eventuell erreichbar.

5 BETRACHTUNG KOGNITIVER ASPEKTE

„Das Ziel der Kognitiven Psychologie besteht darin, das Wesen der menschlichen Intelligenz und ihrer Funktionsweise zu erfassen“. Hierzu beschäftigt sie sich mit „den grundlegenden Mechanismen des menschlichen Denkens“ [Anderson 2001, S. 1 bzw. S. 3]. Die Annäherung an diese Thematik geschieht dabei aus Perspektive der Informationsverarbeitung. Um den menschlichen Informationsverarbeitungsprozess nachzuvollziehen werden hierzu komplexe kognitive Prozesse in Sequenzen von Einzelschritten zerlegt, „in denen eine abstrakte Größe, die Information, verarbeitet wird“ [vgl. ebenda, S. 12]. Die betrachtete Sequenz beginnt hierbei auf grundlegendster Ebene bei der physiologischen Verarbeitung eintreffender Sinnesreize und schreitet fort beispielsweise über Aspekte der Wahrnehmung, Kodierung und Mustererkennung bis hin zu elaborierten kognitiven Konzepten wie der Repräsentation von Wissen, der Funktionsweise des Gedächtnisses oder komplexen Problemlösestrategien.

Aufgrund dieses breiten Gegenstandsbereiches sind kognitive Aspekte für fast alle Bereiche der Usability von Interesse. Erkenntnisse hinsichtlich visueller Wahrnehmung fließen so beispielsweise in Gestaltungsempfehlungen für die Verwendung von Farben und Kontrasten in Benutzerschnittstellen ein. Modelle aus dem kognitionspsychologischen Bereich der Objektwahrnehmung, beispielsweise die „Gestaltgesetzte der Wahrnehmungsorganisation“ [vgl. ebenda, S. 46ff.], werden im Usability-Kontext angewandt, um dargestellte Information zu strukturieren und Zusammengehörigkeiten zwischen Informationseinheiten für den Nutzer nachvollziehbar zu visualisieren [vgl. Nielsen 1993, S. 117f.].

Im Folgenden sind jedoch jene kognitiven Aspekte von Interesse, die aus dem Nutzungskontext mobiler Endgeräte resultieren und die somit die Verwendung dieser Geräte beeinflussen können. Dies betrifft insbesondere Fragen hinsichtlich der limitierten Verfügbarkeit kognitiver Ressourcen sowie zur Auswirkung von parallel ausgeführten und potentiell konkurrierenden kognitiven Prozessen, welche die Multi-Tasking-Anforderungen mobiler Nutzungskontexte mit sich bringen.

Frühe Modelle des Gedächtnisses nehmen eine modularen Aufbau an, der funktional getrennte Einheiten unterscheidet. Jene sensorischen Informationen, denen wir unsere Aufmerksamkeit widmen gelangt dieser Annahme folgend zunächst in ein Kurzzeitgedächtnis, von wo sie erst durch den aktiven Prozess des Memorierens ins Langzeitgedächtnis überführt werden können [vgl. Anderson 2001, S. 174ff.]. Obgleich das Modell des Kurzzeitgedächtnisses als überholt gilt, erfolgt eine Erwähnung, da die resultierende Implikationen sowie die Terminologie – insbesondere im Usability-Kontext – durchaus noch anzutreffen sind. Baddeley und Hitch lassen in ihrem einflussreichen Modell an die Stelle des Kurzzeitgedächtnisses ihr Konstrukt des Arbeitsgedächtnisses treten, in dem eine zentrale Exekutive auf zwei getrennte „Arbeitsspeicher“ für visuell-räumliche (*visuospatial sketchpad*) und phonologische (*phonological loop*) Information zurückgreifen kann. Die zentrale Exekutive steuert hierbei die Manipulation der Inhalte der beiden Arbeitsgedächtnisspeicher [vgl. Pollmann 2008, S. 175ff.]. Wie die Bezeichnung bereits vermuten lässt, übernimmt der Arbeitsspeicher eine aktivere Rolle. Hier werden nicht nur Informationen zwischengespeichert, bevor sie in das Langzeitgedächtnis transferiert werden können. Das Arbeitsgedächtnis hält vielmehr kurzfristig benötigte Information bereit, mit der aktiv gearbeitet wird, d.h. das Arbeitsgedächtnis ermöglicht das „Halten von Information und <ihre> Manipulation“ [ebenda, S. 176]. Abbildung 6 stellt das Gedächtnismodell mit zwischengeschaltetem Kurzzeitgedächtnis dar [entnommen aus Anderson 2001, S. 175], Abbildung 7 illustriert das Modell des Arbeitsgedächtnisses nach Baddeley [entnommen ebenda, S. 180].

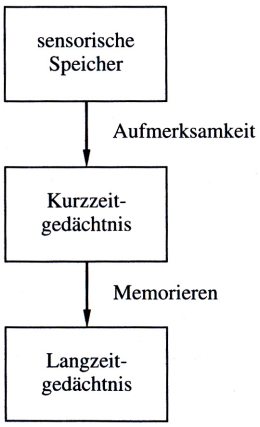

Abbildung 6

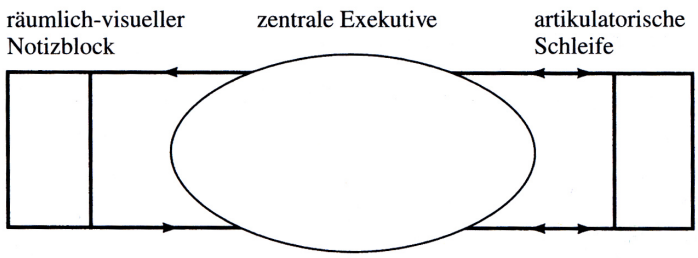

Abbildung 7

Beide Modelle tragen mit ihrer Zweiteilung des Gedächtnisses dem empirischen Befund Rechnung, „daß die Informationsmenge, die Menschen zu einem bestimmten Zeitpunkt memorieren können, beschränkt ist“ [Anderson 2001, S. 178]. Während die Kapazität des Langzeitgedächtnisses grundsätzlich als unbegrenzt angenommen wird, verfügt das Kurzzeitgedächtnis nur über eine sehr beschränkte Kapazität[28]. Gleichzeitig können Informationen im Langzeitgedächtnis potentiell auch zeitlich unbegrenzt gespeichert werden, während Inhalte des Kurzzeit- bzw. Arbeitsgedächtnis sehr flüchtig sind und verloren gehen, sofern sie nicht aktiv memoriert werden: „within current theories of STWM[29], short-term traces are assumed to decay completely, in a time of about 30s, when no rehearsal can be carried out“ [Oulasvirta 2006, S. 944]. Auch das Verfügbarhalten von Inhalten des Arbeitsgedächtnisses setzt somit kognitive Aktivität voraus, die Ressourcen benötigt.

Neben Reizen der Außenwelt können auch Inhalte aus dem Langzeitgedächtnis abgerufen und im Arbeitsgedächtnis verarbeitet werden. Eines der etablierten Modelle des Informationsabrufs aus dem Langzeitgedächtnis ist Andersons ACT-Theorie. Zwei Variablen bestimmen hierbei die Verfügbarkeit von Gedächtnisinhalten und somit die Wahrscheinlichkeit des Zugriffs: *Aktivation* [vgl. Anderson 2001, S. 183ff.] und *Stärke* [vgl. ebenda, S. 188 ff.]. Die Aktivationshöhe wird bestimmt durch die kürzliche Häufigkeit des Zugriffes auf einen bestimmten Gedächtnisinhalt und den letzten Zeitpunkt eines solchen Zugriffs. Dies hat zur Folge, dass kürzlich und häufig verwendete Inhalte eine höhere Zugriffswahrscheinlichkeit auszeichnet. Die höhere Aktivation klingt hierbei nach der Verwendung allmählich wieder ab, so dass ihre Auswirkungen umso stärker sind, je kürzer der letzte Zugriff zurückliegt. Aktivation breitet sich ferner über Netzwerkstrukturen aus [vgl. ebenda, S. 185ff.]. Als Folge hiervon erfahren auch jene Elemente, die mit einem aufgerufenen Gedächtnisinhalt verknüpft sind, eine Erhöhung ihrer Aktivation. Auf diese Weise wird auch ihr Abruf wahrscheinlicher, was gleichbedeutend ist mit einer Erhöhung der Verfügbarkeit oder Zugänglichkeit. Während die Aktivationshöhe zeitlich variabel ist und somit raschen

28 Eine häufig angetroffene Quantifizierung dieser Beschränkung besagt, die Kapazität des Kurzzeitgedächtnisses betrage sieben (mitunter auch +/-2) Elemente [vgl. auch Anderson 2001, S. 175]. Da jedoch der maximale Inhalt dieser Elemente nicht hinreichend definiert ist und die Technik des Chunking es gestattet, komplexere Informationen nicht verbindlich bestimmten Umfangs in ein Element zusammenzufassen, scheint der Aussagegehalt dieser Regel eher gering. Neue Arbeiten stellen sie zudem grundsätzlich in Frage [vgl. Pollmann 2008, S. 178f.].

29 Short-Term Working Memory

Schwankungen unterliegen kann, ist die Stärke ein nur langsam veränderlicher Faktor. Mit jedem Abruf eines spezifischen Inhalts wird dessen Stärke ein wenig erhöht, so dass sie durch wiederholte Übung gezielt und effektiv gesteigert werden kann: „Durch Übung eines Gedächtnisinhalts steigt dessen Stärke nach der Potenzfunktion“ [Anderson 2001, S. 190]. Die Stärke beeinflusst hierbei die maximale Aktivationshöhe und somit die Verfügbarkeit eines Gedächtnisinhalts.

Aktuelle Modelle des Arbeitsgedächtnisses wie das Konzept von Cowan verzichten auf eine Teilung des Gedächtnisses in separate Komponenten und ähneln in gewisser Weise eher einer Netzwerkstruktur, wie sie in der ACT-Theorie beschreiben wird. Der Arbeitsspeicher wird hier verstanden als der gerade aktive Teil des Langzeitgedächtnisses. Innerhalb dieses aktiven Bereichs finden sich wiederum jene Inhalte, die uns aktuell bewusst sind, weil sie sich im Fokus der Aufmerksamkeit befinden [vgl. Pollmann 2008, S. 177]. „Der Fokus der Aufmerksamkeit bestimmt damit auch die Anzahl der Items, die wir im Arbeitsgedächtnis behalten können“ [ebenda]. Neben den Inhalten innerhalb des Fokus gibt es weitere Inhalte in dessen Peripherie, die beispielsweise aufgrund assoziativer Verknüpfung mit aktiven Inhalten, häufiger oder kürzlicher Verwendung ebenfalls eine erhöhte Aktivierung besitzen.

Aufmerksamkeit ist nicht nur in diesem Beispiel ein Faktor, der kognitive Aktivität steuert und fokussiert. Die kognitiven Ressourcen des Menschen sind begrenzt, weshalb es in vielen Situationen erforderlich ist, sich möglichst auf „eine Reizkonstellation und eine Aufgabe“ zu beschränken [ebenda, S. 96]. Eine wichtige Funktion der Aufmerksamkeit besteht somit in der *Selektion* von Aspekten unserer Umgebung: für die Bearbeitung jener Elemente im Fokus der Aufmerksamkeit werden Ressourcen zur Verfügung gestellt, während die umgebenden Elemente gezielt vernachlässigt werden. Diese Selektion kann als Filterprozess verstanden werden, in dem jene Aspekte, die keine Aufmerksamkeit erfahren, ausgefiltert werden. Abhängig von der kognitiven Gesamtbelastung wird diese Filterung in zweierlei Weise vorgenommen. Bei früher Selektion erfolgt keine weitergehende Verarbeitung der gefilterten Reize. Selektion findet dann statt, „wenn die Verarbeitung aufgrund der Vielzahl von Stimuli und/oder Antwortanforderungen ohne Filter überlastet würde“ [ebenda, S. 98]. Besteht die Gefahr einer Überlastung aufgrund moderater Anforderungen der aktuellen Situation nicht, so kann auf eine frühe Selektion verzichtet werden. In diesem Fall

werden auch die irrelevanten Reize „(unbewusst) relativ weitgehend analysiert“ [Pollmann 2008, S. 98].

Aufmerksamkeit hat grundsätzlich eine selektive Funktion, die quasi eine Beschränkung auf das gerade Relevante oder das mit den vorhandenen kognitiven Ressourcen Bewältigbare bedeutet. Interessant ist die Anwendungsbreite des Konzepts der Aufmerksamkeit: von der Verarbeitung eingehender Sinnesreize, beispielsweise in der auditiven und visuellen Aufmerksamkeit, dem Abruf oder der Arbeit mit Gedächtnisinhalten bis hin zur Planung zielgerichteten Handelns: „Goal-directed behavior requires focusing attention on goalrelevant stimuli while ignoring irrelevant distractors“ [Lavie 2004, S. 339]. Anderson beschreibt diese Situation folgendermaßen: „Das Gehirn besteht aus einer Reihe von parallel verarbeitenden Systemen, die zuständig sind für die verschiedenen perzeptuellen Systeme, motorischen Systeme und zentrale Kognition. Jedes dieser Systeme wird durch einen Flaschenhals eingeschränkt, wodurch es seine Verarbeitung auf einen einzelnen Sachverhalt fokussieren muß. Aufmerksamkeit begreift man am besten als einen Prozeß, durch den jedes dieser Systeme den möglicherweise konkurrierenden Anforderungen der Informationsverarbeitung zugeordnet wird“ [Anderson 2001, S. 105].

5.2 AUTOMATISIERUNG

Gleichzeitig ist Aufmerksamkeit selbst eine knappe kognitive Ressource: wir können unsere Aufmerksamkeit nicht beliebig vielen Reizen, Aufgaben oder Prozessen widmen. Eine Möglichkeit, die kognitiven Belastungen eines Prozesses signifikant zu reduzieren, besteht in der Automatisierung des Prozesses, da automatisierte Prozesse „wenig oder keine Aufmerksamkeit erfordern“ [Anderson 2001, S. 101]. Anderson beschreibt die verschiedenen perzeptuellen (z.B. visuelle und auditive) und motorischen (z.B. Artikulation, Handbewegungen) Systeme und die zentrale Kognition als „voneinander unabhängige Systeme, die parallel arbeiten“, unterstreicht jedoch, dass jedes System in sich betrachtet seriell arbeitet: „Man kann nur auf eine Raumregion aufmerksam sein, nur ein Muster von Handbewegungen vollziehen oder nur einen Gedanken denken“ [ebenda]. Die zentrale Kognition als koordinierende

Instanz führt hierbei häufig zu Engpässen, da sie von vielen Aufgaben in Anspruch genommen wird. An dieser Stelle führt Automatisierung zu einer Entlastung. Erreicht wird sie durch wiederholtes Üben: „mit steigender Übung wird eine Aufgabe stärker automatisiert und erfordert dadurch bei ihrer Ausführung zunehmend geringere zentrale Kognition“ [Anderson 2001, S. 101]. Bei der Automatisierung handelt es sich somit um einen graduellen Prozess.

Klassische Beispiele für automatische Prozesse sind das Autofahren oder die Worterkennung beim Lesen, aber auch motorische Prozesse wie das Maschinenschreiben. Während der Lern- und Übungsphase erfordern diese Tätigkeiten wie üblich ein hohes Maß von Aufmerksamkeit und kognitiver Kontrolle. Je weiter die Automatisierung jedoch voranschreitet, desto geringere zentrale Kognition ist erforderlich und umso schwächer wird die Interferenz der automatisierten Prozesse mit parallel laufenden Prozessen. So bereitet es einem geübten Autofahrer in der Regel keine Probleme, das Fahrzeug zu bedienen, den umgebenden Verkehr zu beobachten und eine angeregte Unterhaltung zu führen. Die Vorteile der Automatisierung sind hierbei umso höher, je mehr die aktuelle Situation der geübten entspricht. Bei Abweichungen und in kognitiv besonders belastenden Situationen (z.B. Orientierung im Verkehr einer fremden Großstadt, häufige Spurwechsel in dichtem Verkehrsaufkommen, etc.) benötigen hingegen auch automatische Prozesse zusätzliche zentrale Kognition, so dass parallele Prozesse wie das Führen eines Gesprächs erschwert werden.

5.3 DUAL-TASK PARADIGMA

Bei Dual-Task-Experimenten bearbeiten Probanden zwei Aufgaben gleichzeitig. Hierbei kommt es zur Inteferenz zwischen den Aufgaben, da beide grundsätzlich kognitive Ressourcen beanspruchen. Das Ausmaß dieser Interferenz variiert jedoch und ist abhängig davon, inwiefern beide Aufgaben auf die gleichen Informationsverarbeitungsstrukturen zugreifen. „An obvious example of such a structural distinction is between the eyes (visual processing) and the ears (auditory processing). In many circumstances, dual task performance is poorer when two visual tasks must be time shared than in a configuration in which the equivalent information for one of the tasks

is presented auditorally“ [Wickens 2002, S. 162]. Diese Strukturen, die beispielsweise an Sinnesmodalitäten gekoppelt sind, verhalten sich somit jeweils wie eigenständige Ressourcen. Dadurch gilt grundsätzlich: „cross-modal time-sharing is better than intra-modal time-sharing“ [ebenda, S: 164]. Die Performance zweier parallel bearbeiteter Aufgaben ist somit höher, wenn sie sich unterschiedlicher Sinnesmodalitäten bedienen, während aus der Verwendung identischer Modalitäten für beide Aufgaben erhebliche Interferenzen resultieren. Bildlich gesprochen bildet die von beiden Aufgaben benötigte Ressource einen Engpass („Flaschenhals“ bzw. „Bottleneck“) im Verarbeitungsprozess, da sie zu jedem Zeitpunkt nur für eine der Aufgaben zur Verfügung steht und die Bearbeitung der zweiten Aufgabe entsprechend pausieren muss [vgl. Pollmann 2008, S. 111]. Automatisierte Aufgaben benötigen wie gesehen grundsätzlich nur minimale kognitive Ressourcen, weshalb sie nicht oder nur minimal mit einer parallel bearbeiteten Aufgabe interferieren [vgl. Wickens 2002, S. 161].

Die Anwendungsmöglichkeiten, die sich aus diesem Paradigma ergeben sind vielfältig. Die Auswirkung der Primäraufgabe auf die Performance der Sekundäraufgabe gestattet beispielsweise Rückschlüsse auf das Ausmaß ihrer kognitiven Belastung oder den Grad ihrer Automatisierung. Eine Sekundäraufgabe kann als Distraktor eingesetzt werden, um die Auswirkung kognitiver Belastungen auf die Performance der Primäraufgabe zu untersuchen. Aus dem Grad beobachteter Interferenz können Aussagen über den Deckungsgrad der jeweils benötigten Informationsverarbeitungsstrukturen getroffen werden und vieles mehr.

5.4 AUFGABENWECHSEL

Der Wechsel zwischen Aufgaben verursacht „Wechselkosten“ in Form von Belastungen kognitiver Ressourcen. Diese äußern sich beispielsweise in erhöhtem zeitlichen Aufwand für die wechselnde Bearbeitung zweier Aufgaben im Vergleich zur Summe des Zeitbedarfs der beiden Einzelaufgaben [vgl. Pollmann 2008, S. 114f.]. Es handelt sich dabei nicht nur um Effekte zusätzlicher Belastung des Arbeitsgedächtnisses, in dem die Bedingungen beider Aufgaben verfügbar gehalten werden müssen. Vielmehr kann experimentell gezeigt werden, „dass Aufgabenwechsel genuine Verarbeitungs-

kosten verursachen“ [Pollmann 2008, S. 115]. Dies liegt zum einen an der „Trägheit der Aufgabenkonfiguration“: der Satz von Regeln, der die Reaktion auf Reize in der ersten Aufgabe festlegt, muss für die zweite Aufgabe aktiv inhibiert werden [vgl. ebenda, S. 115]. Zum anderen erfolgt bei jedem Wechsel eine Vorbereitung auf die nachfolgende Aufgabe, was ebenfalls kognitive Ressourcen beansprucht [vgl. ebenda, S. 116].

5.5 UNTERBRECHUNGEN

Wechselkosten treten ebenfalls auf, wenn die Bearbeitung einer Aufgabe durch äußere Einflüsse unterbrochen wird und die Beschäftigung mit einer zweiten Aufgabe somit extern bedingt ist. Sofern keine Vorwarnung auf die neue Aufgabe erfolgt, ist eine Vorbereitung hier nicht möglich. Oulasvirta et al. weisen auf weitere häufige Konsequenzen solcher Unterbrechungen hin: „interruptions pose a serious problem to task performance“ [Oulasvirta 2006, S. 941]. Dies schlägt sich beispielsweise in dem Versäumnis nieder, die ursprüngliche Aufgabe nach der Unterbrechung wieder aufzunehmen oder in der für die Rückkehr zur Bearbeitung der Ausgangsaufgabe benötigten Zeit [vgl. ebenda, S. 941f.].

Diese Problematik resultiert aus den Beschränkungen des Gedächtnisses. Oulasvirta et al. beschreiben die Notwendigkeit des Aufbaus einer sogenannten „retrieval structure“, die „retrieval cues and intra-item associations“ enthält. Diese Struktur wird verwendet, um nach einer Unterbrechung den vorausgehenden Zustand zu rekonstruieren. Der Aufbau einen solchen Struktur (encoding) benötigt jedoch eine gewisse Zeit und wird beeinträchtigt, wenn eine Unterbrechung abrupt erfolgt und einen sofortigen Aufgabenwechsel erfordert [vgl. ebenda, S. 951]. „Weak encoding of retrieval structures (…) should lead to two negative effects: first, decreased overall accuracy because of less encoded materials (…); second, interruption-caused impairment mainly in resuming the task when new materials should be integrated to existing ones“ [ebenda, S. 952]. Die Reaktivierung der Aufgabenrepräsentation nach einer Unterbrechung erfolgt hierbei in erster Linie auf Basis von „proper environmental retrieval cues“ [ebenda, S. 953].

6 IMPLIKATIONEN

6.1 IMPLIKATIONEN FÜR DAS USABILITY ENGINEERING

Die im folgenden dargestellten Empfehlungen sind grundsätzlich auf Basis individueller Applikationen für den mobilen Kontext einsetzbar und können dort jeweils zur Veresserung der Usability beitragen. Wünschenswert erscheint jedoch eine Umsetzung auf Ebene des Betriebssystems, da entsprechende Funktionen dadurch in standardisierter Weise für alle Applikationen der jeweiligen Plattform zur Verfügung gestellt werden können. Die resultierende konsistente Umsetzung etabliert Konventionen, die den Nutzern das Erlernen und Verwenden der Funktionalitäten erleichtern sollten. Für die Entwickler von Applikationen führt die Integration in das Betriebssystem zu Entlastungen, da nicht jeweils individuelle Lösungen für allgemein relevante Situationen entwickelt und getestet werden müssen.

6.1.1 AUTOMATISIERUNG UNTERSTÜTZEN

Der mobile Nutzungskontext stellt vielfältige Anforderungen an den Nutzer, die mit parallelen kognitiven Belastungen einhergehen. Jede Entlastung auf Seiten der Systeminteraktion befreit hierbei kostbare Ressourcen, die somit für die Interaktion mit der Umwelt und parallel ausgeführte Aktionen zur Verfügung stehen. Eine effektive Möglichkeit, kognitive Belastungen zu minimieren, besteht wie gesehen in der Automatisierung von Prozessen. Aus diesem Grunde empfiehlt es sich, diesen Aspekt bei der Konzeption mobiler Geräte zu berücksichtigen. Dies stellt eine gewisse Herausforderung dar, da bei mobilen Geräten nur wenige Interaktionsprozesse so repetetiv ablaufen, dass aufgrund wiederholter Übung ein hoher Automatisierungsgrad erreichbar ist. Wie dargestellt ist Übung der zentrale Faktor für die Automatisierung von Prozessen. Da beispielsweise Mobiltelefone jedoch über einen längeren Zeitraum (häufig die übliche Vertragslaufzeit von 24 Monaten) häufig und intensiv genutzt werden, scheint die Berücksichtigung effektiver Interaktionsformen, von

denen erst fortgeschrittene Nutzer mit steigender Übung profitieren, durchaus angemessen.

Ein relevantes Beispiel für Prozesse mit Automatisierungspotential stellt die Texteingabe dar. So lässt sich bei geübten Nutzer beispielsweise beobachten, dass Kurzmitteilungen (SMS) nicht nur besonders schnell, sondern auch weitgehend „blind“ – also mit minimaler visueller Aufmerksamkeit – verfasst werden können[30]. Aktuelle Entwicklungen mobiler Geräte beeinträchtigen diese Automatisierungsmöglichkeit jedoch. Der Anteil von Geräten mit Touchscreens erfährt derzeit ein rasantes Wachstum, welches die ohnehin dynamische Entwicklung des Smartphone-Sektors noch deutlich übertrifft[31]. Zudem bieten viele Hersteller mittlerweile auch im Preiseinstiegssegment Modelle mit Touchscreen-Bedienung an, so dass mittelfristig eine Dominanz dieses Interaktionsprinzips erwartet werden kann. Hinsichtlich der Automatisierung ist diese Entwicklung zunächst kritisch zu beurteilen, da Touchscreens keine haptischen Informationen zur Position der aktiven Tastenfläche bieten. Den virtuellen Tasten fehlt zudem ein spürbarer Druckpunkt und somit ein haptisches Feedback über das erfolgreiche Betätigen einer Taste. Eine Bedienung ohne visuelle Aufmerksamkeit ist aus diesen Gründen nahezu unmöglich.

Empirische Studien belegen, dass bereits eine Behebung des zweiten Defizits durch Integration haptischen Feedbacks für virtuelle Bedienelemente auf einem Touchscreen (Buttons, Scrollbars, etc.) positive Effekte für die Usability zur Folge haben kann: „Our haptically augmented progress bars and scroll bars led to significantly faster task completion, and favourable subjective reactions“ [Leung 2007, S. 374]. Mittlerweile realisieren bereits einige am Markt verfügbare Geräte ein solches Feedback auf Nutzereingaben (z.B. Rückmeldung auf erfolgten Tastendruck, Verwendung von Scrollbars) auf Basis piezoelektrischer Elemente. Hinsichtlich einer Automatisierung von Prozessen ist auf diese Weise jedoch nur wenig gewonnen, da haptische Information zur *Position* virtueller Elemente weiterhin fehlt.

30 Wege durch komplexe Menüs zu häufig verwendeten Funktionen stellen ein weiteres Beispiel für Automatisierungspotential dar, obgleich in diesen Fällen die grundsätzliche Vermeidung solcher Wege – beispielsweise durch konfigurierbare Schnellwahltasten, ein individuelles Favoriten-Menü o.ä. - zweifellos vorzuziehen wäre.

31 vgl. comScore Pressemitteilung vom 3.11.2009: http://www.mediametrix.com/index.php/layout/set/popup/Press_Events/Press_Releases/2009/11/Touchscreen_Mobile_Phone_Adoption_Grows_at_Blistering_Pace_in_U.S._During_Past_Year

Die kritische Beurteilung dieser Tatsache wird offenbar auch von Apple geteilt. So wurde beim US Patent und Markenamt ein Patent zur Anmeldung gebracht, mit dem eben diese Lücke geschlossen werden soll[32]. Im der Beschreibung des Patents heißt es hierzu: „However, one of a touchscreen's biggest advantages (i.e., the ability to utilize the same physical space for different functions) is also one of a touchscreen's biggest disadvantages. When the user is unable to view the display (because the user is occupied with other tasks), the user can only feel the smooth, hard surface of the touchscreen, regardless of the shape, size and location of the virtual buttons and/or other display elements“. Die angestrebte technische Lösung beinhaltet daher eine „haptic feedback response enabling a user to feel the location of the virtual button“, was explizit eine Bedienung ohne visuelle Aufmerksamkeit ermöglichen soll („non-visual use of an electronic device“). Der Einsatz einer solchen Technologie würde auch für zukünftige Geräte mit Touchscreen eine Automatisierung von Prozessen unterstützen und damit zur kognitiven Entlastung des Nutzers beitragen.

6.1.2 MULTIMODALITÄT ANBIETEN

„Mobile interactive systems (…) raise a usability challenge that is encountered to a lesser degree in stationary systems: Users often try to use such a system while simultaneously performing one or more other tasks that are related to their current environment“ [Jameson 2005, S. 350]. Eine sinnvolle Reaktion auf die Herausforderungen des Multitasking kann in der Unterstützung von Multimodalität bestehen. Wie gesehen interferieren Aufgaben stärker, wenn sie sich der gleichen Sinnesmodalitäten bedienen. Hieraus folgt grundsätzlich die Empfehlung, die Systeminteraktion soweit möglich auf wenig verwendete und somit gering belastete Sinneskanäle auszulagern. Insbesondere die visuelle Aufmerksamkeit ist ein zentraler Faktor bei Orientierung und Koordination im Raum und sollte daher so wenig als möglich durch die Systeminteraktion gebunden werden. Zur Entlastung des visuellen Kanals bieten sich primär akustische und taktile Kanäle an. Dies gilt grundsätzlich sowohl für die Ein-

32 http://appft.uspto.gov/netacgi/nph-Parser?Sect1=PTO2&Sect2=HITOFF&p=1&u=%2Fnetahtml%2FPTO%2Fsearch-bool.html&r=1&f=G&l=50&co1=AND&d=PG01&s1=20090167704&OS=20090167704&RS=20090167704

gabe (z.B. Sprachsteuerung) als auch für die Ausgabe (z.B. akustisches oder taktiles Feedback auf Nutzereingaben, Systemnachrichten, Alarme, etc.). Der Ansatz der Multimodalität sieht dabei vor, dass der Nutzer zwischen verschiedenen Modalitäten wählen kann. Auf diese Weise kann flexibel auf die jeweiligen situativen Anforderungen des Nutzungskontextes reagiert werden.

Die Vorzüge dieser Möglichkeit scheint sich den Nutzern intuitiv zu erschließen: Jameson et al. vergleichen in einer Studie die Auswirkungen von sprachgesteuertem Wählen („ear-based dialing“), bei dem auch das Feedback akustisch erfolgt und dem Wählen über die Tastatur, bei dem das Feedback über den visuellen Kanal aufgenommen wird. Bei Erhebung per Fragebogen wurde ein deutlicher Zusammenhang zwischen Nutzungskontext und Wahl der Modalität festgestellt, mit zunehmender Präferenz für das „ear-based dialing“ in Abhängigkeit von der Bewegungsintensität [vgl. Jameson 2005, S. 371]. Einschränkend muss jedoch festgestellt werden, dass diese geäußerte Präferenz bei der Beobachtung von sechs Testpersonen in der Praxis nicht zu einer entsprechenden Wahl der Modalität führte [vgl. ebenda, S. 369f.]. Die Autoren der Studie betonen dennoch: „Although the responses concerning method choice should not be seen as reliable predictors of actual behavior, they do give us a picture of the relevant knowledge and ideas that people have“ [ebenda, S. 371]. Die beobachtete Diskrepanz zwischen Präferenz und Verhalten ist vermutlich nicht zuletzt auf Usability-Probleme bei der Auswahl von Modalitäten zurückzuführen, was besondere Anstrengungen zur Optimierung in diesem Bereich nahelegt.

Sofern die Wahl geeigneter Modalitäten keine zusätzlichen Beeinträchtigungen mit sich bringt (z.B. Usability-Probleme), ist der Vorteil situativer Flexibilität auf Basis von Multimodalität nicht nur aus kognitionspsychologischer Perspektive gewiss, sondern findet auch Belege in empirischen Studien. So beschreiben Oviatt et al. das Verhalten der Nutzer folgendermaßen: „users of a multimodal interface spontaneously respond to dynamic changes in their own cognitive load by shifting to multimodal communication“ [Oviatt 2004, S. 135]. Sie finden ferner auch in der Praxis eine hohe Akzeptanz multimodaler Interaktion: „With respect to preferences and individual differences, 61.8% of all user interactions with the system involved multimodal communication, confirming past findings of a strong preference for multimodal interaction“ [ebenda]. Die im vorangegangenen Abschnitt aufgegriffene Studie von Leung et al.

kann auch in diesem Kontext als Indiz für „significantly faster task completion, and favourable subjective reactions“ [Leung 2007, S. 374] auf Basis von Multimodalität herangezogen werden.

6.1.3 INTERAKTIONSMÖGLICHKEITEN NUTZEN

Mobile Geräte verfügen über eine Reihe neuartiger Interaktionsmöglichkeiten, für die es bei Desktop-Systemen keine Entsprechung gibt. Fortschrittliche Smartphones verfügen beispielsweise über hochauflösende Sensoren für Bewegung und Beschleunigung. Auf diese Weise wird das gesamte Gerät zu einem Eingabegerät, dessen Bewegung, Neigung oder Drehung erfasst und zur Steuerung verwendet werden kann. Besonders konsequent und spielerisch wurden diese neuen Möglichkeiten von der Unterhaltungsindustrie aufgegriffen. So gibt es beispielsweise Spiele (unter anderem für Apples iPhone und iPod touch), bei denen ein Fahrzeug über die Neigung des Geräts gesteuert wird, wobei das Gerät beidhändig gefasst gleichsam als „Lenkrad“ dient (siehe Abbildung 8, Quelle: Homepage der Firma Apple[33]).

Abbildung 8: Verwendung des Beschleunigungssensors zur Steuerung eines Autorennspiels. Das Gerät selbst wird hierbei zum Lenkrad-Ersatz.

Darüber hinaus finden sich weitere Funktionen innerhalb der normalen Verwendung solcher Systeme, die über Bewegung oder Drehung des Geräts gesteuert werden.

33 Der Screenshot für Abbildung 8 wurde einem der hier verfügbaren Videobeispiele entnommen: http://www.apple.com/de/ipodtouch/games-and-apps/
Weitere Informationen zum Beschleunigungssensor auch unter: http://www.apple.com/de/ipodtouch/what-is/gaming-device.html

So ist es mittlerweile fast zu einer herstellerübergreifenden Konvention geworden, dass eine Drehung des Geräts ins Querformat im Texteingabemodus eine vollwertige QWERTZ-Tastatur quer über die Breite des gesamten Displays erscheinen lässt.

Bisher ist der Einsatz dieser Interaktionsform jedoch noch eingeschränkt, die vorhandenen Möglichkeiten werden kaum ausgeschöpft. Oakley und Park betonen die Notwendigkeit, mobile Interaktionsformen von Grund auf neu zu entwickeln und dem mobilen Kontext anzupassen. Eine aussichtsreiche Komponente hierfür sehen sie in der hier thematisierten Technik, die sie als *motion input* bezeichnen wie folgt definieren: „the act of moving a handheld device in free space in order to issue commands or specify parameters“ [Oakley 2009, S. 516]. In besonderer Weise betonen sie das Potential dieser Technik für eine *eyes-free interaction*, womit sie eine Interaktion ohne Unterstützung durch grafisches Feedback bezeichnen [Oakley 2009, S. 516]. In einer empirischen Untersuchung betrachten sie die Leistungsfähigkeit eines exemplarischen Systems, an welches sie hohe Maßstäben anlegen: „Firstly, it must be usable by novices when they encounter it for the first time. Secondly, it must support rapid, eyes-free use by experienced expert users. Thirdly, it must enable novices to seamlessly develop into experts through nothing more than extended use of the system“ [ebenda]. Trotz dieser Anforderungen gestalteten sich die Ergebnisse dreier durchgeführter Studien sehr vielversprechend. Es konnte gezeigt werden: „that the system can be rapidly learnt“ und zugleich „can be used effectively eyes free“ [ebenda, S. 515]. Die Autoren resümieren: „Together, these three studies rigorously validate the system design and suggest promising new directions for handheld motion-based Uis“ [ebenda].

Ein weiterer neuer Ansatz sieht die Verwendung der Geräterückseite als zusätzliche Eingebefläche vor, auf der – wie auf dem Touchscreen der Vorderseite – Eingaben mit den Fingern erfolgen können. Wobbrock et al. bemerken hierzu: „The idea is obvious enough when one observes the natural placement of the fingers on the back of a device when the device is being held“ [Wobbrock 2008, S. 860]. In einer vergleichenden Studie erzielten sie außergewöhnlich vielversprechende Ergebnisse für die Interaktion auf der Geräterückseite über alle Bedingungen (Studie 1): „For both one and two hands, the index finger performed surprisingly well, especially on the back of the device; in many cases, it had lower movement times and error rates than the

thumb on the front“ [Wobbrock 2008, S. 871]. Die Verwendung der Geräterückseite bietet hierbei zusätzliche Vorteile, da auf diese Weise das Display durch den steuernden Finger nicht verdeckt wird. Zudem muss die natürliche Handhaltung nicht verändert werden [vgl. ebenda, S. 872]. Während bei einhändiger Bedienung eines Geräts mit Touchscreen nur der Daumen für Interaktionen auf der Vorderseite des Geräts zur Verfügung steht, kann auf der Rückseite in selbiger Haltung der Zeigefinger verwendet werden, der in der vorgestellten Studie hinsichtlich Geschwindigkeit und Fehlerrate überlegene Performance bietet.

Mobile Endgeräte bieten somit vielfältige eigenständige und neuartige Interaktionsmöglichkeiten, die hinsichtlich des Usability Engineering berücksichtigt werden sollten. Zum einen liegt hier zweifellos ein großes Potential zur Optimierung der Usability mobiler Geräte. Zum anderen besteht in diesen neuen Möglichkeiten eine Herausforderung für Usability-Experten, da hier noch wenige Erfahrungen existieren und somit umfangreicher Forschungbedarf besteht, beispielsweise um neue angemessene Heuristiken und Empfehlungen für zukünftige Generationen mobiler User Interfaces zu entwickeln.

6.1.4 CONTEXT-AWARENESS ERWÄGEN

Die Nutzung mobiler Geräte ist gekennzeichnet von der Vielfalt an Nutzungskontexten, mit denen jeweils unterschiedliche Anforderungen an Nutzer und Systeminteraktion einhergehen. Als Reaktion hierauf könnten mobile Geräte mit entsprechenden Modi ausgestattet werden, welche die jeweiligen Anforderungen berücksichtigen. Die Wahl des geeigneten Modus könnte hierbei durch den Nutzer selbst getroffen werden. Alternativ könnten jedoch auch die Geräte in die Lage versetzt werden, den aktuellen Nutzungskontext zu erschießen – eine Fähigkeit, die als *Context-Awareness* bezeichnet wird. Auf dieser Grundlage kann ein System automatisch jenen Modus aktivieren, der den aktuellen kontextuellen Gegebenheiten entspricht.

Moderne Geräte, insbesondere Smartphones, verfügen bereits über eine Reihe von Sensoren und Funktionen, die zum Zwecke der Context-Awareness eingesetzt wer-

den könnten (Lage- und Beschleunigungssensoren, Umgebungslichtsensoren, Ermittlung der Umgebungslautstärke über eingebautes Mikrofon, Positionsbestimmung über GPS, etc.). So ließe sich beispielsweise in Bewegung ein reduzierter Anzeigemodus verwenden, der weiterführende Informationen, zusätzliche Funktionen, Konfigurationsoptionen und dergleichen ausblendet und auf der somit gewonnenen Displayfläche die zentralen Informationen größer und kontrastreicher darstellt. Auf diese Weise könnten relevante Inhalte auch in Bewegung mit einem raschen Blick erfasst werden kann. Eine solche Reduktion auf das für die aktuelle Situation Notwendige ließe sich ferner auch mit Verweis auf den Gestaltungsgrundsatz der *Aufgabenangemessenheit* begründen. Ähnlich könnten Menüs auf die zentralen kontextrelevanten Optionen reduziert werden, die entsprechend größer dargestellt werden. Auf diese Weise wären sie leichter zu erfassen, einzelne Menüeinträge wären zudem einfacher anzuwählen, also in Bewegung leichter zu treffen. Insbesondere Geräte mit Touchscreen können von derartigen Modi in besonders flexibler Weise profitieren, da hier auch die aktiven Eingabeflächen (die Analogie zu mechanischen Tasten) in Position und Größe nach Bedarf frei variierbar sind.

Im Rahmen dieser Überlegungen kommt es zu einem grundsätzlichen Konflikt mit Nielsens Empfehlung: „Avoid Modes“ [Nielsen 1993, S. 146ff.]. Im Bereich mobiler Geräte sind Modi jedoch häufig anzutreffen und – aufgrund begrenzter Ein- und Ausgabefläche und hoher Funktionsvielfalt auf engem Raum – kaum vermeidbar. So wechselt man beispielsweise bei mobilen Browsern in einen Texteingabe-Modus, um eine Ziel-URL einzugeben. Bei Smartphones mit Touchscreen ist ein Wechsel in einen alphabetischen Tastaturmodus möglich, der durch Drehung des Geräts häufig in einen weiteren Modus überführt werden kann, der eine quer angeordnete Tastatur über die gesamte Displayfläche darstellt. Darüber hinaus gibt es grundsätzlich vielfältige Modi für spezielle Funktionen, beispielsweise Foto-, Video- oder Medienwiedergabemodi. Insofern kann davon ausgegangen werden, dass Nutzer mobiler Geräte mit der Verwendung von Modi grundsätzlich vertraut sind.

Ein automatischer Moduswechsel auf Veranlassung des Systems könnte für den Nutzer jedoch potentiell problematisch sein, nicht zuletzt, da ein solches Verhalten zunächst unerwartet ist und der Regel der *Steuerbarkeit* zu widersprechen scheint. Im realen Nutzungskontext könnten dennoch die Vorteile einer solchen Funktion

deutlich überwiegen, sofern bei Konzeption und Implementierung die Aspekte *Selbstbeschreibungsfähigkeit* und *Steuerbarkeit* berücksichtigt werden. Dies bedeutet: das System muss jederzeit deutlich darüber Auskunft geben, in welchem Modus es sich gerade befindet und der Nutzer muss jederzeit die Möglichkeit haben, den aktuellen Modus manuell zu deaktivieren bzw. zu wechseln sowie die Automatik der Context-Awareness generell abzuschalten.

Wie häufig in Fragen der Usability gilt es, Vor- und Nachteile sorgsam gegeneinander abzuwägen. Eine Design-Entscheidung erfolgt idealer Weise auf Basis valider Evaluationsergebnisse. Insbesondere erzielbare Performancegewinne und subjektive Beurteilung repräsentativer Nutzer sollten hierbei Berücksichtigung finden.

6.1.5 WIEDERAUFNAHME ERLEICHTERN

Da häufige Unterbrechungen die Nutzung mobiler Geräte kennzeichnen, sind (retrieval-)cues für möglichst leichten Wiedereinstieg in eine Aufgabe nach einer Unterbrechung vermutlich hilfreich (vgl. Abschnitt 5.5). Dies gilt insbesondere bei komplexeren Aktionen, die mehrere Handlungen in sequentieller Abfolge erfordern. Als solche cues scheinen Informationen über die gerade ausgeführte Aktion, die aktuelle Position innerhalb der Handlungssequenz und die noch ausstehenden Handlungsschritte geeignet. Diese Informationen sollten bei entsprechenden Aktionen entweder grundsätzlich dargestellt oder auf einfache Weise eingeblendet werden, um dem Nutzer einen raschen Überblick zu verschaffen und die Wiederaufnahme der abgebrochenen Handlung bzw. Handlungskette zu erleichtern. Die Gestaltung einer solchen Information kann sich eventuell am Check-out eines Warenkorbs orientieren, wie er aus dem Web-Bereich bereits als Konvention vertraut ist. Hierbei wird ebenfalls eine gerade ausgeführte Handlungssequenz sowie die aktuelle Position innerhalb dieser Sequenz übersichtlich veranschaulicht.

Die grundsätzliche Problematik bei einer analogen Umsetzung im Bereich mobiler Geräte besteht in der geringen verfügbaren Bildschirmfläche für diese zusätzliche Information. Bei Touchscreen-Modellen mit vollflächigem Bildschirm ist eine Umset-

zung tendenziell leichter realisierbar. In anderen Fällen stellt wie angedeutet das Einblenden bei Bedarf als Reaktion auf eine entsprechende Nutzeraktion eventuell einen praktikablen Lösungsansatz dar. Auch hinsichtlich der retrieval-cues gilt es letztlich, Vor- und Nachteile einer solchen Funktion sorgfältig abzuwägen. Die Grundlage der Entscheidung sollten empirische Testergebnisse bilden, mittels derer beispielsweise überprüft werden kann, ob und inwieweit Performancegewinne (z.B. reduzierte Bearbeitungszeiten) durch die Zurverfügungstellung von retrieval-cues erzielbar sind. Auch die subjektive Bewertung der Nutzer sollte in den Entscheidungsprozess einfließen.

6.2 IMPLIKATIONEN FÜR DIE USABILITY EVALUATION

Der Nutzungskontext mobiler Endgeräte stellt klassische Methoden der Usability Evaluation – insbesondere die nutzerzentrierten Methoden – vor Herausforderungen. Die Evaluation im kontrollierten Labor weicht erheblich von der realen Nutzungssituation ab, was wie die externe Validität der Ergebnisse wie dargestellt beeinträchtigt. Feldstudien unter weitgehend realen Bedingungen werfen jedoch Fragen hinsichtlich der internen Validität auf, da die Experimentalbedingungen kaum kontrolliert werden können. Ein möglicher Ausweg aus dieser Situation besteht in der Konzeption neuer Herangehensweisen, die eine Simulation realer Nutzungssituationen im Labor zum Ziel haben. Ergänzend zu etablierten Methoden können derartige Simulationsstudien dazu beitragen, die Realitätsnähe der Testsituation zu erhöhen, ohne auf kontrollierbare Laborbedingungen und wiederholbare Testdurchläufe verzichten zu müssen.

Die denkbaren Simulationsszenarien unterscheiden sich hierbei in ihrer Komplexität und somit in der Anzahl der Eigenschaften des Nutzungskontextes, die nachgebildet werden. Die einfachste Variante besteht in der Simulation einzelner Teilaspekte der mobilen Nutzung, beispielsweise der freihändigen Nutzung in Bewegung. Hierzu gilt es zunächst, statische und stark eingeschränkte Testsituationen (z.B. Testgerät fixiert oder freihändige Nutzung, jedoch sitzend am Tisch und Gerät in definiertem Bereich unterhalb einer Kamera zu halten, etc.) aufzubrechen und die Nutzung freihändig

und in Bewegung zu ermöglichen. Zur kontrollierbaren Umsetzung von Bewegung im begrenzten Laborraum können hierzu Hilfsmittel wie beispielsweise Laufbänder eingesetzt werden. Dieser Ansatz berücksichtigt nicht die kognitiven Anforderungen des realen Nutzungskontextes. Es können jedoch bereits ergänzende Erkenntnisse über die motorischen und koordinatorischen Konsequenzen der Nutzung in Bewegung gewonnen werden. So ist für Geräte mit einer Bedienung per Touchscreen beispielsweise denkbar, dass Menüeinträge an einem Tisch sitzend problemlos bedient werden können, während sie sich in Bewegung als zu klein erweisen und häufig die versehentliche Auswahl benachbarter Items zur Folge haben.

Kjeldskov et al. entwickeln diesen Ansatz weiter und stellen ein Framework vor, in dem sie verschiedene Abstufungen von Bewegungsintensität und kognitiver Belastung durch Bewegungskoordination berücksichtigen [vgl. Kjeldskov 2004, S. 604].

The five configurations based on motion and need for navigation

Body motion	Attention needed to navigate	
	None	Conscious
None	1. Sitting at a table or standing	n/a
Constant	2. Walking on a treadmill with constant speed or stepping on a stepping machine	4. Walking at constant speed on a track that is changing because obstructions are moved
Varying	3. Walking on a treadmill with varying speed	5. Walking at varying speed on a track that is changing because obstructions are moved

Abbildung 9: Fünf unterschiedene Bewegungskonditionen nach Kjeldskov et al., Abbildung entnommen aus [Kjeldskov 2004, S. 604].

Zusätzlich untersuchten Kjeldskov et al. eine Simulationsbedingung, die auf die Induzierung kognitiver Belastung im Rahmen des Dual-Task Paradigma abzielte. Die Probenden mussten in dieser Experimentalbedingung parallel zur Interaktion mit dem Mobilgerät eine kognitiv aufwendige Zweitaufgabe bewältigen. Diese wurde mittels eines Computerspiels operationalisiert, welches über Bewegungen gesteuert wird: „When playing this game, the user steps on different active areas on a 'dance mat' according to sequences shown on a monitor“ [Kjeldskov 2004, S. 608].

Während die grundsätzlich angestellten Überlegungen interessante Anregungen für die Praxis der Usability-Evaluation bieten, sind an der durchgeführten Studie einige Punkte zu kritisieren. Zum einen wurden kaum signifikante Ergebnisse erzielt, so dass keine verlässliche Unterscheidung zwischen Szenarien ermöglicht wird. Rückschlüsse über konkrete Eigenschaften und entsprechende Eignung der Situationen sind so kaum möglich. Die Problematik mangelnder Signifikanz liegt hierbei voraussichtlich in hohem Maße an der geringen Gruppengröße der Probanden pro Bedingung (4-6 Personen pro Bedingung, vgl. [Kjeldskov 2004, S. 618]). Ferner verglichen die Autoren in beiden Teilexperimenten (Bewegungssituationen und Computerspielbedingung) die im Labor erzielten Werte mit Ergebnissen einer Feldstudie mit Probanden in einer Straßensituation. Diese diente als Vergleichswert, den es im Laborkontext zu replizieren galt. Eine möglichst starke Annäherung der Laborsituation an diesen Wert wurde somit zum Gütekriterium der Beurteilung der Simulationsbedingung. Während eine solche Baseline zunächst als sinnvoll betrachtet werden kann, erscheint die Vergleichbarkeit der Ergebnisse aus Feld- und Laborstudie ebenso wie die tatsächliche Realitätsnähe der Feldsituation aufgrund des Vorgehens [vgl. ebenda, S. 616f.] jedoch zweifelhaft.

Abschließend ist auch die für den Simulationsfall gewählte Aufgabe nicht repräsentativ für den realen Nutzungskontext. Die Beobachtung vorgegebener Bewegungssequenzen und Reaktion bzw. Nachahmung auf gewisse Cues entspricht zweifellos nicht der üblichen freien und selbstbestimmten Bewegung im öffentlichen Raum. Aus der Aufgabenstellung resultieren ferner mit hoher Wahrscheinlichkeit abweichende kognitive Anforderungen. Grundsätzlich kann dennoch für den Einsatz einer solchen Technik argumentiert werden, da immerhin hinsichtlich der Tatsache, *dass* eine parallele kognitive Belastung vorliegt eine Annäherung an die reale Nutzungssituation erfolgt. Diese Überlegung liegt beispielsweise auch Studien zugrunde, die mobile Geräte in Autosimulatoren testeten, den Probanden also durch paralleles Steuern eines Fahrzeuges einer kognitiven Belastung aussetzten. Obgleich diese Situation sich im realen Nutzungskontext aus pragmatischen ebenso wie aus juristischen Erwägungen nicht empfiehlt, führen Kjeldskov et al. zu den Vorzügen eines solchen Vorgehens aus: „it facilitates the evaluation of mobile system use while simultaneously engaged in a demanding cognitive activity. The simulator-based technique

facilitated both quantitative and qualitative data to be collected and a huge number of test sessions to be conducted within limited time frames“ [Kjeldskov 2004, S. 602].

Ein solches Vorgehen mag bereits eine sinnvolle Ergänzung zu statischer Evaluation darstellen und ist dem grundsätzlichen Verzicht auf eine Berücksichtigung des Kontextes eventuell vorzuziehen. Da unterschiedliche Informationsverabeitungsstrukturen sich jedoch wie eigenständige kognitive Ressourcen verhalten können, scheint es dennoch angebracht, mehr Sorgfalt auf eine realitätsnahe Simulation der tatsächlichen kognitiven Belastungen zu verwenden. Kjeldskov et al. mögen bei der Auswahl ihrer Simulationsaufgabe auch von der Tatsache geleitet worden sein, dass die Interaktion mit der Simulationsumgebung – im vorliegenden Falle dem Computerspiel – über Bewegungen erfolgte und somit motorischer Natur war. Aus Perspektive der kognitiven Ressourcen ist dies jedoch keineswegs zwingend notwendig. Der Aspekt der Bewegung sollte wie oben ausgeführt im Hinblick auf die motorische Steuerung – beispielsweise bei der Auswahl von Menüitems, etc. – berücksichtigt werden. Hinsichtlich der kognitiven Belastung ist jedoch nicht die motorische Komponente entscheidend – das Gehen kann als so weitgehend automatisiert angesehen werden, dass es keinerlei kognitive Belastung hervorruft. Es ist vielmehr die Orientierung und Bewegungskoordination im Raum, die kognitive Anforderungen mit sich bringt. Sie erfordert beispielsweise die Bestimmung umgebender Objekte und ihrer Eigenschaften, z.B. als Hindernisse auf einer Route, ggf. die Berücksichtigung ihrer Geschwindigkeit, Geschwindigkeitsänderungen, Bewegungsvektoren, sowie die resultierende Ableitung ihrer zukünftigen Position im Raum zur Koordination eigener Bewegungen und vorausschauender Kollisionsvermeidung und viele weitere Faktoren mehr. Es erscheint folglich als sinnvolle Option, sich in der Simulation auf diese Aspekte des Bewegungskontextes zu konzentrieren und beispielsweise realitätsnahe belebte Straßenszenen, Flughafenhallen oder Bahnsteige zu simulieren, durch die der Proband navigieren muss. Eine große Herausforderung besteht in diesem Fall in der einer geeigneten Schnittstelle, über die der Proband mit der Simulation interagiert.

Abschließend ist darauf hinzuweisen, dass aus der tendenziell geringen externen Validität statischer Laborstudien keineswegs folgt, dass diese klassische Herangehensweise für mobile Geräte keine Anwendung finden sollte. Viele Usability-Probleme sind grundsätzlicher Natur (z.B. Konsistenz, verständliche Benennung, Farben

und Kontraste, Metaphern, etc.) und können durchaus in stationärer Evaluation ermittelt werden. Hierzu können gerade auch expertenzentrierte Methoden einen signifikanten Beitrag leisten. Jene Aspekte eines Systems, die dem Nutzer bereits unter statischen Bedingungen, Abschirmung von störenden Einflüssen und voller Konzentration Probleme bereiten, lassen naturgemäß auch im Einsatz unter weniger günstigen realen Bedingungen Komplikationen erwarten. Aufgrund der vollen Konzentration auf das zu testende System überrascht es daher auch nicht, wenn bei Kjeldskov et al. in der statischen Laborbedingung (am Tisch sitzend) signifikant mehr Usability-Probleme aufgedeckt werden, als in den verglichenen alternativen Bedingungen [vgl. Kjeldskov 2004, S. 610].

Es gilt jedoch zu bedenken, dass zwar alle im Labor ermittelten Usability-Probleme mit hoher Wahrscheinlichkeit auch in der realen Nutzung relevant sind – dies jedoch im Umkehrschluss nicht bedeutet, dass alle in der Nutzungspraxis relevanten Probleme auch im Labor zu ermitteln sind. Aspekte, die unter idealen Laborbedingungen unkritisch sind, können sich im realen Nutzungskontext durchaus als problematisch erweisen. Weitere Probleme ergeben sich womöglich erst aus Interferenzen zwischen den Anforderungen der Systeminteraktion und den Anforderungen des Nutzungskontextes. Um auch derartige Problemkategorien im Evaluationsprozess berücksichtigen zu können und resultierende Usability-Probleme aufzudecken, empfiehlt es sich, zusätzlich Anstrengungen zu unternehmen, um möglichst realitätsnahe Testbedingungen zu schaffen.

7 FAZIT

Mobile Endgeräte stellen Usability-Experten vor vielfältige Herausforderungen. Während im Bereich der Desktop-Systeme seit Etablierung des WIMP-Paradigmas in den achtziger Jahren kaum grundlegende Neuerungen zu beobachten sind, entfaltet der Sektor mobiler Geräte eine vergleichsweise atemberaubende Dynamik. Neuartige Interaktionsformen entstehen, für die keine Erfahrungswerte und Heuristiken existieren, technische Einschränkungen und Begrenzungen gilt es zu überwinden und der mobile Nutzungskontext bringt ganz eigene Anforderungen für Nutzer und Systeme mit sich.

Für eine Auseinandersetzung mit dieser dynamischen Entwicklung erweist sich eine integrierte Betrachtungsweise von Geräteeigenschaften, Nutzungskontext und den resultierenden kognitiven Anforderungen grundsätzlich als fruchtbarer Ansatz. Es lassen sich sowohl konkrete Designempfehlungen als auch Impulse für neue Evaluationsverfahren aus entsprechenden theoretischen Überlegungen ableiten. Anwendbarkeit, Praxistauglichkeit und realen Beitrag zur Optimierung der Usability mobiler Endgeräte bedürfen zunächst zweifellos einer kritischen Überprüfung auf Basis experimenteller Studien. Grundsätzlich jedoch verknüpft die vorgestellte Herangehensweise interdisziplinäre Perspektiven in produktiver Weise und unterstützt die Entwicklung innovativer Konzepte. Sie kann somit als Ausgangspunkt für weiterführende Arbeiten dienen und in Kombination mit empirischen Methoden somit zur Optimierung von Usability Engineering und Evaluation beitragen.

8 ANHANG

8.1 ÜBERSICHTSKARTEN ZUR ABDECKUNG UMTS/HSDPA

Den nachstehenden Übersichtskarten der beiden großen Netzanbieter T-Mobile und O2 ist die aktuelle flächenmäßige Abdeckung mit breitbandigen mobilen Datenverbindungen (UMTS sowie UMTS mit HSDPA) zu entnehmen, dunkle Flächen kennzeichnen hierbei jeweils die Verfügbarkeit. (Stand: März 2010)

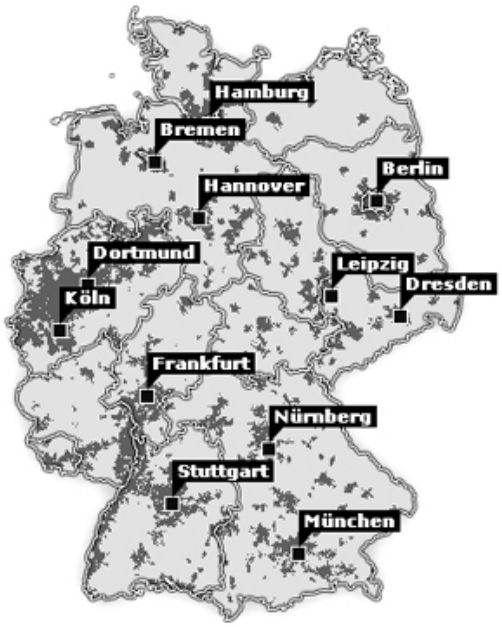

Abbildung 10: UMTS/HSDPA-Verfügbarkeit im Netz von O_2

Abbildung 11: UMTS/HSDPA-Verfügbarkeit im Netz von T-Mobile

Quellen: http://www.o2online.de/nw/support/mobilfunk/netz/index.html
http://www.t-mobile.de/funkversorgung/inland/0,12418,15400-_,00.html
(letzter Aufruf jeweils 17.03.2010)

9 QUELLENVERZEICHNIS

Anderson, John: *Kognitive Psychologie*. Berlin: Spektrum, 2001.

Bristow, Huw; Baber, Chris et al.: *Defining and evaluating context for wearable computing.* International Journal of Human-Computer Studies. Volume 60, Issues 5-6, May 2004, Seiten 798-819.

Dahm, Markus: *Grundlagen der Mensch-Computer-Interaktion*. München: Pearson, 2006.

Dey, Anind; Abowd, Gregory et al.: *A Conceptual Framework and a Toolkit for Supporting the Rapid Prototyping of Context-Aware Applications*. Human-Computer Interaction. Volume 16, Issues 2-4, December 2001, Seiten 97-166.

Faulkner, Laura: *Beyond the five-user assumption: Benefits of increased sample sizes in usability testing*. Behavior Research Methods, Instruments, & Computers. Volume 35, Issue 3, August 2003, Seiten 379-383.

Jameson, Anthony; Klöckner, Kerstin: *User Multitasking with Multimodal Systems*. In: Spoken Multimodal Human-Computer Dialogue in Mobile Environments. Minker, Wolfgang; Bühler, Dirk et al. (Hrsg.). Dordrecht: Springer, 2005. Seiten 349-377.

Kjeldskov, Jesper; Stage, Jan: *New techniques for usability evaluation of mobile systems.* International Journal of Human-Computer Studies. Volume 60, Issues 5-6, May 2004, Seiten 599-620.

Lavie, Nilli; Hirst, Aleksandra et al.: *Load Theory of Selective Attention and Cognitive Control.* Journal of Experimental Psychology: General. Volume 133, Issue 3, September 2004, Seiten 339-354.

Leung, Rock; MacLean, Karon et al.: *Evaluation of haptically augmented touchscreen gui elements under cognitive load*. In: International Conference on Multimodal Interfaces. Proceedings of the 9th international conference on Multimodal interfaces, Nagoya, Aichi, Japan. New York: ACM, 2007. Seiten 374-381.

Nielsen, Jakob: *Usability Engineering*. San Francisco: Morgan Kaufmann, 1993.

Nielsen, Jakob: *Enhancing the Explanatory Power of Usability Heuristics.* In: Celebrating interdependence: CHI '94 conference proceedings. Adelson, Beth (Hrsg.). New York: ACM Press, 1994, Seiten 152 – 158.

Oakley, Ian; Park, Junseok: *Motion marking menus: An eyes-free approach to motion input for handheld devices*. International Journal of Human-Computer Studies. Volume 67, Issue 6, June 2009, Seiten 515-532.

Oulasvirta, Antti; Saariluoma, Pertti: *Surviving task interruptions: Investigating the implications of long-term working memory theory.* International Journal of Human-Computer Studies. Volume 64, Issue 10, October 2006, Seiten 941-961.

Oviatt, Sharon; Coulston, Rachel et al.: *When Do We Interact Multimodally? Cognitive Load and Multimodal Communication Patterns*. In: Sixth International Conference on Multimodal Interfaces (ICMI 2004). New York: ACM, 2004. S. 129-136.

Pollmann, Stefan: *Allgemeine Psychologie*. München: Reinhardt, 2008.

Sarodnick, Florian; Brau, Henning: *Methoden der Usability Evaluation. Wissenschaftliche Grundlagen und praktische Anwendung*. Bern: Huber, 2006.

Weiss, Scott: *Handheld Usability*. Chichester: Wiley, 2002.

Wickens, Christopher: *Multiple resources and performance prediction.* Theoretical Issues in Ergonomics Science. Volume 3, Issue 2, April 2002, Seiten 159-177.

Wobbrock, Jacob; Myers, Brad: *The performance of hand postures in front- and back-of-device interaction for mobile computing*. International Journal of Human-Computer Studies. Volume 66, Issue 12, December 2008, Seiten 857-875.

York, Judy; Pendharkar, Parag: *Human-computer interaction issues for mobile computing in a variable work context.* International Journal of Human-Computer Studies. Volume 60, Issues 5-6, May 2004, Seiten 771-797.

Zhang, Dongsong; Adipat, Boonlit: *Challenges, Methodologies, and Issues in the Usability Testing of Mobile Applications.* International Journal Of Human–Computer Interaction. Volume 18, Issue 3, 2005, Seiten 293-308.

ONLINEQUELLEN

Cisco Visual Networking Index: Forecast and Methodology, 2008–2013. Cisco Systems, 2009.
http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns827/white_paper_c11-481360.pdf
(verfügbar als PDF, letzter Aufruf 17.03.2010)

Gens, Frank: *IDC Predictions 2010: Recovery and Transformation.* IDC, 2009.
http://cdn.idc.com/research/predictions10/downloads/Top10Predictions.pdf
(verfügbar als PDF, letzter Aufruf 17.03.2010)

Morgan Stanley: *The Mobile Internet Report.* Dezember 2009.
http://www.morganstanley.com/institutional/techresearch/pdfs/2SETUP_12142009 _RI.pdf
(verfügbar als PDF, letzter Aufruf 17.03.2010)

Ramsay, Marc; Nielsen, Jakob: *WAP Usability Déjà Vu: 1994 All Over Again - Report from a Field Study in London, Fall 2000.* Nielsen Norman Group, 2000.
http://www.nngroup.com/reports/wap/WAP_usability.pdf
(verfügbar als PDF, letzter Aufruf 17.03.2010)

Sigurdson, Jon: *WAP OFF - Origin, Failure and Future.* Entwurf für Teilnahme an Japanese-European Technology Studies (JETS), 2001.
http://www.telecomvisions.com/articles/pdf/wap-off.pdf
(verfügbar als PDF, letzter Aufruf 17.03.2010)

van Dam, Andries: *Post-WIMP User Interfaces.* Communications of the ACM, Volume 40, Issue 2, February 1997, Seiten 63-67.
http://portal.acm.org/citation.cfm?id=253708
(verfügbar als PDF, letzter Aufruf 17.03.2010)

Vuolle, Maiju; Aula, Anne et al.: *Identifying Usability and Productivity Dimensions for Measuring the Success of Mobile Business Services.* Advances in Human-Computer Interaction, Volume 2008, Article ID 680159, 9 Seiten.
http://downloads.hindawi.com/journals/ahci/2008/680159.pdf
(verfügbar als PDF, letzter Aufruf 17.03.2010)

VERWENDETE WEBSEITEN

(letzter Aufruf sofern nicht anders angegeben: 25.01.2010)

Albrecht, Georg (Ansprechpartner): App Store von Apple übertrifft 3 Milliarden Downloads. Apple Pressemitteilung, 5.01.2010.
http://www.apple.com/de/pr/library/2010/01/05appstore.html

Bangeman, Eric: *Smartphones are leaving PDAs in the dust*. ars technica, 10.10.2006.
http://arstechnica.com/gadgets/news/2006/10/7944.ars

Canibol, Hans; Schuster, J.: *Pager haben ausgequixt*. FOCUS. Nr. 26, 1998.
http://www.focus.de/finanzen/news/telekommunikation-pager-haben-ausgequixt_aid_170722.html

Lipsman, Andrew (Ansprechpartner): *Touchscreen Mobile Phone Adoption Grows at Blistering Pace in U.S. During Past Year.* comScore Pressemitteilung, 3.11.2009. (letzter Aufruf 21.03.2010)
http://www.mediametrix.com/index.php/layout/set/popup/Press_Events/Press_Releases/2009/11/Touchscreen_Mobile_Phone_Adoption_Grows_at_Blistering_Pace_in_U.S._During_Past_Year

Nielsen, Jakob: *Mobile Usability.* 20.07.2009 (letzter Aufruf 21.03.2010)
http://www.useit.com/alertbox/mobile-usability.html

Sauter, Michael (Ansprechpartner): *Europaweit anhaltender Umsatzrückgang bei Technischen Gebrauchsgütern*. GfK TEMAX Pressemitteilung, 28.8.2009.
http://www.gfk.com/group/press_information/press_releases/004548/index.de.html

Schaad, Maurice (Ansprechpartner): *Jeder vierte Mobilfunknutzer plant Kauf eines neuen Handys*. BITKOM Pressemitteilung, 29.10.2009.
http://www.bitkom.org/de/presse/62013_61548.aspx

Schaad, Maurice (Ansprechpartner): *IT- und Telekommunikations-Trends 2010*. BITKOM Pressemitteilung, 13.01.2010.
http://www.bitkom.org/62121_62116.aspx

Schaad, Maurice (Ansprechpartner): More than four billion mobile phone users worldwide. European Information Technology Observatory (EITO) Pressemitteilung, 7.8.2009.
http://www.eito.com/pressinformation_20090807.htm

amazon – Produkteigenschaften *kindle*
http://www.amazon.com/Kindle-Wireless-Reading-Display-Generation/dp/B0015T963C/

Apple – Produkteigenschaften *iPod touch*
http://www.apple.com/de/ipodtouch/features/

Apple – technische Daten *iPad*
http://www.apple.com/de/ipad/specs/

Apple – Übersichtsseite *iPad App Store*
http://www.apple.com/de/ipad/app-store/

Apple – Übersichtsseite *Spiele und Apps* zum iPod touch (letzter Aufruf 22.03.2010)
http://www.apple.com/de/ipodtouch/games-and-apps/

Apple – Übersichtsseite *Was ist der iPod touch?* (letzter Aufruf 22.03.2010)
http://www.apple.com/de/ipodtouch/what-is/gaming-device.html

Archos – Übersichtsseite *Internet Tablets*
http://www.archos.com/products/imt/index.html

mobile media and communication lab (m2c-lab) der FH Aachen: *Usability Evaluation von Smartphones.*
http://www.m2c-lab.fh-aachen.de/smartphoneusability/content/einleitung.php

NTT docomo – Operating Data zu den *i-Mode Kundenzahlen*
http://www.nttdocomo.com/about/operating/

O_2 – *Netzausbau Deutschland* (letzter Aufruf 17.03.2010)
http://www.o2online.de/nw/support/mobilfunk/netz/index.html

Opera – Übersichtsseite unterstützter Geräte des mobilen Browsers mit Schwerpunkt auf dem *Nintendo DSi* Browser und Links zu weiterführenden Informationen
http://www.opera.com/devices/

T-Mobile – *Funkversorgung Deutschland* (letzter Aufruf 17.03.2010)
http://www.t-mobile.de/funkversorgung/inland/0,12418,15400-_,00.html

United States Patent & Trademark Office, Patent Application: Multi-touch display screen with localized tactile feedback, Apple Inc. (letzter Aufruf 21.03.2010)
http://appft.uspto.gov/netacgi/nph-Parser?Sect1=PTO2&Sect2=HITOFF&p=1&u=%2Fnetahtml%2FPTO%2Fsearch-bool.html&r=1&f=G&l=50&co1=AND&d=PG01&s1=20090167704&OS=20090167704&RS=20090167704

Wikipedia – Eintrag zum *Universal Mobile Telecommunications System (UMTS)*
http://de.wikipedia.org/wiki/Universal_Mobile_Telecommunications_System

Wikipedia – Eintrag zu Sonys *PlayStation Portable*
http://en.wikipedia.org/wiki/PlayStation_Portable

10 DANKSAGUNGEN

An erster Stelle danke ich meinen Eltern für ihre beständige, bedingungslose und liebevolle Unterstützung in allen wichtigen Dingen, die umfassende Förderung, die mir vom ersten Tag auf diesem Planeten zuteil wurde, ihre Geduld, ihr Vertrauen und nicht zuletzt die Gewissheit, mich jederzeit und in jeder Weise auf diese beiden Menschen verlassen zu können. All dies ist weit mehr als irgendjemand erwarten darf. Meines Glückes bin ich mir täglich bewusst, doch mangelt es im Alltag mitunter an Anlässen, dies zu artikulieren. Dabei sollte keine Gelegenheit ungenutzt verstreichen. Ich danke Euch. Von Herzen. Für alles.

Ein weiterer Dank gebührt meiner Liebsten Sandra für ihre Geduld und Unterstützung in anstrengenden und mitunter schwierigen Wochen. Neuhaus, Caro-Kaffee und Buddenbrooks-Lesungen halfen, den Anschein geistiger Gesundheit zu wahren. Deine Gegenwart, gemeinsame Fluchten in phantastische Refugien und die gelegentlichen Ausbrüche noch mehr. Vielen Dank.

Abschließend möchte ich Frau Ilse Harms danken – nicht nur für die Übernahme der Betreuung dieser Arbeit und ihre Förderung während meiner langjährigen Studienzeit. Vielmehr hat sie – ohne dies zu ahnen – durch ihre menschliche, beherzte Art, ihren unermüdlichen Einsatz, durch Offenheit, Neugier und klare Worte in manch dunkler Stunde nicht unerheblich dazu beigetragen, mich immer wieder mit meinem Studium zu versöhnen. Dafür an dieser Stelle und in aller Form: herzlichen Dank.