enSektion Medizinische Psychologie

Prof. Dr. Harald C. Traue

Psychophysiologische Parameter der Software-Usability: Experimentelle Studie zur Korrelation zwischen Eye-Tracking-Parametern und Software-Designdomänen

Dissertation zur Erlangung des Doktorgrades der Humanbiologie

der Medizinischen Fakultät der Universität Ulm

(Dr. biol. hum.)

Gertrud Grünwied

aus Vilsbiburg

2007

Amtierender Dekan:	**Prof. Dr. Klaus-Michael Debatin**
1. Berichterstatter:	**Prof. Dr. Harald C. Traue**
2. Berichterstatter:	**Prof. Dr. Michael Weber**
Tag der Promotion:	16. Juli 2007

Inhaltsverzeichnis

Abkürzungsverzeichnis 5

1 Einleitung 6

1.1 Software-Usability 6
1.1.1 Usability-Anforderungen 6
1.1.2 Usability-Probleme 8
1.1.3 Experten- und benutzerorientierte Messverfahren 10
1.2 Eye-Tracking 13
1.2.1 Visuelle Aufmerksamkeit 13
1.2.2 Augen- und Blickbewegungen 15
1.2.3 Eye-Tracking-Anwendungsoptionen 17
1.2.4 Eye-Tracking-Parameter 19
1.2.5 Gütekriterien von Eye-Tracking-Tests 20
1.3 Fragestellung 24

2 Methode und Experiment 26

2.1 Kombination eines Referenzmodells mit Eye-Tracking 26
2.1.1 Referenzmodell WebSCORE 26
2.1.2 Eye-Tracking-relevante Designdomänen 27
2.2 Eye-Tracking-Parameter für Usability 30
2.2.1 Usability-relevante Augen- und Blickbewegungen 30
2.2.2 Usability-relevante Eye-Tracking-Parameter 31
2.3 Experiment 34
2.3.1 Gegenstand der Mensch-Computer-Interaktion 34
2.3.2 Testumgebung und Situation 36
2.3.3 Testaufgabe 37
2.3.4 Testpersonen 41
2.3.5 Hardware und Software für Eye-Tracking 42
2.3.6 Testablauf 44
2.4 Analyse und Berechnungen 46
2.4.1 Analyse der Rohdaten 46
2.4.2 Bewertung der Reliabilität der Tests 50

3 Ergebnisse 54

3.1 Leistungsdaten zur Effizienzmessung 54
3.1.1 Bearbeitungsdauer bzw. Blickpfaddauer 54
3.1.2 Weitere Leistungsdaten zur Effizienzmessung 56
3.2 Experten-Bewertung der Problemstellen 58
3.2.1 Teilaufgaben 58
3.2.2 Problemstellen je Oberflächen-Komponente 61
3.2.3 Gewichtung der Problemstellen je Oberflächen-Komponente 66
3.2.4 Oberflächen-Komponenten korreliert mit Designdomänen 68
3.3 Ergebnisse pro Parameter 70
3.3.1 Suchzeit 70
3.3.2 Kumulierte Übergangszeit 73

3.3.3 Relative Häufigkeit der Fixationen von Objekten ... 77
3.3.4 Richtungsänderungen des Blickpfades ... 84
3.3.5 Übergangshäufigkeiten ... 92
3.3.6 Übergangsdichte ... 95
3.4 Zusammenfassung der Ergebnisse ... 96
3.4.1 Designdomänen, Oberflächen-Komponenten und Parameter ... 96
3.4.2 Aussagekraft der Parameter ... 97
3.5 Ergebnisse des retrospektiven Protokolls mit Nutzerzitaten ... 98

4 Diskussion und Ausblick ... 100

4.1 Diskussion ... 100
4.1.1 Auswahl der relevanten Parameter und deren Wertebereiche ... 100
4.1.2 Blickpfaddauer ... 102
4.1.3 Suchzeit ... 103
4.1.4 Kumulierte Übergangszeit ... 105
4.1.5 Relative Häufigkeit der Fixationen von Objekten ... 107
4.1.6 Übergangshäufigkeiten und Übergangsdichte ... 109
4.1.7 Richtungsänderungen ... 110
4.2 Kritische Bewertung der Gütekriterien ... 112
4.3 Ausblick ... 114

5 Zusammenfassung ... 115

6 Literatur ... 117

7 Danksagung ... 121

8 Anhang: Bildschirmseiten ... 122

Abkürzungsverzeichnis

(B)	Benutzeraktion
(S)	Systemreaktion
°/s	Winkelgrad pro Sekunde
AOI	Area of Interest
AVI	Audio Video Interleave
cm	Zentimeter
DIN	Deutsches Institut für Normung
E-Commerce	Electronic Commerce
EN	European Nations
E-Procurement	Electronic Procurement
E-Shops	Electronic Shops
EVD	Event Data
FXD	Fixation Data
HTML	HyperText Markup Language
Hz	Hertz
IAO	Institut für Arbeitswirtschaft und Organisation
ISO	International Organization for Standardization
kg	Kilogramm
Max	Maximum
Min	Minimum
MIT	Massachusetts Institute of Technology
ms	Millisekunden
PC	Personal Computer
s	Sekunden
TFT	Thin File Transistor
URL	Unified Resource Locator
USB	Universal Serial Bus
VGA	Video Graphics Array

1 Einleitung

Gegenstand der vorliegenden Arbeit ist die Interpretierbarkeit von Eye-Tracking-Parametern zur Evaluation von Software-Usability. Kapitel 1 führt in die Themen Software-Usability und Eye-Tracking ein und endet mit der Formulierung der Fragestellung.

1.1 Software-Usability

1.1.1 Usability-Anforderungen

„Usability“ (dt. Gebrauchstauglichkeit) ist eine Produkteigenschaft und weiterhin ein Ziel der Ergonomie-Wissenschaft.

> *„Gebrauchstauglichkeit (Usability) ist das Ausmaß, in dem ein Produkt durch bestimmte Benutzer in einem bestimmten Nutzungskontext genutzt werden kann, um bestimmte Ziele effektiv, effizient und zufrieden stellend zu erreichen.“ (DIN-EN-ISO 1999, S. 4)*

Die Kriterien, um zu beurteilen, ob eine Gestaltungslösung gebrauchstauglich ist, sind:

- Der Benutzer muss seine Aufgaben mit dem Dialogsystem grundsätzlich erledigen können (Effektivität des Dialogsystems bei der Aufgabe). Ein Messverfahren hierzu ist die erfolgreiche Bearbeitung einer Aufgabe.
- Der Benutzer muss die Aufgaben mit möglichst wenig Zeitaufwand und mit möglichst wenigen Arbeitsschritten erledigen können (Effizienz des Dialogsystems bei der Aufgabe). Geeignete Messgrößen hierzu sind die Anzahl benötigter Mausklicks und die Bearbeitungszeit.
- Der Benutzer muss mit der Benutzung des Dialogsystems für seine Aufgabe zufrieden sein (Zufriedenheit des Benutzers). Ein Prüfverfahren hierzu sind Fragebogen-Bewertungen oder offene Interviews.

Auf dem Trendtag 2006 in Hamburg forderte Prof. Wippermann „Simplexity“ als die neue „effiziente Klugheit“ (Wippermann 2006). Nach Wippermann wird in Zukunft die „*gefühlte Einfachheit der High-Tech-Komplexität vorgezogen werden*“. In diesem Zusammenhang betont er auch die Einfachheit der Bedien-Oberflächen durch intuitive Benutzerführung und logische Konsistenz.

Eine Software besitzt eine niedrige Gebrauchstauglichkeit, wenn sich der Anwender an vorgegebene Abläufe adaptieren muss. „*Der Be-Nutzer wurde Be-Diener der Arbeits-*

mittel“ (Sarodnick u. Brau 2006, S. 11). Was macht ein System für den Benutzer leicht bedienbar und leicht zu verstehen? Folgende Faktoren sind für die Benutzerfreundlichkeit von zentraler Bedeutung (Nielsen 1993, Wickens et al. 1998):

- **Aufgabenangemessenheit:** Das System sollte funktional sein, d. h. es sollte hinsichtlich der Komplexität den erforderlichen Aufgaben angepasst sein.
- **Lernförderlichkeit:** Das System soll dem Benutzer ohne Anspruch auf große Vor- und Fachkenntnisse einen raschen und einfachen Gebrauch ermöglichen.
- **Selbstbeschreibungsfähigkeit:** Das System soll den Benutzer sowohl über mögliche Funktionen als auch über den aktuellen Systemstatus informieren, d. h. es soll klar erkennbar sein, ob das System den Input des Benutzers erhalten hat und sein Befehl ausgeführt wird.
- **Erwartungskonformität:** Das System soll dem Benutzer vertraut sein, d. h. möglichst auf dem Benutzer bereits bekannten Konzepten, Handlungsabläufen etc. aufbauen. Zudem sollte es auch systemintern konsistent sein: gleiche Gestaltungselemente sollten durchgängig mit denselben Funktionen gekoppelt sein.
- **Individualisierbarkeit:** Der Benutzer soll die Mensch-System-Interaktion und die Darstellung von Informationen ändern können, um sie an seine individuellen Bedürfnisse anzupassen. Ein Beispiel sind die verschiedenen Ansichten des Datei-Explorers als Liste, als Symbole oder als Miniaturbilder.
- **Fehlertoleranz:** Das System soll Fehler erkennen, verständlich zurückmelden und konstruktive sowie rasche Fehlerbehebungen anbieten.
- **Steuerbarkeit:** Der Benutzer soll das System an seine Bedürfnisse anpassen können. Beispielsweise sollen Menüs eine individuelle Auswahl ermöglichen.

Anforderung der Intuitivität

Eine Anforderung ist, dass die Systeme intuitiv zu gestalten sind. Das Lexikon der Psychologie definiert „Intuition“:

> *„Die plötzliche Eingebung; unmittelbares, ganzheitliches Erkennen oder Erfahren eines Gegenstandes ohne Reflexion; intuitives Denken steht im Gegensatz zum schlussfolgernden, diskursiven Denken.“ (MEYERS-Lexikon-online 2007)*

Übertragen auf Ergonomie und Design von Software bedeutet intuitive Benutzbarkeit, die Softwarefunktionen erschließen sich einem Nutzer ohne aufwändige Reflexionen und Interpretationen. Dieser Erkenntnisprozess ist zumeist ohne Bewusstheit.
Von den Grundsätzen der Dialoggestaltung lassen sich die Anforderungen „Selbstbeschreibungsfähigkeit“ und „Lernförderlichkeit“ auf intuitive Benutzung abbilden (Glaser et al. 2005). Selbstbeschreibungsfähigkeit überschneidet sich mit intuitiver Bedienung, da diese durch Analogien oder durch Aufzeigen eines Kontextes das vorhandene Wissen des Benutzers aufruft. Das fehlende Wissen erschließt sich dem Benutzer, ohne dass er diesen Prozess bewusst wahrnimmt. Dabei handelt es sich um einen Lernprozess, der bei intuitiven Oberflächen minimalen Aufwand erfordern sollte.

1.1.2 Usability-Probleme

Bevor man sich mit Usability-Evaluationsmethoden beschäftigt, ist zu klären, wie Usability-Probleme definiert sind. Denn nicht alle Schwierigkeiten, die während einer Systemnutzung auftreten, sind auf den Benutzer und auf die Bedienung zurückzuführen.

> *„Ein Usability-Problem liegt vor, wenn Aspekte eines Systems es Nutzern mit hinreichender Domänenerfahrung unangenehm, ineffizient, beschwerlich oder unmöglich machen, in einem typischen Anwendungskontext die Ziele zu erreichen, für deren Erreichung das System erstellt wurde.“ (Sarodnick u. Brau 2006, S. 22)*

Dieses Zitat zeigt auf, wann ein „echtes“ Usability-Problem vorliegt und wann nicht. Demnach sind zum einen Personen-immanente Benutzungsprobleme, wie z. B. die fehlende Schulung in eine Branchensoftware, keine personen-übergreifende Bedienprobleme. Zum anderen kann auch nur dann von einem Usability-Problem gesprochen werden, wenn das System tatsächlich für den bestimmungsgemäßen Zweck verwendet wird und dabei die Anforderungen an Effektivität, Effizienz und Zufriedenheit nicht erfüllt. Um Systeme zu entwickeln, die möglichst wenige Usability-Probleme beim Anwender verursachen, gilt es, die klassische Produkt-Entwicklung um die ergonomische Gestaltung zu ergänzen. Software-Entwickler besitzen zwar die Kompetenz der technischen Umsetzung und erreichen damit funktionierende Systeme. Die technische Realisierung allein ist jedoch nicht ausreichend, um eine gute Gebrauchstauglichkeit zu erzeugen. Sarodnick definiert im obigen Zitat, Anwender haben nicht aufgrund individueller Eigenschaften Nutzungsprobleme, sondern die Probleme sind vielmehr in der Software begründet. Da sie beim Großteil der Anwender auftreten, liegt es nahe, sich mit der allgemeinen menschlichen Wahrnehmung und der Kognition bei Arbeitstätigkeiten zu

befassen. Erfahrungen aus der Laborpraxis haben gezeigt, *„dass ca. 90 % aller Nutzungsprobleme vor der Ausführung einer Handlung stattfinden“* (Schimpfky 2003, S. 16).

Die Interaktion zwischen Anwender und Computer läuft in mehreren Phasen ab (Abbildung 1). Vor der Ausführung liegen die zwei Phasen „Orientierung“ und „Verarbeitung“. Während der Orientierungsphase nehmen Betrachter zunächst einfache visuelle Reize auf einer Oberfläche wahr. Die Orientierungsphase entspricht der groben syntaktischen Wahrnehmung der Ist-Situation. Ein Beispiel ist die Wahrnehmung von Schriftzügen auf der Bildschirmoberfläche. Dabei wird der gesehene Text im Einzelnen (noch) nicht erkannt. Es wird nur seine reine Existenz wahrgenommen und eine Vorauswahl der im Folgenden näher zu entschlüsselnden Informationen getroffen. Die Verarbeitungsphase ist der Interpretation des Wahrgenommenen gewidmet. Den syntaktischen Objekten wird eine Bedeutung verliehen. Als Ergebnis dieses Prozesses wird eine Aktivität geplant. Die semantische Analyse zum Beispiel der Menü-Beschriftungen sowie die Entscheidung, auf einen bestimmten Menüpunkt zu klicken, ist dieser Vorbereitungsphase zuzuordnen. Die Ausführungsphase entspricht dann der tatsächlichen Benutzeraktivität. Die geplante Aktivität setzt der Anwender über seine Hände sowie über die Eingabegeräte (Tastatur, Maus) um. Beispielsweise klickt der Nutzer einen Menüpunkt an oder bewegt den Cursor über ein Bild oder erfasst Buchstaben in einem Feld.

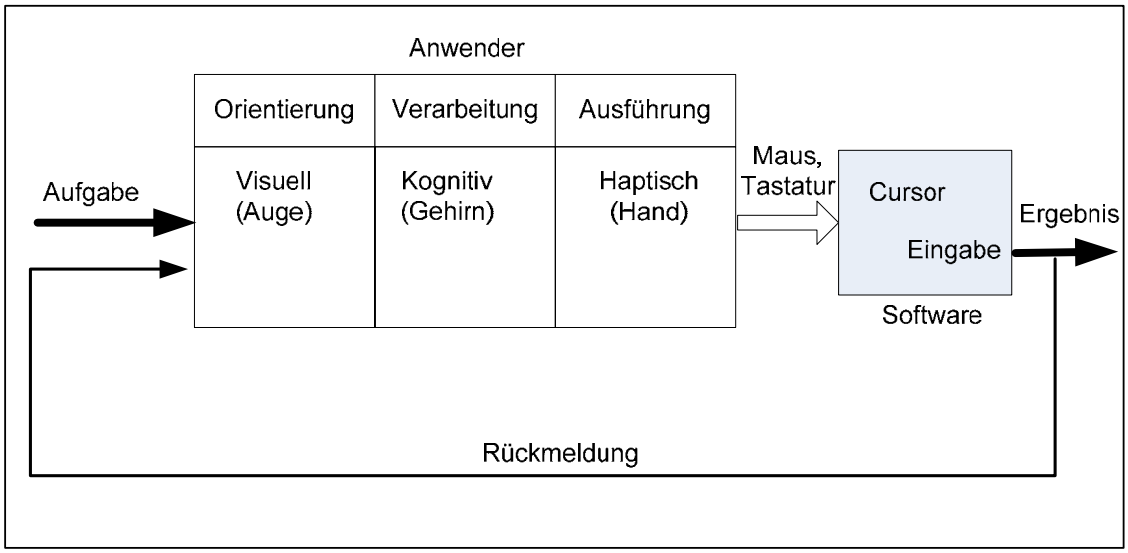

Abbildung 1: Phasenmodell bei der Anwender-Software-Interaktion

Geis et al. erweitern das Phasenmodell um eine zusätzliche Bewertungsphase, die zeitlich nach der Ausführung und zeitgleich zur nächsten Orientierungsphase stattfindet (Geis et al. 2004). Bei der Bewertung prüft der Anwender über die Rückmeldung des Systems, ob das gewünschte Resultat auch tatsächlich eingetreten ist.
Krömker spricht im Zusammenhang der verschiedenen Phasen eines Dialogschrittes davon, *„dass die Software-Bedienung nur ein Minimum an Zeit und Beanspruchung erfordern solle“* (Krömker 1999, S. 30).

1.1.3 Experten- und benutzerorientierte Messverfahren

Messverfahren zur Software-Usability prüfen, wie effektiv, effizient und zufriedenstellend Anwender ihre Aufgaben lösen können. In Abhängigkeit von der Testergruppe (Experten oder „echte“ Benutzer) ist eine der grundlegenden Unterscheidung die zwischen den benutzer- und den expertenorientierten Verfahren (Heuer 2003, Krings 2003, Schweibenz u. Thissen 2003).

Bei den expertenorientierten Verfahren halten sich Experten häufig an Normen, Gesetze, Verordnungen, Heuristiken und Problemtopologien der Usability-Evaluation. Eine zentrale Stellung unter den Normen besitzt die Ergonomie-Normenreihe DIN EN ISO 9241 „Ergonomie der Mensch-System-Interaktion“. Teile der Norm werden seit 2006 reformiert und modernisiert. Der Vorteil eines Experten-Gutachtens ist, dass sein Umfang variabel gestaltet werden kann. Die Bandbreite der Untersuchung reicht vom Aufdecken der wichtigsten Problemstellen der Software bis hin zur Zerlegung einer Seite in einzelne Oberflächenelemente. Schwächen von Experten-Reviews zeigte eine schweizerische Studie auf (Blume et al. 2005). Sie ergab, dass die Analysen verschiedener unabhängig voneinander arbeitender Experten sich widersprechende Aussagen lieferten. Diesen Verfahren fehle der realistische Kontext, so lautet ein weiterer Schluss der Autoren. Der expertenorientierte Ansatz für sich allein ist demnach unzureichend und kann lediglich am Anfang eines Entwicklungs- oder Qualitätssicherungsprozesses stehen, der in seinem weiteren Verlauf auch Benutzertests einschließt. Eine Richtlinie für das Ineinandergreifen von expertenbasierten Evaluationen und Benutzerbefragungen stellt beispielsweise der ERGOguide dar (Dzida et al. 1993).

Bei den benutzerorientierten Tests (Abbildung 2) werden Benutzer vor, während oder nach der Produktnutzung beobachtet. Das Nutzerverhalten kann entweder „Offline“

oder „Online“ getestet werden: „Offline“ bedeutet, eine Befragung findet zu einem späteren Zeitpunkt statt als die Produktnutzung, bei „Online“-Tests laufen Datenerhebung und Benutzung gleichzeitig ab. „Online“-Verfahren lassen sich wiederum in quantitative und qualitative Verfahren unterscheiden. Bei qualitativen Verfahren können Fragen nach dem „Warum“ mit Hilfe von Verbalisierungen der Testpersonen geklärt werden. Weit verbreitet und am häufigsten praktiziert unter den „Online“-Verfahren ist das „Laute Denken“. Bei dieser Methode sprechen die Probanden während der Erledigung einer Aufgabe ihre Gedanken und Emotionen laut aus und werden dabei mit Video aufgezeichnet und von Experten beobachtet. Eye-Tracking zählt zu den quantitativen „Online“-Verfahren unter den Benutzertests.

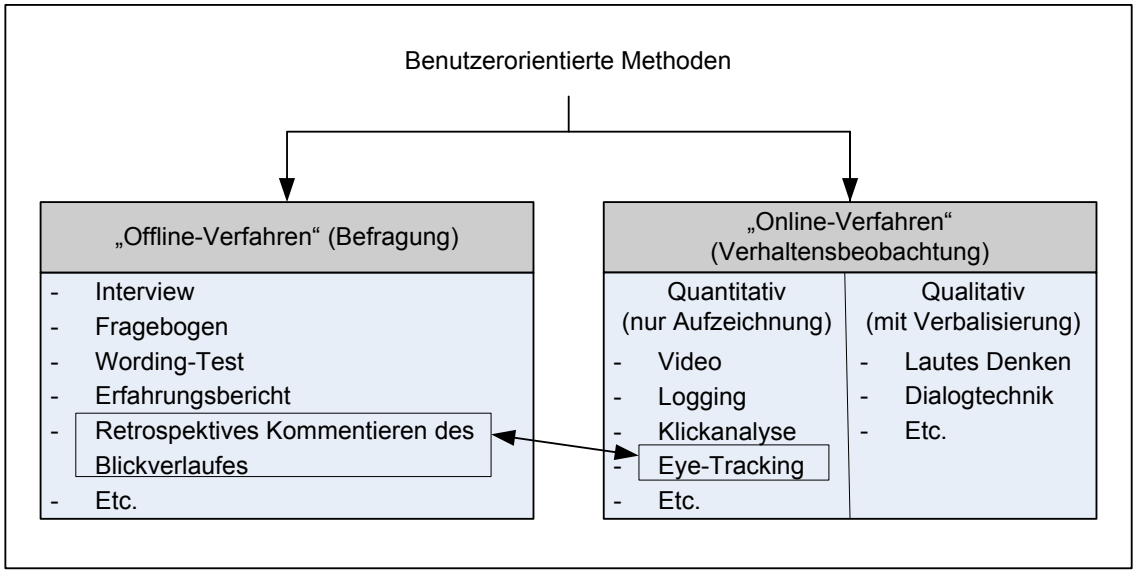

Abbildung 2: Benutzerorientierte Evaluations-Methoden

Es ist ein in der empirischen Forschung bekanntes Problem, dass oftmals die Gründe, die Testpersonen in einem Nachgespräch angeben, nicht die tatsächlich handlungsführenden Gründe waren. Vielmehr sind sich die Testpersonen der relativ künstlichen Testsituation bewusst und versuchen, bestimmten Erwartungen an Intelligenz und Sozialverhalten zu entsprechen. Dadurch sind bei den Verbalisierungs-Methoden Verzerrungen der Ergebnisse nicht auszuschließen. Es hängt wesentlich vom Versuchsleiter ab, die Situation zu erklären und zu entspannen.

Die gebundene Reproduktion, d. h. das retrospektive laute Kommentieren, gestützt durch den eigenen Blickverlauf, bildet eine Mischung aus „Lautem Denken“ und Eye-Tracking. Als der Benutzung zeitlich nachgelagertes „Offline“-Verfahren zählt es nach

diesem Schema zu der Gruppe der Befragungen. Eine Methodenforschung zum retrospektiven Protokoll findet sich in (Eger et al. 2005).

In vier Studien der Humboldt Universität Berlin wurde nachgewiesen (Schiessl et al. 2003), dass mit Eye-Tracking eine höhere Validität der Daten im Vergleich zu den Selbstberichten der Testpersonen erreicht werden kann. Die introspektiven Methoden „Lautes Denken“ und Fragebögen lieferten sogar gegensätzlich Ergebnisse zu Eye-Tracking-Daten. Frauen lesen Textinformationen auf einer Webseite sehr sorgfältig und Männer erfassen eher bildorientiert die Informationen, so lautet ein Ergebnis von Eye-Tracking-Studien. Bei der Befragung gaben Männer jedoch an, sie brauchen keine Bilder auf Webseiten. Im Abgleich mit anderen Studien stimmt jedoch die Interpretation, dass Frauen die Texte einer Webseite lesen und Männer eher nicht. Die Berliner Studie zeigte damit, dass die Anwendung von Eye-Tracking eine validere Testmethode in Usability-Tests ist als konventionelle Methoden, da Eye-Tracking einen tiefen Einblick in die Ursachen von Benutzungsproblemen gibt. Zu den Usability-Messungen, die von den Testpersonen bewusst reflektiert und gesteuert werden, zählen unter anderem der retrospektive Erfahrungsbericht, Fragebögen, „Lautes Denken“ und die Klickanalyse. Der alleinige Einsatz dieser Methoden führt jedoch zu einem Validitätsproblem, da Aufmerksamkeitsprozesse nicht ausschließlich von bewusster Kontrolle abhängen.

Bereits im Jahr 1986 wurde die Effektivität einer Methodenkombination an Fallbeispielen nachgewiesen (Moll 1987). Moll wendet mehrere Methoden aus „Online“- und „Offline“-Benutzerverfahren an. Eine Kategorie dient dem Erfassen von subjektiven Kognitionen. Die gewählten Methoden sind dabei das „Laute Denken“, die Videokonfrontation sowie Befragungen. Die andere Kategorie hat die objektive Verhaltensbeobachtung zum Ziel. Eine Methode dazu stellt die Logfile-Registrierung dar. Wichtig sei die Mischung aus subjektiven und objektiven Methoden.

1.2 Eye-Tracking

1.2.1 Visuelle Aufmerksamkeit

Zunächst soll eine Definition der menschlichen Aufmerksamkeit gegeben werden.

> *„Aufmerksamkeit ist definiert als ein Zustand konzentrierter Bewusstheit, begleitet von einer Bereitschaft des zentralen Nervensystems, auf Stimulation zu reagieren. Man kann sich Aufmerksamkeit als eine Brücke vorstellen, über die einige Bestandteile der äußeren Welt – die ausgewählten Aspekte, auf die die Aufmerksamkeit konzentriert ist – in die subjektive Welt des Bewusstseins gebracht werden, so dass Verhaltenskontrolle ermöglicht wird." (Zimbardo 1992, S. 195)*

Auf welche Bestandteile der Außenwelt Menschen ihre Aufmerksamkeit richten, hängt von mehreren äußeren und inneren Faktoren ab. Äußere Reize gehen von etwas Neuem, Unerwartetem oder Auffälligem aus. In Bezug auf Software-Oberflächen geht der Blick, wo immer möglich, auf Farben, Bilder, animierte Elemente oder große (fast bildhafte) Überschriften. Auch unerwartet auftretende Fehlermeldungen ziehen die Aufmerksamkeit unwillkürlich an. Zu den inneren Reizen gehören physiologische Zustände z. B. wie Hunger, der die Aufmerksamkeit auf Lebensmittel lenken lässt.

Nach Pashler gibt es zwei wesentliche, durch wissenschaftliche Experimente robust gemachte, Theorien zur Aufmerksamkeit (Pashler 1997).

- Selektive Aufmerksamkeit (Filtertheorie): Menschen filtern aus der Vielfalt der durch die Sinne einfallenden Informationen die wichtigsten Bereiche aus. Diese Bereiche betrachten sie genauer, andere Bereiche werden ignoriert.
- Geteilte Aufmerksamkeit: Mehrere gleichzeitig eintreffende Informationen über einen oder mehrere Sinne werden parallel wahrgenommen. Die Aufnahmekapazität scheint insbesondere dann unbegrenzt groß zu sein, wenn die Informationen Bekanntes enthalten und nicht bewusst verarbeitet werden müssen.

Heute stützen viele Untersuchungen den Schluss, dass Aufmerksamkeit nicht, wie in der Filtertheorie postuliert, alles andere ignoriert, sondern eher ein graduell abgestufter Vorgang ist. Jeder visuelle Input wird vom Betrachter analysiert, wenn auch nur teilweise und nicht bewusst. Ein Beispiel für nicht bewusste Verarbeitungsprozesse ist, ein Auto zu steuern. Diese Tätigkeit verlangt ständige Anpassung an Reizbedingungen, denen der Fahrer jedoch wenig oder keine bewusste Aufmerksamkeit widmet. Dieses

Phänomen der „Geistesabwesenheit“ ermöglicht es, eine Flut von Informationen parallel zu bewältigen und nicht ein Signal nach dem anderen verarbeiten zu müssen.

Ein in der Literatur häufig zitierter Mechanismus der Aufmerksamkeit ist Posners „spotlight“. Demnach beschreibt Posner (Posner et al. 1980) die visuelle Orientierung mit der Metapher eines Scheinwerfers. Ein Teil des visuellen Feldes wird erhellt, während der Rest im Dunkeln bleibt. Eine Unterscheidung, die Posner trifft, ist dabei für das Verständnis grundlegend notwendig. Posner unterscheidet zwei Aspekte der visuellen Aufmerksamkeit, nämlich die „Orientierung“ (orienting, „where“) und die „Entdeckung“ (detection, „what“). Im Gegensatz zur Detektion, die selektiv, also nur hintereinander und kontextsensitiv erfolgt, funktioniert die Orientierung parallel und umfassend, wie vergleichsweise der Kegel eines Scheinwerfers. Die Orientierung ist demnach nicht zwangsläufig mit dem scharfen Sehen im fovealen Bereich verbunden, sondern eher mit dem peripheren und unscharfen Sehen. Die Orientierung kann somit auch unabhängig von Augenbewegungen funktionieren. Der „Scheinwerfer“ beleuchtet parallel mehrere Teile der Umgebung und die Aufmerksamkeit bestimmt dabei, welche Informationen weiterverarbeitet werden und was verworfen wird. Wenn sich die Aufmerksamkeit einem neuen Gegenstand zuwendet, endet die Verarbeitung des vorherigen und die Verarbeitung des neuen Materials beginnt. Zwischen den Sprüngen der Aufmerksamkeit wird keine Information verarbeitet. Die Orientierung während der Aufmerksamkeit funktioniert parallel und eilt dem neuen Zielort, der Aufdeckung, dabei voraus. Ein Beispiel aus der Linguistik soll dies illustrieren:

> *„Wenn Sie den folgenden Punkt kurz fixieren •, hat Ihre Aufmerksamkeit schon bis zu zwölf Buchstaben rechts vom Punkt vorverarbeitet. Beim Lesen hat der Scheinwerfer also eine stark asymmetrische Form.“ (Scheier u. Heinsen 2003, S. 155)*

Das Erklärungsmodell der parallelen Wahrnehmung dazu ist: sobald ein Wort eindeutig identifiziert ist, d. h. die Syntax erkannt ist, stellt dies die Auslösebedingung für das Weiterspringen des Blickes zum nächsten Wort dar. Die Bedeutung des Wortes und die Einordnung in den Text müssen dabei noch nicht vollkommen klar sein.

In welchem Zusammenhang stehen Aufmerksamkeit und Augenbewegungen, die Gegenstand der Blickmessung sind? Der Blick müsste entsprechend den oben genannten dualen Annahmen ein zyklischer Prozess sein, der sich aus folgenden Schritten zusammensetzt (Duchowski 2003):

1. Der Einstieg in den betrachteten Stimulus ist vorwiegend parallel und vom peripheren Sehen bestimmt. In dieser Initialphase können auffällige Bereiche des Materials bewirken, dass sich die Aufmerksamkeit vorbereitend auf diese Bereiche ausrichtet für eine weitere detaillierte Betrachtung.
2. Die Augen bewegen sich schnell auf die erste auffällige Region hin.
3. Sobald die Augen ihre Zielposition erreicht haben, wird der foveale Bereich auf das Ziel gerichtet und fest in den Blick genommen.

Die Verarbeitung von Informationen kann auch außerhalb des fixierten Objektes stattfinden. Die Annahme, dass das fixierte Objekt auch Gegenstand der kognitiven Verarbeitung ist, trifft damit nur in den meisten Fällen zu. Die Aufmerksamkeit selbst kann zudem innerhalb des fovealen Bereiches verschoben werden (Anderson 2001). Festzuhalten ist, dass der Aufmerksamkeitsscheinwerfer stark mit den Augenbewegungen korrespondiert, dies aber nicht in jedem Fall so sein muss. Die Aufmerksamkeit kann weit ab vom Fixationsort sein oder kann innerhalb des Fixationsfeldes verschoben werden.

1.2.2 Augen- und Blickbewegungen

Das gesamte Blickfeld eines Auges umfasst einen Kegel von etwa 100°, wobei der Bereich scharfen Sehens deutlich kleiner ist. Schärfstes Sehen ist in einem Kegelöffnungswinkel von ca. 2° um den Fixationsort möglich (Fovealer Bereich). Zur Veranschaulichung: der foveale Bereich entspricht der Größe des Daumenfingernagels bei ausgestrecktem Arm oder ca. 2,1 cm Bildschirmfläche bei 60 cm Abstand zwischen Monitor und Benutzer. Die Augen müssen also Bewegungen (Blicksprünge) durchführen, um sie auf den zu verarbeitenden Stimulus im fovealen Bereich zu platzieren. Die Sehschärfe nimmt im parafovealen Bereich, der eine Ausdehnung von 2° bis 10° hat, und dem Bereich des peripheren Sehens außerhalb von ca. 10° stark ab. Das periphere Sehen ermöglicht die Orientierung im Raum sowie die Wahrnehmung von Bewegungen und Helligkeits- und Kontraständerungen.
Psychophysiologische Untersuchungen grenzen Augen- und Blickbewegungen voneinander ab:

> *„Augenbewegungen sind alle Bewegungen des Auges, die allein durch Beobachtung des Auges erfasst und interpretiert werden können. Im Gegensatz dazu werden als Blickbewegungen solche Bewegungen des Auges bezeichnet, die in Verbindung mit den vom Auge aufgenommenen Informationen interpretiert werden.“ (Rötting 2001, S. 14)*

In dieser Arbeit liegt der Schwerpunkt auf den bewussten Blickbewegungen, die mit den Methoden des Eye-Tracking gemessen werden können. Ein Beispiel für die Blickbewegung ist die Fixationsdauer auf einem Wort in einem Text oder auf einem Menüpunkt in einer Software.
Es lassen sich im Wesentlichen drei Gruppen von Augen- und Blickbewegungen unterscheiden. Eine detaillierte Erläuterung im Anschluss ist notwendig, um die in der vorliegenden Untersuchung berücksichtigten Augenbewegungen einzuführen.

A. Ausrichtung der Fovea zu neuen Sehobjekten: Die für die detaillierte Informationsaufnahme notwendigen Augenbewegungen verlaufen nicht langsam, linear und gleichmäßig, sondern sprunghaft mit ständigen Ruhepunkten (Sakkadensprünge und Fixationen). Die raschen Sprünge, Sakkaden genannt, dauern zwischen 15 ms und 100 ms und haben eine Geschwindigkeit von bis zu 700°/s. Die Auslöser für Sakkaden sind mittels des peripheren Sehens für interessant befundene Sehobjekte oder Bewegungen. Sakkaden dienen dazu, die Fovea auf neue Ziele auszurichten und sind darum der visuellen Suche zuzuschreiben. Sakkaden werden von beiden Augen gleichzeitig ausgeführt und sind ballistische Bewegungen, d. h. einmal initiiert, können sie nicht mehr von ihrem programmierten Kurs abgebracht und unterbrochen werden. Die Zeit, die zwischen Erscheinen eines visuellen Stimulus und dem Beginn einer Sakkade vergeht, wird als Sakkadenlatenz bezeichnet. Während einer sakkadischen Augenbewegung ist die Sensitivität für visuellen Input stark herabgesetzt.

B. Augen- und Blickbewegungen, die das Fixieren von Objekten ermöglichen: Zu diesen Bewegungen, die scharfes Sehen und damit eine Informationsaufnahme ermöglichen, zählen Fixationen und Blickfolgebewegungen. Zwischen den Sakkaden treten Ruhepunkte (Fixationsperioden) auf. Während der Fixation ruht das Auge, und der fixierte Reiz, z. B. ein Wort, kann verarbeitet werden. Das fixierte Objekt wird für eine bestimmte Dauer im fovealen Bereich der Retina abgebildet. Ein Sonderfall von Fixationen sind Augenbewegungen, die eintreten, wenn der Stimulus ein bewegtes Objekt ist. Die Augen verfolgen das Objekt mit gleitenden Augen-Folgebewegungen.

C. Mikrobewegungen: Das Auge ruht auch bei einer Fixation nie gänzlich, sondern führt Mikrobewegungen aus. Diese werden im Rahmen dieser Arbeit jedoch aufgrund der schweren Messbarkeit nicht einbezogen.

1.2.3 Eye-Tracking-Anwendungsoptionen

Eine wissenschaftliche Methode zu Benutzertests ist die Registrierung der Blickbewegungen (Eye-Tracking, Okulographie). Dabei werden die Blickbewegungen einer Person aufgezeichnet. Mit Eye-Tracking lässt sich die visuelle Erfassung eines Objektes entschlüsseln. Eye-Tracking-Daten lassen sich mit Hilfe zahlreicher räumlicher und zeitlicher Parameter beschreiben: Fixationsdauer, Fixationsort, kumulierte Fixationen, Blickpfade, um nur einige Beispiele zu nennen. Eye-Tracking wird in verschiedenen Wissenschaftsdisziplinen angewendet, z. B. in den Neurowissenschaften, der Wahrnehmungs- und Kognitionspsychologie, der Marktforschung und der Sicherheitstechnik. Für diese Arbeit sind die Eye-Tracking-Ergebnisse der Leseforschung in der Linguistik und der Usability-Forschung in der Software-Ergonomie von Bedeutung. Aus diesem Grund soll an dieser Stelle kurz erläutert werden, wie die Augenbewegungen beim Lesen von Text auf Papier oder Online und bei der Betrachtung von Webseiten verlaufen.

Blickverlaufsmuster beim Lesen auf Papier oder am Bildschirm

Die Methode Eye-Tracking in der Linguistik soll die Fragen klären, wie sich der Prozess des Lesens vollzieht und was Texte mehr oder weniger verständlich macht. Eye-Tracking liefert eine Messmethode, die es erlaubt zu analysieren, wie Leser einen Text wahrnehmen. Im westlichen Kulturkreis verfolgen die Leser eine Leserichtung, bei der die Sakkaden auf einer Diagonale, von links nach rechts und oben nach unten, verlaufen. Zusätzlich kommt es zu Regressionen. Dabei handelt es sich um Sakkaden, die der Leserichtung entgegengesetzt sind und die zu bereits gelesenem Text zurückführen. Empirische Studien zeigen, wie sich die messbaren Parameter unterscheiden: Mit zunehmender Schwierigkeit des Textes und sinkender Vorhersagbarkeit der visuellen Information werden die Sakkaden kürzer und die Fixationsphasen länger. Auch bei mehrdeutigen und missverständlichen Textinhalten nimmt der Anteil an Regressionen zu.
Weiterhin lässt sich unter anderem über Eye-Tracking ermitteln, wie sich die Lesezeiten beim Lesen auf Papier oder auf dem Bildschirm unterscheiden. Die Lesezeit am Bildschirm wird mit dem Faktor drei länger angegeben als die Lesezeit auf Papier.

Blickverlaufsmuster bei der Betrachtung von Webseiten

Die Kenntnisse über die Blickbewegungen bei Webseiten stammen in erster Linie aus den empirischen Studien zur Web-Usability des Poynter-Instituts (Outing u. Ruel 2004). Die ersten Fixationen der Augen auf einer Homepage finden danach am häufigsten in

der linken oberen Ecke der Seite statt. Der Blick wandert dann von links nach rechts bis zur Mitte der Seite, zurück in die obere Mitte und dann vertikal in der Mitte der Seite nach unten links. Am Ende wandert der Blick unten von links nach rechts und am rechten Rand vertikal von unten nach oben. Das Blickmuster variiert abhängig von verschiedenen Bildschirm-Layouts. Das Muster (Abbildung 3) zeigt eine einfache Repräsentation der häufigsten Blickbewegungsfolge über mehrere Homepage-Designs (Outing u. Ruel 2004).

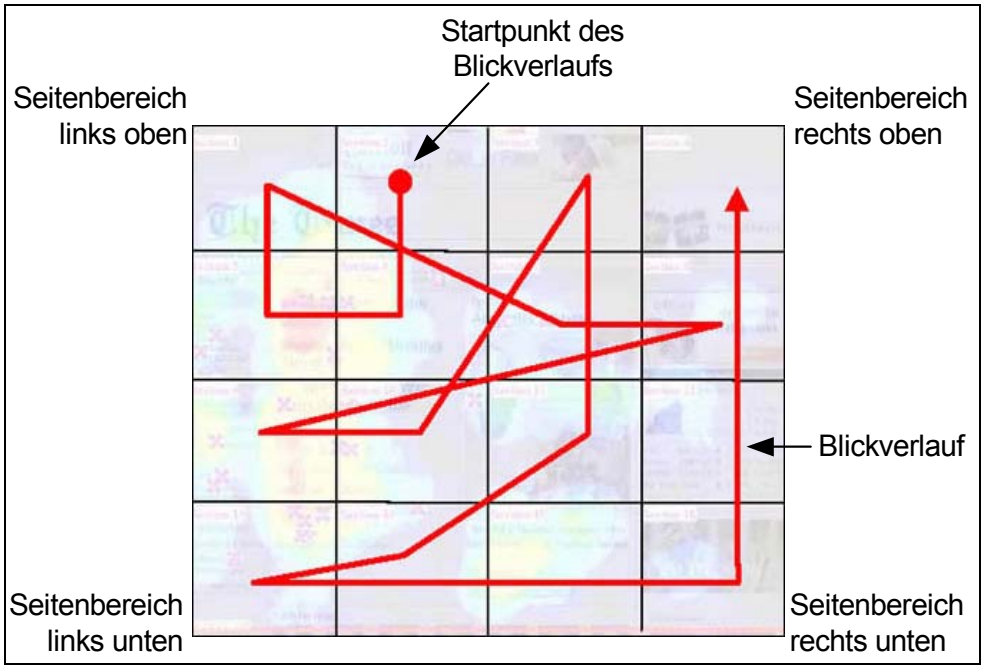

Abbildung 3: Typischer Blickverlauf auf einer Homepage

Vor- und Nachteile von Eye-Tracking-Methoden

Blickmessung in Usability-Studien ermöglicht (Scheier u. Heinsen 2003):

- Identifikation so genannter „toter Zeiten“, in denen der Benutzer keine Aktion (Maus, Tastatur) ausführt, sondern sich mit dem Blick orientiert.
- Messen der Aufmerksamkeits-Verteilung auf bestimmte Bereiche des Bildschirms (Navigation, Werbebanner etc.).
- Messen von Blickpfaden, die sich aus der Abfolge von Fixationen und Sakkaden ergeben.
- Ermitteln von Suchzeiten und Sichtbarmachen der Suchstrategie bei Problemlösungs-Aufgaben.
- Ergänzen qualitativer Methoden, wie zum Beispiel Interviews.

- Minimale unerwünschte Methodeneffekte im Vergleich zum bekannten Versuchsleiter-Effekt bei Nutzerbefragungen.
- Anschauliche Belege (Blickverlaufsvideos und grafische Darstellungen) für die Präsentation der Ergebnisse.

Es gibt jedoch in der Praxis auch Schwierigkeiten mit dieser Messmethode. Die zentralen Probleme liegen sowohl in dem hohen technischen und finanziellen Aufwand der Systeme als auch in der schwierigen Analyse und Interpretation der Daten. Daneben lassen sich weitere Probleme angeben:

- Kleine Anzahl von Probanden aufgrund des hohen Aufwandes möglich.
- Ausfallquote der Probanden (Brillen, Sehfehler etc.).[1]
- Laboreffekte bei invasiven Systemen, die zu einer körperlichen Belastung der Testpersonen führen, z. B. Kopf-basierte Helmsysteme.
- Fehlende Standards zur Bestimmung und Interpretation von Eye-Tracking-Parametern bezogen auf den Anwendungsfall.

Eine detaillierte Beschreibung der methodischen Herausforderungen bei der Eye-Tracking-Methode sowie seine Vor- und Nachteile schildert (Jacob u. Karn 2003).

1.2.4 Eye-Tracking-Parameter

Es gibt eine ganze Reihe von Parametern, die sich zur Erfassung und Interpretation von Blickbewegungen eignen. Eine Durchsicht der gegenwärtigen Forschungsarbeiten zur Blickregistrierung zeigt, dass die erfassten und analysierten Parameter häufig auf den spezifischen Anwendungskontext und die damit einhergehenden technischen Bedingungen abgestimmt sind.

Rötting (2001) bezieht Eye-Tracking auf die Beanspruchung bei arbeitswissenschaftlichen Aufgaben. Er unterscheidet dabei:

- **Parameter der Fixationen**, z. B. Anzahl der Fixationen, Fixationsdauer
- **Parameter der Sakkaden**, z. B. Anzahl der Sakkaden, Sakkadenweiten
- **Parameter des Wechsels aus Fixationen und Sakkaden**, z. B. Blickpfad

[1] Der Hersteller Tobii (www.tobii.com) gibt max. 10 % Ausfallquote für sein Remote-System 1750 an.

Jacob & Karn (2003) geben einen umfassenden Überblick über 24 Usability-Studien und die darin verwendeten Eye-Tracking-Parameter. Aus dieser Menge extrahieren sie die sechs am häufigsten eingesetzten Parameter, wie z. B. die Anzahl der Fixationen. Von der Häufigkeit der Nutzung von Parametern in Studien kann jedoch nicht zwingend abgeleitet werden, sie seien die aussagekräftigsten. Ebenfalls bedeutungsvoll sein können mit diesen Parametern kombinierte Parameter. Ein Beispiel ist der Blickpfad und die Übergangswahrscheinlichkeit zwischen Interessenbereichen auf der Bedienoberfläche, den sogenannten „Areas of Interest".

Schweigert (2003) untersuchte das Blickverhalten von Autofahrern beim Führen eines Fahrzeuges. Schweigert unterscheidet:

- **Zeitbezogene Parameter:** Das sind die wohl am häufigsten benutzten Parameter des Blickverhaltens. Diese zeitlichen Parameter spiegeln die Erfüllung der Effizienz-Forderung für gute Software-Usability wider.
- **Häufigkeits-Parameter:** Diese Parameter beziehen sich auf absolute und relative Häufigkeiten von Fixationen. Wie bei den zeitbezogenen Parametern kann sich die Analyse auf einzelne „Areas of Interest" beziehen oder summiert über alle Bereiche die Gesamtanzahl bestimmt werden. Die Häufigkeits-Parameter können mittels deskriptiver Statistik beschrieben werden.
- **Bezogene Parameter:** Darunter versteht man die Anzahl von Fixationen, die hinsichtlich der Zeit oder des Ortes normiert werden. Ein Beispiel für eine bezogene Größe ist die Anzahl von Fixationen pro Objekt. Auch können Parameter aus der Kombination von Anzahl und Zeitdauer berechnet werden. Beispiel: die Anzahl von Fixationen auf dem Navigationsbereich pro Minute.
- **Sequenzberücksichtigende Parameter:** Sie berücksichtigen die chronologische Reihenfolge des Auftretens von Fixationen. Für viele Fragestellungen ist es aufschlussreich, den Weg des Auges zu verfolgen. Die Parameter können durch Übergangswahrscheinlichkeiten in Markov-Ketten angegeben werden.

1.2.5 Gütekriterien von Eye-Tracking-Tests

Tests und deren Evaluationsmethoden müssen nach den Definitionen von Lienert und Raatz (Lienert u. Raatz 1994) die Testgütekriterien der Objektivität, Reliabilität, Validi-

tät und Praktikabilität (Ökonomie) erfüllen. In diesem Kapitel werden die allgemeinen wissenschaftlichen Gütekriterien auf Usability-Tests mit Blickerfassung übertragen.

Objektivität der Blickregistrierung

Die Objektivität einer Evaluationsmethode bezeichnet das Ausmaß, in dem die Ergebnisse vom Untersuchungsleiter unabhängig sind. Das Gütekriterium der Objektivität von Tests fordert die Unabhängigkeit der Ergebnisse vom Versuchsleiter. Verschiedene Testleiter dürfen keine unterschiedlichen Ergebnisse erzielen. Dies gilt sowohl bei der Durchführung des Tests als auch bei der Analyse und Auswertung des Tests.

Eye-Tracking zählt zur Kategorie der Benutzertests. Die Anwesenheit von Versuchsleitern, die die Benutzer durch die Studie begleiten, ist unvermeidbar. Eine Verletzung der Objektivität, der so genannte Versuchsleitereffekt (Rosenthal 1966), tritt ein, wenn die Leiter Suggestionseinflüsse ausüben oder die Testpersonen aufgrund persönlicher Einstellungen unterschiedlich motiviert sind und dadurch Verzerrungen bei den Daten entstehen. Versuchsleitereffekte sind darum bei Eye-Tracking-Studien nicht gänzlich auszuschließen.

Ausreichende Objektivität kann bei der Blickregistrierung nur erreicht werden (Mussgnug u. Stowasser 2003):

- Wenn die Blickregistrierung standardisiert eingesetzt wird, z. B. durch die Einhaltung von technischen Vorgaben der Eye-Tracking-Versuchsanordnung.
- Wenn die Auswahl der Parameter der Blickregistrierung (Fixationsanzahl, Sakkadenweite etc.) vorab definiert wird.
- Wenn die Testleiter für Eye-Tracking-Untersuchungen qualifiziert sind und ihr Verhalten bei der Testdurchführung standardisiert ist.

Reliabilität der Blickregistrierung

Eine hohe Reliabilität liegt bei Evaluationen vor, wenn die Ergebnisse von mehreren Evaluationen, die mit derselben Methode und auf demselben System durchgeführt wurden, in hohem Maße übereinstimmen. Die Reliabilität bzw. Zuverlässigkeit gibt den Grad der Messgenauigkeit eines Messsystems an. Die Messgenauigkeit wird auch als Präzision bezeichnet.

Bei der Blickregistrierung sind aufgrund des komplexen technischen und methodischen Aufwandes verschiedene Messfehler möglich, die nicht außer Acht bleiben dürfen. Die durch die Apparatur und Methode bedingten möglichen Messfehler entstehen in der Aufzeichnungsphase. Bei der Erfassung der Blickbewegungen sind hierbei Diskrepanzen zwischen tatsächlichen und gemessenen Blickbewegungen möglich. Die Ursachen können sowohl nutzungsbedingt als auch messsystembedingt sein.

Zu den nutzungsbedingten Messfehlern zählen probandenspezifische Eigenschaften. Typisch bei Eye-Tracking-Tests sind z. B. Sehfehler. Weiterhin kann eine ungenaue Kalibrierung die Genauigkeit der erfassten Blickdaten ungünstig beeinflussen. Eine weitere Einschränkung liegt in der Bewegungseinschränkung. Starke Kopf- und Körperbewegungen dürfen Personen selbst bei einem nicht-invasiven, d. h. berührungslosen Eye-Tracking-System nicht ausführen, da sie sich sonst außerhalb des Kamera-Blickwinkels befinden. Messfehler bezüglich der Tiefe des Sehens, so genannte Paralaxe-Fehler, entstehen durch den veränderten Abstand zwischen Augen und Bildschirm im Vergleich zum Abstand während der Kalibrierung.

Auf der anderen Seite können Messfehler in den apparativen Merkmalen des Messsystems begründet liegen. Jedes verwendete Messsystem besitzt Grenzen hinsichtlich seiner zeitlichen und räumlichen Auflösung. Um Sakkadengeschwindigkeiten auswerten zu können, sind videobasierende Auswertesysteme aufgrund der zu geringen Bildwiederholfrequenz von 50 Hz ungeeignet. Sie sind jedoch für die Auswertung von Fixationsdauern geeignet. Die räumliche Auflösung für eine Auswertung von Fixationsorten ist je nach Fragestellung verschieden. Sie sollte für die Untersuchung der Blickbewegungen bei der Computer-Interaktion mindestens 5° betragen.
Eine Rolle spielen zudem die technische Qualität des optischen Systems, die Präzision der Lichtreflexionsmethode, die Genauigkeit der Digitalisierung der Blickdaten und die Datenübertragung der digitalen Blickdaten in die Auswertungssoftware.

Zusammengefasst zeigt die Blickerfassung eine große Anfälligkeit für Messfehler, die während der Aufzeichnung entstehen und ist darum unter Umständen wenig zuverlässig. Wenn Fehler bei Evaluationen eingetreten sind, sind diese von der weiteren Verarbeitung auszuschließen, damit sich die Fehler nicht fortsetzen.

Validität der Blickregistrierung

Unter Validität versteht man die Gültigkeit einer Messmethode und deren Parametern. Die Validität gibt an, wie gut die Methode geeignet ist, genau das zu messen, was sie messen soll.

Bei der Beurteilung der Validität der Blickregistrierung muss berücksichtigt werden, dass bei der Aufzeichnung Messfehler möglich sind und die registrierten Blickpositionen dadurch in vielen Fällen mit unzureichender Zuverlässigkeit erfasst wurden. Bei der weiteren Bearbeitung werden die Daten auf verschiedene Arten verdichtet. Damit unterschiedliche Verhaltensweisen interpretiert werden können, müssen eine Vielzahl von Parametern berechnet werden. Die Parameter können Einzelgrößen wie örtliche oder zeitliche Parameter sein, oder zusammengesetzte Parameter wie der Wechsel von Fixationen und Sakkaden. Die Auswahl der Indikatoren muss auf den Anwendungsfall abgestimmt sein. Zum Beispiel interessieren bei der Untersuchung von Werbeplakaten mittels Eye-Tracking vor allem die ersten unwillkürlichen Fixationen. Bei Software-Interaktionen sind dagegen die späteren Fixationen aufschlussreich, die bewussten Problemlösungsprozessen unterliegen. Eine fehlerhafte Auswahl der Parameter sowie fehlerhafte Verdichtungen lassen sich bei der Vielzahl der Parameter und deren Gewinnungsmöglichkeiten nicht grundsätzlich ausschließen.

Da der zentrale Blickparameter der Fixation lediglich auf Plausibilitätsüberlegungen basiert, stellt dies einen weiteren kritischen Aspekt der Validität dar. Im Kontext von Software-Usability werden die Blickdaten so gefiltert, dass eine Fixation zum Beispiel ab Fixationsradius 30 mm und ab Fixationsdauer 100 ms entsteht. Möchte man dagegen Bildmaterial untersuchen, wird eine Fixation häufig mit einem Radius von 20 mm und einer Dauer von 40 ms definiert.

Praktikabilität der Blickregistrierung

Hardware und Software der Registrierung der Blickbewegungen nehmen demnach eine entscheidende Rolle ein, ob ein Test die genannten Voraussetzungen der Wirtschaftlichkeit erfüllt. Im Falle des Systems Tobii dauert ein Test geringe Zeit, da die Testpersonen sich nur vor den Tobii-Monitor mit der Infrarot-Augenkamera zu setzen brauchen, ohne dass weitere Hardwarekomponenten angeschlossen werden müssen. Zudem genügt bereits eine 5-Punkt-Kalibrierung beider Augen vor der Aufzeichnung.

Über die Ad-hoc-Auswertungsmöglichkeiten können die wichtigsten Analysen wie Blickverlaufsvideo, statische Blickverteilung je Aufzeichnung und Häufigkeitsanalyse sehr einfach erzeugt werden. Jedoch sind Anzahl und Dauer der Auswertungen, die ClearView bei der Auswertung berücksichtigt, begrenzt. Ein spezielles Tool innerhalb von ClearView unterstützt die Auswertung der Blickpunkte bezogen auf einzelne Interessensbereiche in der Software. Ein Ascii-Export der Blick- und Event-Daten gibt den Weg frei für Auswertungen in handelsüblicher Statistik-Software.

1.3 Fragestellung

Aufgrund des aktuellen Forschungsbedarfs zur Nutzung von Eye-Tracking im Bereich der Software-Usability ist die weitere Erforschung wünschenswert. Gleichzeitig sind andere Eye-Tracking-Forschungsgebiete wie in der Linguistik und in der Systemtechnik vorangeschritten, deren Erkenntnisse sich teilweise auch auf Software übertragen lassen. Die technischen Messapparaturen wurden in den letzten Jahren wesentlich verbessert. Ein großer Fortschritt in der Messung von Software-Usability sind einfach handhabbare und platzsparende Eye-Tracking-Systeme. Diese messen die Blickbewegungen der Person berührungslos und helfen Methodeneffekte zu reduzieren. Die Aufzeichnung und die Synchronisation der Blickdaten mit dem Stimulus werden durch entsprechende Software gut unterstützt.

Die zentrale Fragestellung bei der Auswertung und Interpretation der Blickverlaufsdaten besteht darin, einen Zusammenhang zwischen bestimmten Verlaufsmustern und softwarebedingten Usability-Problemen nachzuweisen. Es besteht ein spezieller Untersuchungsbedarf dafür, welche Kriterien im Einzelnen untersucht werden müssen, um Eye-Tracking-Messungen zur Beurteilung von Software-Usability sinnvoll und methodisch nachvollziehbar interpretieren zu können.

Bisher gibt es noch keine allgemein-anerkannten Maßstäbe dafür, welchen Nutzen bestimmte Eye-Tracking-Parameter hinsichtlich Usability-Anforderungen haben. Nach Goldberg (Goldberg u. Wichansky 2003) besteht Forschungsbedarf darin, Korrelationen zwischen anerkannten Usability-Metriken und Eye-Tracking-Daten herzustellen. Hier würden, folgt man Goldberg, Minimal-Standards für die Häufigkeit der Kalibrierung, der Abtastraten, der Fixationsdauern und der Ausstattungen es wesentlich erleichtern, Eye-Tracking-Daten über verschiedene Studien einheitlich zu interpretieren. Auch

Mussgnug & Stowasser (Mussgnug u. Stowasser 2004) sprechen von einer aufwändigen und fachlich schwierigen Auswertung von Eye-Tracking-Daten und geben als Ursache vor allem die unterschiedlichen Auswertungsmöglichkeiten und die Unsicherheiten bei der Interpretation der Blickbewegungen an.

In dieser Arbeit wird ein integrierter Ansatz zur Interpretation der Eye-Tracking-Daten vorgestellt. Im Einzelnen wird untersucht, wie die Metriken der Experten mit den Eye-Tracking-Parametern der Benutzertests korrelieren. Hierzu werden die Kriterienkategorien von Experten-Leitfäden den Eye-Tracking-Parametern und deren Werten gegenübergestellt.

Dabei stehen zwei Fragen im Zentrum:

1. **Fragestellung: Frage nach der Auswahl relevanter Parameter**

   Mit welchen Eye-Tracking-Parametern lässt sich die Gebrauchstauglichkeit einer Software prüfen? Aus welchen Parametern lassen sich Kriterien zur Effizienz einer Software ermitteln? Welche Parameter geben Aufschluss über den Bedienaufwand, das heißt über das Aktivitäts-Niveau und die kognitiven Vorgänge bei der Softwarenutzung?

2. **Fragestellung: Frage nach dem Wertebereich der relevanten Parameter**

   Wie müssen die Werte der relevanten Parameter ausfallen, um einen hohen oder niedrigen Grad von Usability anzuzeigen?

Die Analyse von Blickbewegungen im Kontext der Usability ist damit das zentrale Thema dieser Arbeit. Mit diesem speziellen Ansatz soll diese Arbeit einen Beitrag zur Überprüfung der Gebrauchstauglichkeit von Software leisten.

Die vorliegende Arbeit ist im Landesforschungsschwerpunkt „Der Computer als Dialogpartner - Perzeption und Interaktion in Mehrbenutzerumgebungen“, Universität Ulm, angesiedelt. In diesem Forschungsschwerpunkt werden mehrere Arbeiten zu aktuellen und vielfach auch interdisziplinären Themenbereichen durchgeführt. Das Forschungsvorhaben hat das Ziel, Systemarchitekturen und Algorithmen für eine intelligente Mensch-Computer-Interaktion in Mehrbenutzerumgebungen zu entwickeln und hinsichtlich ihrer psychobiologischen Reaktionen zu bewerten.

2 Methode und Experiment

Dieses Kapitel erklärt die methodische Vorgehensweise bei der Studie. Weiterhin wird beschrieben, wie das Experiment durchgeführt wurde.

2.1 Kombination eines Referenzmodells mit Eye-Tracking

2.1.1 Referenzmodell WebSCORE

WebSCORE ist ein Referenzmodell für die Gestaltung und Evaluation von Webseiten. Das integrierte Methodeninventar WebSCORE wurde vom Fraunhofer Institut für Arbeitswirtschaft und Organisation (IAO) entwickelt und dient dort als intern verwendetes Validierungswerkzeug für Expertenbewertungen. Das Konzept des Modells sowie seine Anwendungsbereiche sind in großen Teilen öffentlich zugänglich (Heidmann 2004, Heidmann u. Ziegler 2002, Hermann 2003, Peissner et al. 2003). Das Modell unterscheidet zwischen einer Design- und einer Evaluationsdomäne. Die Designdomäne umfasst alle Komponenten einer Web-Oberfläche, deren Ergonomie von der Evaluationsdomäne aus betrachtet wird. Zu den Evaluationsdomänen gehören Ziel- und Aufgabenfaktoren, Nutzerfaktoren, sozioökonomische Faktoren sowie technische Rahmenbedingungen. Die konkreten Angaben zu den Evaluationsdomänen müssen in jedem Usability-Test je nach Fragestellung und Nutzungskontext definiert werden. Zudem sieht eine Evaluation nach dem Modell WebSCORE einen standardisierten Ablauf des Usability-Tests vor.

Die Auswahl des WebSCORE-Modells unter einer Vielzahl von Experten-Leitfäden ist darin begründbar, dass das Modell speziell die Benutzerfreundlichkeit von Webseiten und typischen Web-Anwendungen wie E-Commerce-Software in das Zentrum stellt. Andere Ergonomie-Normen setzen einen fensterorientierten Aufbau der Software-Anwendung voraus. Für Desktop-Anwendungen ist dies zutreffend, in Web-Anwendungen sind Fenster unüblich. Eine Website besteht aus mehreren vernetzten Webseiten, wobei jede Seite in Frames (das sind separate Bereiche einer Seite) aufgeteilt ist. Weiterhin handelt es sich bei dem Referenzmodell um eine neue Methode. Sie greift Bestandteile des 2003 erschienenen ISO-Standards 14915 „Software-Ergonomie für Multimedia-Benutzungsschnittstellen“ auf. Interaktive Elemente wie Hyperlinks, Baumstrukturen oder sensitive Bilder sowie Animationen, Audio- und Video-Elemente können mit dieser Methode evaluiert werden. Der zurzeit im Entwurf befindliche Nor-

menteil DIN EN ISO 9241-151 trägt den Titel „Leitlinien zur Gestaltung von Benutzungsschnittstellen für das World Wide Web.“

In dieser Arbeit interessieren insbesondere die Designdomänen von WebSCORE (Abbildung 4), die als Bewertungsschema für die Blickbewegungsanalyse herangezogen werden sollen. Das Schema unterscheidet vier zentrale Designdomänen. Unter „Strategie“ ist zu verstehen, welche strategischen Ziele mit der Web-Anwendung erreicht werden sollen. „Content & Funktionalität“ behandelt die Frage, inwieweit wichtige Inhalte und Funktionalitäten auf der Website vorhanden sind. Die Domäne „Navigation & Interaktion“ fasst die wichtigen Aspekte von klarer Struktur, Erreichbarkeit von Seiten und interaktiven Bedienelementen zusammen. Über „Mediendesign & Präsentation“ lässt sich prüfen, wie benutzungsfreundlich und intuitiv die Gestaltung der Inhalte ist.

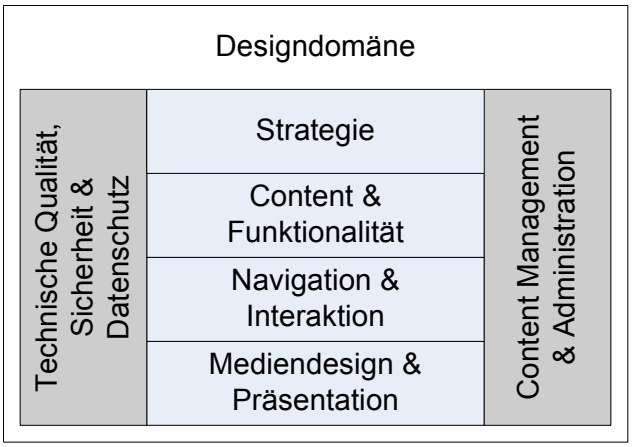

Abbildung 4: WebSCORE Referenzmodell

2.1.2 Eye-Tracking-relevante Designdomänen

Zwei der Domänen, nämlich „Strategie“ sowie „Content & Funktionalität“ lassen sich mit der quantitativen Evaluierungsmethode Eye-Tracking nicht messen. Strategische und inhaltlich-funktionale Aspekte einer Software gehen aus diesem Grund nicht in das Bewertungsschema dieser Arbeit ein.

Die beiden für Eye-Tracking relevanten Domänen sollen hier ausführlicher erläutert werden.

- Navigation & Interaktion
- Mediendesign & Präsentation

Navigation & Interaktion

Die Domäne „Navigation & Interaktion“ ist für die Korrelation mit Eye-Tracking-Daten relevant. Sie enthält wichtige Fragestellungen zur Erreichbarkeit und Strukturierung der Inhalte. WebSCORE unterscheidet dabei folgende Unterbereiche:

- Informationsarchitektur:
  Zielgruppen- und aufgabenbezogene Strukturierung und Benennung der Inhalte
- Navigationskonzept:
  Klare und einheitliche Konzepte unterstützen Navigation und Orientierung
- Interaktive Elemente / Bedienelemente
- Einsatz, interaktives Verhalten, grafische Gestaltung und Positionierung sinnvoll und erwartungskonform

Mediendesign & Präsentation

Die zweite für Blickbewegungsanalysen wichtige Domäne ist „Mediendesign & Präsentation“. Sie stellt wichtige Fragestellungen zur Aufbereitung und Umsetzung der Inhalte bereit. WebSCORE gliedert diese Domäne feiner in:

- Gestaltung der Texte
- Sprachliche Gestaltung, Textumfang pro Screen, Schriftarten und –größen, grafische Strukturierung
- Intuitives Design
- Klare und einheitliche Layout-Templates, Visualisierung struktureller Beziehungen und erwartungskonforme Gestaltung
- Medieneinsatz
- Emotionale Wirkung, Integration unterschiedlicher Medien
- Ästhetische und emotionale Aspekte
- Farbgestaltung, Kommunikationsstil, Trust und Credibility

Eye-Tracking-Designdomänen

Das WebSCORE-Referenzmodell wird für den Rahmen dieser Arbeit leicht angepasst. Das ist notwendig, um die durch das Experiment mittels Eye-Tracking als problema-

tisch ermittelten Oberflächen-Komponenten besser zuordnen zu können. Weiter oben wurde bereits erwähnt, dass zwei der vier WebSCORE-Domänen, nämlich „Strategie“ sowie „Content & Funktionalität“, nicht mit Eye-Tracking evaluierbar sind. Diese beiden Domänen werden damit im weiteren Verlauf dieser Arbeit nicht mehr berücksichtigt. Die verbleibenden beiden Domänen werden an die Eye-Tracking-Optionen angepasst. Die Art der Anpassung wird hier im Folgenden erläutert.

Die WebSCORE-Designdomäne „Navigation & Interaktion“ soll für Eye-Tracking-Untersuchungen in zwei Bereiche aufgeteilt werden. Ein Bereich bezeichnet sich als „Struktur“. Darunter sind Oberflächen-Komponenten zu verstehen, die über- und untergeordnete Elemente aufweisen. Ein Beispiel ist eine Baumstruktur, wie zum Beispiel der Katalog für die Artikelsuche. Bäume werden heute zumeist in expandierbarer Form dargestellt. Hauptanforderung an Strukturen ist ein klar erkennbares Ordnungsprinzip zur Strukturierung sowie die klare und eindeutige Benennung der Informationselemente.

Der zweite Bereich nennt sich „Interaktion“. Im Interaktions-Bereich werden die Oberflächen-Komponenten untersucht, an denen der Benutzer die Funktionen der Software ausführt. Zu den häufigsten Bedienelementen zählen Menüs zur Auswahl von Software-Bereichen sowie Buttons zur Befehlsausführung und Links zur Navigation in andere Bereiche. Die typischen Elemente der Benutzer-Interaktion wie Menüs, Suche, Buttons etc. müssen gewährleistet sein. Die Positionierung und ein konsistentes Layout der Bedienelemente und eine einheitliche inhaltliche Belegung spielen eine wichtige Rolle bei der einfachen Handhabung.

Die zweigliedrige WebSCORE-Designdomäne „Mediendesign & Präsentation“ soll für Eye-Tracking-Untersuchungen in einen einzigen Bereich zusammengefasst werden. Dieser Bereich nennt sich „Präsentation“. Unter Präsentation ist alles Gestalterische zusammengefasst. Die Gestaltungsaspekte erstrecken sich über alle Komponenten, die auf einer Oberfläche vorhanden sind. Mit Eye-Tracking lässt sich insbesondere das Seitenlayout anhand der Gestaltgesetze untersuchen. Bei der Textgestaltung sind die Textlänge und der Umfang eines Informationselementes wichtige Aspekte. Unter piktorialen Aspekten lassen sich die Gestaltung von Icons und anderen grafischen Elementen auf Verständlichkeit und Intuitivität untersuchen.

Folgende Übersicht zeigt die Korrelation zwischen WebSCORE- und Eye-Tracking-Domänen (Tabelle 1).

Tabelle 1: Zuordnung WebSCORE-Domänen zu Eye-Tracking-Domänen

WebSCORE-Designdomäne	**Eye-Tracking-Designdomäne**
Strategie	---
Content & Funktionalität	---
Navigation & Interaktion	Struktur
	Interaktion
Mediendesign & Präsentation	Präsentation

Im Folgenden werden nur noch die Eye-Tracking-Designdomänen „Struktur“, „Interaktion“ und „Präsentation“ behandelt. Nach der hier vorgenommenen Anpassung werden diesen drei Domänen verschiedene Oberflächen-Komponenten, wie zum Beispiel Baum oder Menüs, zugeordnet werden (siehe Tabelle 10: Oberflächen-Komponenten und ihre Designdomänen, Seite 69). Die Oberflächen-Komponenten schließlich sind die zu evaluierenden Elemente, um die Usability-Qualität einer Software zu prüfen. Um ein Beispiel zu nennen: Möchte man wissen, wie gut die Interaktion der Software gelungen ist, kann dies über das Oberflächen-Element „Button“ vorgenommen werden.
Als Ergebnis schließlich sollen den Oberflächen-Elementen die jeweils relevanten Eye-Tracking-Parameter zugeordnet werden. Um das Beispiel fortzusetzen, würde sich die Frage stellen, welcher Parameter Aufschluss über die Güte der Buttons respektive die Interaktionsqualität der Software gibt (siehe Tabelle 20: Ergebnistabelle Designdomäne - Komponente - Parameter, Seite 96).

2.2 Eye-Tracking-Parameter für Usability

Im Folgenden soll basierend auf Literaturstudien ein Satz von Parametern definiert werden, die potenziell für Usability-Studien geeignet sind. Die Aufgabe dieser Forschungsarbeit ist es, aus dieser Vorauswahl die aussagekräftigsten Parameter zu ermitteln. Die Ermittlung der bedeutungsvollen Parameter erfolgt über ein Experiment.

2.2.1 Usability-relevante Augen- und Blickbewegungen

Wie bereits beschrieben wurde, wird zwischen Augen- und Blickbewegungen unterschieden. Im Kontext von Software-Usability sind vor allem die Blickbewegungen interessant, also Augenbewegungen, die mit bewusster Informationsaufnahme gekoppelt sind. Auch wenn Blickbewegungen immer mit Augenbewegungen gekoppelt sind, wird im Folgenden zur Vereinfachung nur noch von Blickbewegungen gesprochen.

Für den Anwendungsfall der Software-Usability sind in erster Linie Fixationen, die allgemein den informationsreichen Stellen gelten, relevant. Die Testperson hat fast immer den gleichen Abstand zur Bildschirmanzeige (ca. 60 cm). Dadurch treten kaum Vergenzbewegungen zum Ausgleichen von Nähe und Entfernung auf. Die betrachteten Bildschirminhalte sind im Falle der vorliegenden Untersuchung statisch und nicht dynamisch wie etwa Animationen oder Filme. Dadurch wird das Auftreten von Folgebewegungen ausgeschlossen. Die vestibulären Bewegungen dagegen treten immer auf, wenn der Kopf bewegt wird. Sie sind aber für Usability-Analysen mit Eye-Tracking nicht relevant, da diese Augenbewegungen relativ zum Kopf stattfinden. Für den Zweck dieser Arbeit sind aber nur Augenbewegungen relativ zum Objekt relevant.

An dieser Stelle soll auf zwei in der Literatur und in Forschungsarbeiten uneinheitlich verwendete Begriffe eingegangen werden. Die Begriffe „Blick“ und „Fixation“ werden oft als Synonyme verwendet. Andere Autoren hingegen unterscheiden Blick und (Einzel)-Fixation. Ein Blick beinhaltet sukzessive Fixationen auf ein und dasselbe Objekt oder auf dieselbe Objektgruppe. Der Blick und insbesondere seine Dauer sind demnach definiert als die Zeit zwischen zwei großen Sakkaden, wobei kleinere Korrektursakkaden dazwischen vorkommen dürfen. In dieser Arbeit werden die Begriffe voneinander unterschieden. Wenn von Fixationen die Rede ist, sind Einzelfixationen gemeint. Die Folge mehrerer Fixationen, unabhängig davon, ob kleinere oder größere Sakkaden dazwischen liegen, bildet einen Blickpfad. Allerdings erhalten bei den Analysen und Auswertungen die Fixationen, die auf ein und demselben Objekt stattgefunden haben, unter dem Begriff „Area of Interest“ (AOI) eine große Bedeutung.

Die Blickbewegungen wurden in bisherigen Forschungsarbeiten üblicherweise in zwei Dimensionen, zeitlich und räumlich, analysiert. Der räumliche Parameter des Fixationsortes kann mittels Koordinaten oder seinem zugehörigen Objektbereich (z. B. Button, Menüpunkt) beschrieben werden. Zu der zeitlichen Dimension gehört die Verweildauer einer Fixation auf einem bestimmten Bereich.

2.2.2 Usability-relevante Eye-Tracking-Parameter

Bei der Softwarenutzung sind vorwiegend nur die mit der aufgenommenen Information in Zusammenhang stehenden Blickbewegungsparameter, insbesondere die Fixationen, von Belang. Sakkadische Bewegungen, während denen keine Informationsaufnahme

stattfindet, sind nur relevant, indem sie zu neuen Fixationsorten führen. Besonders große Sakkadenweiten werden indirekt erfasst, indem die Koordinaten des neuen Fixationsortes von dem vorherigen stark abweichen.

Im Folgenden werden die in der Literatur gefundenen Usability-relevanten Eye-Tracking-Parameter skizziert (Tabelle 2). Dieser Satz von Parametern wurde in bisherigen Forschungsarbeiten und Usability-Studien am häufigsten eingesetzt. Sie sollen auch für diese Arbeit als aussagekräftige Parameter angenommen werden und durch das anschließende Experiment validiert werden.

Tabelle 2: Usability-relevante Eye-Tracking-Parameter

Parameter	**Wertebereich und Interpretation**
Fixationsdauer	Zeitraum zwischen zwei Sakkaden. Die Fixationsdauer ist die Zeit der Informationsaufnahme und gibt das Ausmaß des mentalen Verarbeitungsaufwandes von Informationen wieder. **Wertebereich:** • Durchschnitt beim Lesen 225 ms • Durchschnitt bei Bildwahrnehmung 330 ms • Durchschnitt bei visuellen Suchaufgaben 275 ms • Fixationsdauer Min: ca. 100 ms, Max: mehrere s
Kumulierte Fixationsdauer	Die Fixationsdauern mehrerer Fixationen lassen sich auf bestimmte Objektbereiche „Areas of Interest“ (z. B. Navigationsbereich), Versuchsabschnitte oder die gesamte Versuchsdauer zuordnen und kumulieren. Die Bildung von Mittelwerten, Modalwerten, Perzentilen, zeitlichen Häufigkeitsverteilungen etc., also die deskriptive Statistik, findet hier ihre Anwendung.
Suchzeit	Zeit vom Erscheinen eines Stimulus (z. B. Homepage einer Website) bis zur ersten Fixation auf dem Zielobjekt der Suchaufgabe (z. B. Kontakt-Link). Bei Usability-Studien ist abweichend von dieser Definition die erste Fixation nicht ausschlaggebend, da diese nicht zwingend sofort zur Entscheidung/Handlung führt. Erst der Mausklick ist ausschlaggebend dafür, dass Benutzer eine betrachtete Stelle als Zielstelle identifizieren. Die Suchzeit ist ein Parameter für die Art der Repräsentation (Informationsverarbeitung) oder des Layouts (Informationsaufnahme) einer Information. **Wertebereich:** in hohem Maße abhängig von der Aufgabe und dem Gegenstand der Untersuchung

Parameter	Wertebereich und Interpretation
Blickpfaddauer	Ein Blickpfad setzt sich aus einer Aufeinanderfolge von mehreren Fixationen zusammen. Die Blickpfaddauer wird definiert durch die Zeit vom ersten bis zum letzten Messpunkt innerhalb eines Blickpfades. In der Gesamtzeit sind die Fixationsdauern, die Sakkadendauern und Zeiten für andere Bewegungen oder Messartefakte enthalten. Die Blickpfaddauer ist ein Kriterium für die Effizienz der Informationsgestaltung. **Wertebereich:** wenige Sekunden bis viele Minuten
Kumulierte Übergangszeit	Darunter versteht man die Summe der Zeiten, bei denen sich der Blick von einem Blickobjekt zum nächsten bewegt. Während dieser Sakkadenzeiten findet keine Informationsaufnahme statt, es sind „tote Zeiten“. Die Übergangszeiten hängen im Wesentlichen von den Sakkadenweiten ab. Die kumulierte Übergangszeit ist ein Maß für die Informationsgestaltung. **Wertebereich:** Ganzzahl [Millisekunden, Sekunden] oder prozentualer Anteil an der Gesamtdauer, z. B. 30 % Übergangszeiten für eine visuelle Suchaufgabe
Anzahl der Fixationen	Hierbei handelt es sich um die Anzahl der Fixationen innerhalb eines bestimmten Zeitintervalls. Das Zeitintervall ist skalierbar: es kann die Zeit bis zur Lösung einer Teilaufgabe oder der gesamten Aufgabe sein. Die Deutung der ermittelten Anzahl der Fixationen hängt in hohem Maße von der Fragestellung ab. Allgemein ist die Häufigkeit der Fixationen ein Maß für die Verarbeitungs-Anforderungen. **Wertebereich:** Ganzzahl
Relative Häufigkeit der Fixationen von Objekten	Darunter versteht man die Anzahl der Fixationen eines Objektes („Area of Interest“) im Verhältnis zur Gesamtzahl der Fixationen während eines bestimmten Zeitraumes. **Wertebereich:** Prozentualer Wert zwischen 0 % bis 100 %
Anzahl Fixationen pro Zeiteinheit	Auch Fixationsfrequenz genannt. Die Zeiteinheit ist normiert. Sie kann pro Sekunde, pro Minute etc. angegeben werden. Sie ist ein Maß für die relative Wichtigkeit eines angeblickten Objektes. Diese Angabe ist maßgeblich davon abhängig, wie eine Fixation vordefiniert ist. **Wertebereich:** 0,5 bis maximal 10 Fixationen/Sekunde
Übergangshäufigkeiten	Das gesamte Stimulusmaterial wird in regelmäßige oder nach Aufgabenaspekten ausgewählte „Areas of Interest“ unterteilt. Die Anzahl der Übergänge von einem „Area of Interest“ zum anderen wird dann bestimmt. Die Übergangshäufigkeit ist ein Maßstab für die Güte der räumlichen Anordnung von Informationsobjekten, z. B. Instrumente im Cockpit eines Flugzeuges. **Wertebereich:** Matrix mit besetzten und unbesetzten Zellen.

Parameter	Wertebereich und Interpretation
	Die einzelnen Zellen enthalten Ganzzahlen (Anzahl Übergänge) oder Prozentwerte (Prozentualer Anteil der gesamten Übergänge). Die Übergangshäufigkeiten können gerichtet oder ungerichtet angegeben werden. Die Matrix der Übergangshäufigkeiten betrachtet immer nur zwei aufeinanderfolgende Fixationen.
Übergangsdichte	Die Übergangsdichte ist ein Wert, der sich aus der Matrix der Übergangshäufigkeiten, die alle möglichen Übergänge enthält, berechnen lässt. Die Dichte ist der Anteil der mit mindestens einem Wert gefüllten Zellen in Prozent der möglichen Übergänge. **Wertebereich:** Prozentwert zwischen 0 % bis 100 %. Für die Gestaltungsgüte von Software-Oberflächen wurde ein Mittelwert von 9,5 % (gut) und 10,2 % (schlecht) gefunden.
Richtungsänderungen	Eine Richtungsänderung des Blickpfades liegt vor, wenn eine Sakkade von der vorhergehenden Sakkade in der Richtung abweicht. Den Grad der Winkelabweichung des Vektors bei aufeinanderfolgenden Sakkaden definieren Autoren in ihren Forschungsarbeiten unterschiedlich. Richtungsänderungen weisen bei der visuellen Suche auf Diskrepanzen zwischen den Erwartungen der Person und der vorgefundenen Situation hin. In der Leseforschung sind Rücksprünge in der gleichen Zeile (Regressionen) ein Maß für die Schwierigkeit des Textes.

2.3 Experiment

2.3.1 Gegenstand der Mensch-Computer-Interaktion

Als experimenteller Untersuchungsgegenstand wurde eine E-Commerce-Software ausgewählt. Einer der Gründe für diese Wahl liegt darin, dass die Transaktionen in E-Commerce-Software prozedural einfach sind, verglichen mit den komplexen Aktionen in Betriebssystem-Software oder Spezialanwendungen. Eine methodische Untersuchung lässt sich folglich auch mit ungeschulten Teilnehmern durchführen. Die Mehrzahl der kognitiven Prozesse, die bei der Nutzung von E-Commerce involviert ist, lässt sich auf beinahe beliebige Software-Interaktionen übertragen.

Im Jahr 2006 erlangte der private E-Commerce im Web ein Wachstum um 44 Prozent auf 46 Milliarden Euro. Der Handel im Business-to-Business-Bereich stieg um 36 Prozent auf 392 Milliarden Euro (BITKOM 2007). Dieser sprunghafte Anstieg zeigt, in welchem Umfang immer mehr Verbraucher ihre Dienstleistungen und Waren über das

Internet einkaufen. Ein Drittel der Nutzer findet das gesuchte Produkt im E-Commerce-Shop nicht, obwohl der Artikel im Shop vorhanden ist (Nielsen u. Tahir 2001). Dadurch ist belegt, dass zahlreiche Bedienprobleme vorliegen und sich eine detaillierte Untersuchung lohnt.

Für die Studie wurde eine E-Procurement-Software ausgewählt. E-Procurement zählt zu den Web-Anwendungen. Von Desktop-Anwendungen kann Software-Usability, von Web-Anwendungen kann Web-Usability abgeleitet werden (Schweibenz u. Thissen 2003). Deshalb wurde in dieser Arbeit auch WebSCORE für Webseiten und Web-Anwendungen als Experten-Referenzmodell ausgewählt.

Die Kernfunktion von E-Procurement ist die Durchführung elektronischer Bedarfsanforderungen. Mitarbeiter eines Unternehmens können den täglichen Materialbedarf elektronisch anfordern. Funktional hat E-Procurement einen ausgeprägten Prozess-Charakter und umfasst vorrangig zwei Kernprozesse. Am Anfang des Beschaffungsprozesses steht der Bedarfsanforderer. Durch das Auslösen einer Bedarfsanforderung wird der zweite Prozess, die Genehmigung durch den Budgetverantwortlichen in Gang gesetzt. Das Ende des Genehmigungsprozesses bildet eine Schnittstelle in das zentrale Einkaufssystem des Unternehmens.

Im Experiment wird der erste Kernprozess verwendet, also die Bedarfsanforderung des Mitarbeiters (Abbildung 5). Dieser Ablauf entspricht den Bestellabläufen in E-Shops. Nach der Anmeldung kann der Anforderer zwischen zwei Suchmethoden wählen. Entweder sucht er den gewünschten Artikel über einen Katalog oder über die Produkt-Textsuche. Bei beiden Recherchearten werden die Ergebnisse in einer Artikelliste angezeigt. Bei Auswahl eines Artikels wird dieser im Detail angezeigt und Angaben wie z. B. die Menge, können angepasst werden. Von der Anforderung des Artikels gelangt man anschließend in die Kontierung. Sie verlangt die Eingabe einer Kostenstelle. Die Kontierung führt weiter zum Warenkorb. Dort werden die angeforderten Artikel mit Artikeldaten, Mengen- und Preisangaben zur Überprüfung angezeigt. Im Warenkorb kann die Anforderung ausgelöst werden. An manchen Stellen sind Rücksprünge auf vorherige Bereiche des Workflows möglich.

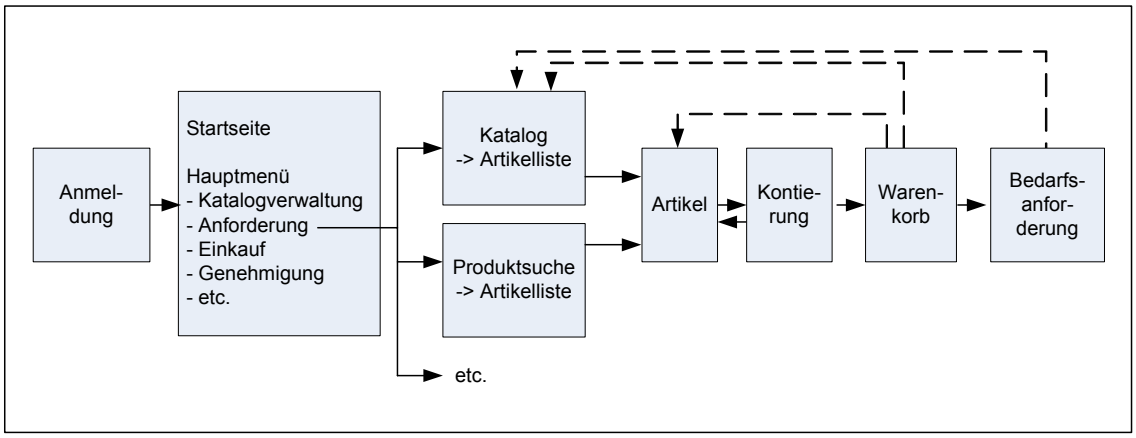

Abbildung 5: Schematischer Ablauf einer Anforderung in E-Procurement

2.3.2 Testumgebung und Situation

Das Experiment wurde auf der internationalen Computermesse CeBIT 2006 in Hannover und dort auf dem Messestand der Wilken GmbH durchgeführt. Der Zeitraum umfasste sieben Tage in der Zeit vom 9.-15. März 2006. Bezüglich des Ortes der Studie handelt es sich nicht um einen klassischen Labortest. Vielmehr handelt es sich um das Testen mit einer Zufallspopulation. Der Test fand auf ca. zwei Quadratmetern am Rande des Messestandes statt (Abbildung 6). Die technische Ausstattung bestand aus einem Eye-Tracking-Monitor, einer Tastatur und einer Maus zur Dateneingabe und einem Farbdrucker (Rückseite der Stele). Zur Ausstattung gehörte ein höhenverstellbarer Stuhl, auf dem die Testperson Platz nahm. Eine Stele mit dem Aufdruck „Wie guckst Du?“ informierte über den Zweck und den Hintergrund der wissenschaftlichen Studie.

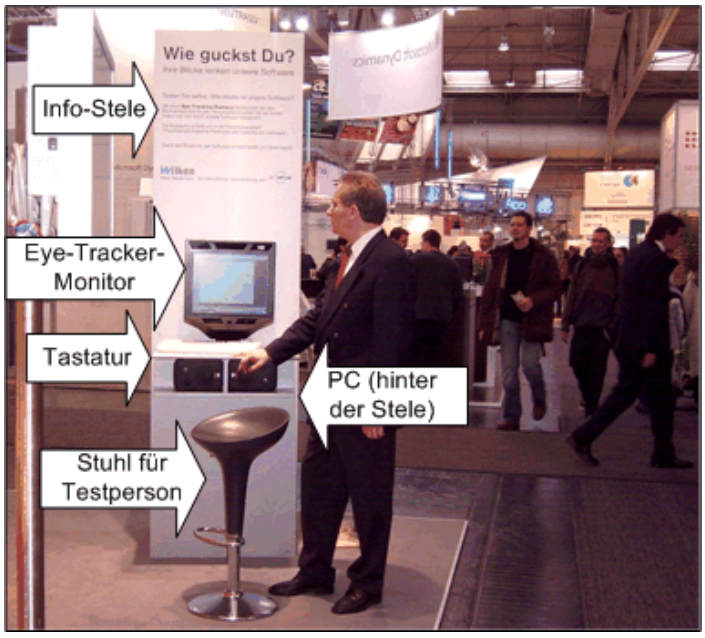

Abbildung 6: Fotografie des experimentellen Aufbaus, CeBIT 2006

2.3.3 Testaufgabe

Da die Gebrauchstauglichkeit einer Software wesentlich davon abhängt, welchen persönlichen Hintergrund (Domänenwissen, Erfahrung etc.) bei einem Nutzer vorhanden ist, wird ein allgemeinverständliches Software-Beispiel (hier: Bestell-Software) verwendet. So ist sichergestellt, dass die Aufgaben auch von Personen ohne besondere Vorkenntnisse gelöst werden können. Individuell unterschiedliche Vorkenntnisse wirken sich damit nicht maßgeblich auf die Messergebnisse aus.

Testaufgaben können auf zwei Arten gestellt werden: zur Ermittlung von Detailproblemen oder von allgemeinen Benutzerproblemen, die auf weite Teile des Produktes verallgemeinert werden können. In diesem Experiment sollen allgemeine Benutzerprobleme untersucht werden, da die zentrale Forschungsfrage dieser Arbeit darauf gerichtet ist, die Usability-abhängigen Kennzahlen der Blickregistrierung zu ermitteln. Testaufgaben sollen einen realen Kontext aufweisen, damit sie die tatsächlichen Nutzungsprobleme wiedergeben können. Die realitätsnahe Aufgabe lautet somit:

„Fordern Sie einen Wilken-Kugelschreiber aus dem Katalog ‚Guilbert' an."

Bei dieser Testaufgabe handelt es sich um eine handlungsorientierte Formulierung. Dazu trägt das Verb „fordern Sie […] an" bei, ebenso die persönliche Anrede, die die Testperson zum Handeln motiviert. In der Aufgabe werden konkrete Objekte (Wilken Kugelschreiber, Katalog ‚Guilbert') benannt, die nur im Rahmen einer Handlung umgesetzt werden können. Die handlungsorientierte Formulierung wurde gewählt, da die Testpersonen über ein geringes fachspezifisches Vorwissen verfügen. Im Gegensatz dazu erfordern problemorientierte Formulierungen ein hohes Vorwissen.

Anmerkung zur Terminologie der Instruktion:
Bei der Instruktion wird das Wort „anfordern" verwendet, da der relevante Menüpunkt in der Software „Anforderung" lautet. Gängiges Vokabular für diesen Vorgang in Internet-Shops ist dagegen „bestellen" oder „einkaufen". Bei E-Procurement spricht man von „anfordern", da der ausgewählte Artikel noch nicht bestellt wird, sondern erst über Genehmigungsinstanzen im Unternehmen freigegeben werden muss. Der Test zeigt, dass viele Benutzer dennoch in der Software nach „bestellen" oder „einkaufen" suchen, trotz der anders lautenden Instruktion.

Für Software-Systeme ist es typisch, dass nicht nur ein einziger Lösungsweg möglich ist, sondern mehrere Wege zur Lösung der Aufgabe führen. Auch bei der Lösung der Testaufgabe können die Testpersonen zwei verschiedene Wege einschlagen.

Gemäß den allgemeinen Anforderungen an Testaufgaben (Sarodnick u. Brau 2006) erfüllt die Testaufgabe folgende Anforderungen:

- Die Aufgabe präsentiert inhaltlich die gesamte Bandbreite des Systems, da sich die Aufgabe auf den Kernprozess bezieht. Die Testaufgabe umfasst alle Webseiten der Software.
- Die Aufgabe stammt aus der Praxis, da die Anforderung eines Kugelschreibers zu den typischen Bestellvorgängen von Sachbearbeitern zählt. ‚Guilbert' ist zudem ein bekannter Büroartikelversand.
- Da mehrere Lösungswege möglich sind, ist die Aufgabe nicht trivial und lässt das Erkennen komplexer Probleme zu.
- Die Aufgabe ist in geeigneter Zeit zu bearbeiten. Die Aufgabe kann in ca. zwei Minuten gelöst werden.[2] Damit liegt die Aufgabe in der für Eye-Tracking typischen Zeitskala von 1 bis 10 Minuten.
- Zum Eingewöhnen ist die erste Aufgabe einfach, da sie sich ausschließlich auf das Hauptmenü bezieht.

Unterteilung der Testaufgabe in Teilaufgaben

Die Bedienung einer Software ist eine Folge von Handlungen. Den Handlungen wiederum gehen jeweils Orientierungs- und Verarbeitungsphasen voraus, um eine Entscheidung herbeizuführen.

> *„Eine komplexe Aufgabe wird hierbei in Teilaufgaben zerlegt, die sich wie die Glieder einer Kette aneinander reihen. Bricht ein Glied der Kette, dann wird die Gesamtaufgabe nicht oder nicht richtig ausgeführt. In jeder Teilaufgabe stecken somit potenzielle Ursachen für Bedienprobleme." (Mussgnug u. Stowasser 2003, S. 51)*

Gerade die einzelnen Details eines interaktiven Systems bergen in sich Probleme und müssen darum hinsichtlich ihrer Bedienbarkeit bewertet werden. Die Testaufgabe des

[2] Jeder Test dauerte mindestens 10 Minuten, d. h. inklusive Vorbereitung und nachträglichem Interview.

Experimentes wurde darum in Teilaufgaben unterteilt (Tabelle 3). Die erste Teilaufgabe, das Anmelden in der Software mit Benutzername, Passwort und Mandant ist ausgespart, da sie durch die Testleiterin selbst durchgeführt wurde. Das Abbilden von alternativen Teilaufgaben bei der Artikelrecherche ist notwendig, da die Software diese zur Unterstützung verschiedener Suchstrategien der Benutzer anbietet. Dabei konnten zusätzlich die bevorzugten Suchstrategien ermittelt und bewertet werden. In der zweiten und dritten Spalte der Tabelle werden die Benutzeraktionen und die darauf folgenden Systemreaktionen dargestellt.
Ausdrucke der Bildschirmseiten je Teilaufgabe finden sich im Anhang.

Tabelle 3: Teilaufgaben zu der Testaufgabe

Teilaufgabe	**Benutzeraktion (B)**	**Systemreaktion (S)**
1. Artikel-Recherche starten	**B1-1:** Startseite ansehen	**S1-1:** Hauptmenü oben, zentraler Darstellungsbereich zeigt eine Hintergrund-Grafik.
	B1-2: Klick auf Hauptmenü „Anforderung“	**S1-2:** Untermenü öffnet sich.
2a. Artikelauswahl über Katalog	**B2a-1:** Klick auf Untermenü „Katalog“	**S2a-1:** Seite „Katalog“ öffnet sich. Links: Unter „Kataloge“ sind zwei Kataloge sichtbar. Rechts: Artikelliste zeigt eine Tabelle mit Kopfzeile und einer leeren Zeile.
	B2a-2: Klick auf das Pfeilsymbol vor der Kataloggruppe „Guilbert Katalog 2001“	**S2a-2:** Drei untergeordnete Katalogruppen werden angezeigt.
	B2a-3: Klick auf das Pfeilsymbol vor der Kataloggruppe „Bürobedarf“	**S2a-3:** 14 untergeordnete Katalogruppen (Liste ist weiter scrollbar) werden angezeigt.
	B2a-4: Klick auf das Pfeilsymbol vor der Kataloggruppe „Schreiben und Korrigieren“	**S2a-4:** Sechs untergeordnete Kataloggruppen (Liste ist weiter scrollbar) werden angezeigt.
	B2a-5: Klick auf das Ordnersymbol vor der Kataloggruppe „Kugelschreiber“	**S2a-5:** In der Artikelliste rechts werden zehn Kugelschreiber angezeigt. Der erste Eintrag ist der gesuchte „Wilken-Kugelschreiber“.
2b. Artikelauswahl über Produktsuche	**B2b-1:** Klick auf Untermenü „Produktsuche“	**S2b-1:** Seite „Produktsuche“ öffnet sich. Links: Vier Felder Rechts: Die Artikelliste zeigt Tabelle mit Kopfzeile und einer leeren Zeile.

Teilaufgabe	Benutzeraktion (B)	Systemreaktion (S)
	B2b-2: - Im Feld „Katalog“ *Guilbert* auswählen - Im Feld „Bezeichnung“ *Kugelschreiber* eingeben - Kontrollkästchen *Langtext* aktivieren - Klick auf Button „Suche“	**S2b-2:** Im Erfolgsfall der Produktsuche werden in der Artikelliste rechts zehn Kugelschreiber angezeigt. Der erste Eintrag ist der gesuchte „Wilken-Kugelschreiber“. Bei fehlerhafter Suche wird eine Fehlermeldung „Keine Ergebnisse gefunden!“ in der Statuszeile unten ausgegeben.
3. Artikel aus Artikelliste auswählen	**B3-1:** Klick auf Bild oder ID des Wilken-Kugelschreibers (erste Zeile in Artikelliste)	**S3-1:** Seite „Anforderung Artikel“ öffnet sich.
4. Artikeldetail prüfen, Menge eingeben und Aufgabe fortsetzen	**B4-1:** - Artikeldaten lesen, Bild betrachten - ggf. Menge anpassen - ggf. „Gew. Liefertermin“ erfassen - ggf. weitere Register anklicken und die Angaben lesen, jedoch keine Änderung notwendig	**S4-1:** Angeklickte Register öffnen sich.
	B4-2: - Klick auf Button „Weiter“	**S4-2:** Seite „Kontierung“ öffnet sich
5. Kontierung durchführen und Aufgabe fortsetzen	**B5-1:** Die Testleiterin gibt hier die Instruktion, auf den Button „Verteilen“ zu klicken, da für diesen Schritt Expertenwissen notwendig ist.	**S5-1:** In der Tabelle „Kontierungen“ wird eine Position angezeigt.
	B5-2: Klick auf Button „Zum Warenkorb“	**S5-2:** Seite „Warenkorb“ öffnet sich. Dort befindet sich eine Position.
6. Warenkorb prüfen, Aufgabe fortsetzen und beenden	**B6-1:** Position „Wilken-Kugelschreiber“ betrachten	---
	B6-2: Klick auf Button „Anfordern“	**S6-2:** Seite „Bedarfsanforderung“ öffnet sich. Dort endet der Ablauf.
	B6-3: Beenden der Aufzeichnung (mit der Taste F10) durch die Testleiterin	**S6-3:** Die E-Procurement-Software schließt und ClearView ist wieder aktiv.

2.3.4 Testpersonen

Die 120 Testpersonen wurden aus den Messebesuchern und Ausstellern ausgewählt. Die Probanden blieben namentlich anonym. Die Testpersonen wiesen folgende Profile auf (Tabelle 4):

Tabelle 4: Probandenprofile

Merkmal des Profils	**Merkmalsausprägung**
Anzahl Probanden	120
Geschätzte[1)] Altersspanne	Ca. 18 bis 70 Jahre
Geschlecht (Strichliste[2)])	männlich (70 %), weiblich (30 %)
Landessprache	deutsch
Beobachteter[1)] Grad der Erfahrung am PC	Fortgeschrittene, Experten

1) von der Studienleiterin geschätzt bzw. beobachtet
2) von der Studienleiterin mittels einer Strichliste ermittelt

Vorkenntnisse und Erfahrung der Benutzer zur Beispiel-Software wurden von der Versuchsleitung nicht explizit abgefragt, sondern beobachtet. PC-Kenntnisse waren bei den Testpersonen grundsätzlich vorhanden. Die Studienleiterin beobachtete, dass sich Experten im Vergleich zu den Fortgeschrittenen dadurch auszeichneten, dass sie die Links sofort identifizierten und mit Meldungen schneller umgehen konnten. Keine der Testpersonen hatte bereits mit einer E-Procurement-Software gearbeitet, allerdings war der Umgang mit vergleichbarer Beschaffungs-Software und E-Shops bekannt.

Die Testpersonen nahmen freiwillig an dem Experiment teil. Die Personen konnten ablehnen, nachdem sie über die voraussichtliche Testdauer von ca. 15 Minuten informiert worden waren.

Für die Auswertung fand keine Subgruppen-Bildung statt. Der Grund liegt darin, dass das Forschungsziel nicht die Untersuchung der konkreten Beispiel-Software zum Ziel hatte, sondern Korrelationen zwischen Software-Problemen und Eye-Tracking-Parametern herzustellen waren. Die Daten erhalten ihre Bedeutung in ihrer Aussagekraft hinsichtlich der Usability, nicht in der Unterteilung nach Probanden-Spezifika.

2.3.5 Hardware und Software für Eye-Tracking

Eye-Tracking-Hardware

Die technischen Leistungen der Mess-Systeme haben einen erheblichen Einfluss auf die Reliabilität und Validität der gemessenen Daten. Aus diesem Grund wurde für das Experiment in diesem Promotionsvorhaben ein System ausgewählt, das auf Usability-Tests spezialisiert ist. Die wesentlichen Anforderungen an die Messanordnung dieses Experimentes waren:

- Freiheit für Kopf- und Körperbewegungen, um mit der Software interagieren zu können (Vorbeugen, Mausklicks, Tasteneingaben)
- Testperson sitzt vor dem Bildschirm, um die Testaufgabe zu lösen
- Keine Belastung der Testperson durch invasive Geräte, um eine natürliche Benutzungssituation zu erzeugen
- Mobil einsetzbares und leicht zu transportierendes Gerät
- Robustes Gerät, d. h. keine System-Abstürze während der Testdurchführung
- Einfache und schnelle Kalibrierung des Systems
- Ad-hoc-Auswertungen als Basis für das retrospektive, gebundene Interview

Bei dem Gerät, das diese Anforderungen erfüllte und das für das Experiment ausgewählt wurde, handelte es sich um das System Tobii 1750 Eye-Tracker (Abbildung 7).

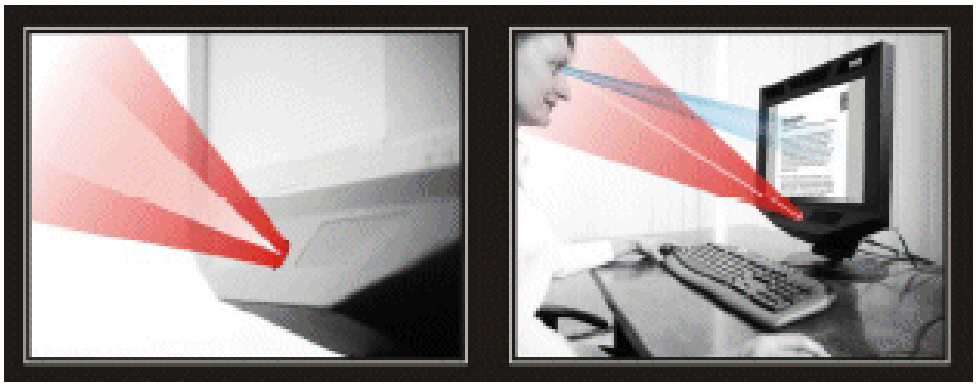

Abbildung 7: Tobii Eye-Tracking-Hardware

Dieses System wurde speziell für die Untersuchung von On-Screen-Material entwickelt. Der Eye-Tracker ist in einen 17“ TFT Monitor integriert. Die Messung erfolgt nach der Lichtreflexionsmethode. Das verwendete Eye-Tracking-System hat keinen Kontakt zu

Körperteilen der Testperson. Kontaktlose Systeme haben den Vorteil, dass sich die Versuchsperson relativ frei an einem Schreibtisch bewegen kann (Tabelle 5).

Tabelle 5: Technische Daten des Tobii-Systems 1750

Genauigkeit (accuracy)	0,5°
Drift	< 1°
Räumliche Auflösung	0,25°
Freiheit für Kopfbewegung	L 30 cm (d. h. 15 cm nach links oder rechts bewegen) H 16 cm (d. h. 8 cm nach oben oder unten bewegen) B 20 cm (d. h. 10 cm nach vorne oder hinten bewegen) Augenabstand zum Monitor: 60 cm
Binocular Tracking	ja
Abtast-Rate	50 Hz
Monitor	17“ TFT, Auflösung 1280*1024 Pixel
Schnittstellen	Firewire, USB, VGA, Power
Gewicht	Ca. 10 kg
Maximaler Blickwinkel (d. h. Winkel der Stellung des Auges zur Kamera)	+/- 40° auf dem Halbkreis oberhalb der Kamera +/- 10° direkt unterhalb der Kamera

Eye-Tracking-Software

Die gesamte Tobii-Apparatur umfasst eine Software für das Aufzeichnen und die Analyse von Blickbewegungsdaten, die zusammen mit dem Eye-Tracker eingesetzt wird. Außerdem wird ein PC zur Speicherung der Daten benötigt. ClearView ist die Software zur Vorbereitung, Durchführung und Analyse von Eye-Tracking-Studien, abgestimmt auf die Tobii-Hardware 1750, 2150 und x50.

Für das Experiment wurde in der ClearView-Software der Stimulustyp „Web“ ausgewählt, der speziell an die Anforderungen zur Untersuchung von Webseiten und Web-Anwendungen angepasst ist. Bei dem Web-Stimulus werden aufgezeichnet: Blickdaten, Augenpositionen, Pupillengröße, Film des Bildschirminhaltes, Bitmap jedes HTML-Frames, Browser-Window-Position, HTML-Frame-Position, Wechsel der Webseite, Tastendruck, Mausklick und Scrollen.

Die ClearView-Software bietet für die Auswertung der Daten mehrere Auswertungshilfen. Diese sind:

- **Blickverlauf in einer Aufzeichnung (Gaze-Replay):** Das Video zeigt die Blickposition einer einzigen Testperson auf der Software durch einen Punkt an.

Je länger die Fixationsdauer ist, umso größer ist der Punkt. Der Blickverlauf wird dabei mit einer Linie angezeigt.

- **Blickpunkte in einer Aufzeichnung (Gaze-Plot):** Der Gaze-Plot zeigt alle Blickpunkte (Fixationen) einer einzigen Testperson an, die auf einer Software-Seite stattgefunden haben. Es handelt sich dabei, im Gegensatz zum Gaze-Replay, um eine statische Anzeige des Blickpfades auf einer Bildschirmseite. Die Transparenz der Darstellung der Blickpunkte ist einstellbar. Die Verweildauer wird über die Größen der Blickpunkte repräsentiert. Ein Blickpfad ergibt sich aus der Abfolge von mehreren Blickpunkten. Ziffern innerhalb der Blickpunkte geben die Reihenfolge der Blickpunkte innerhalb eines Blickpfads an.
- **Kumulierte Blickpunkte über mehrere Aufzeichnungen (Hotspot):** Eine weitere Visualisierungsart der Blickbewegungen bieten Hotspots. Die Bereiche, die Anwender besonders häufig oder besonders lange betrachtet haben, werden je nach Intensität unterschiedlich eingefärbt. Unbeachtete Bereiche bleiben hingegen dunkel. Hotspots sind besonders aussagekräftig, wenn sie die Blickdaten über mehrere Aufzeichnungen kumulieren.
- **Interessengebiete („Areas of Interest“, kurz: AOI):** Die Interessengebiete lassen sich mit dem AOI-Tool als Rechtecke oder Polygone definieren. Je Webseite oder Softwaremaske lassen sich beliebig viele AOIs definieren. Es kann z. B. jeder Menüpunkt einer Menüleiste einen eigenen AOI darstellen. AOIs können über Text- und Excel-Export weiterverarbeitet werden. Aus diesen Daten lässt sich dann die Verteilung der Blickdauern je Interessengebiet ableiten.
- **Text-Export (je Aufzeichnung mehrere Textdateien):** Exportiert werden können die Blickdaten als Text- oder Excel-Dateien. Aus den Text- und Excel-Dateien lassen sich weitere Parameter rechnerisch ableiten.

2.3.6 Testablauf

Jeder Test wurde nach demselben Schema durchgeführt. Die Tests folgten zeitlich teilweise direkt aufeinander, teilweise gab es Pausen von ein bis zwei Stunden. Die technischen Systeme, Eye-Tracking-Monitor und Aufzeichnungssoftware, konnten ohne Zwischenschritte nach erfolgtem Test sofort wieder neu eingesetzt werden.

Alle Testdurchläufe wurden von der Verfasserin dieser Arbeit als Testleiterin geleitet. Für ein umfassendes Usability-Testing in der Laborpraxis werden jedoch mehrere Personen für unterschiedliche Aufgaben benötigt (Rubin 1994). Bei der Durchführung dieses Experiments nahm die Autorin mehrere dieser Funktionen ein. Sie leitete den Test und beobachtete das Problemlöseverhalten. Eine separate Person zur Datenerfassung war nicht notwendig, da die Leiterin die Testpersonen und Zeitpunkte direkt in der Aufzeichnungssoftware verwaltete. Für technische Probleme mit der Software waren Systemberater des Herstellers auf dem Messestand verfügbar.

Der Testablauf untergliederte sich in folgende Schritte:

1. Begrüßung und Erläuterung der Studie und Technik

2. Vorbereitung

Der Stuhl wurde so eingestellt, dass die Testperson im korrekten Abstand von 60 cm und in richtiger Höhe vor der Augenkamera saß. Dabei sollte der Proband bereits seine typische Arbeitshaltung am PC einnehmen. Damit wurde ausgeschlossen, dass sich die Körperhaltung bei der Kalibrierung nicht von der Haltung bei der späteren Aufgabenbearbeitung unterscheidet. Wenn die Person sich während der Bedienung der Software zu stark nach vorne beugt, entstehen so genannte Paralaxe-Fehler. Durch den veränderten Abstand zwischen Augen und Bildschirm liegen die Blickpunkte dann nicht auf den tatsächlich betrachteten Stellen des Untersuchungsmaterials, sondern bis zu 1 cm oberhalb oder unterhalb. Dann wurden mit einer Fünf-Punkt-Kalibrierung die Augenbewegungen auf den Bildschirm normiert.

3. Durchführung

Die Durchführung startete mit der mündlichen Instruktion zur Testaufgabe. Anschließend wurde die E-Procurement Software im Aufzeichnungs-Modus gestartet. Die Testaufgabe war beendet, wenn der geforderte Artikel angefordert war, wenn die Person erfolglos abbrechen musste oder wenn das zu testende E-Procurement einen Systemabsturz hatte. Nach der Aufzeichnung der Blickbewegungen fand eine Verbalisierungsphase in Form eines gebundenen Interviews statt. Die Testperson wurde aufgefordert, besonders auffällige Stellen im Blickverlauf laut zu kommentieren. Dieses Interview hatte das Ziel, subjektive Ursachen über die Bedienprobleme zu erfahren und bestimmte Teile der Software von der Testperson näher analysieren zu lassen. Die Versuchsleiterin notierte sich die zentralen Aussagen schriftlich.

2.4 Analyse und Berechnungen

2.4.1 Analyse der Rohdaten

Die Aufzeichnungssoftware ClearView erlaubt den Export der Daten in verschiedene Formate, um andere Software, z. B. Statistik-Software, für die quantitative Analyse einsetzen zu können. Für jedes Recording wird eine Reihe von verschiedenen Textdateien erzeugt. Alle Exportdateien enthalten im Kopfbereich Informationen über Aufzeichnungsdatum und –uhrzeit, Studienname, Name des Subjects (Testperson), Name der Aufzeichnung, Bildschirmauflösung und Einheit der Koordinaten, hier Pixel. Für die Auswertung dieser Studie wurden folgende Exportdateien ausgewertet:

FXD-Datei: Fixationen

In diesem Export ist eine Zeile pro Fixation enthalten. Die Filter für eine Fixation können in den Einstellungen der Software definiert werden. Blickdaten, die nicht zu einer Fixation gehören, sind ausgelassen.

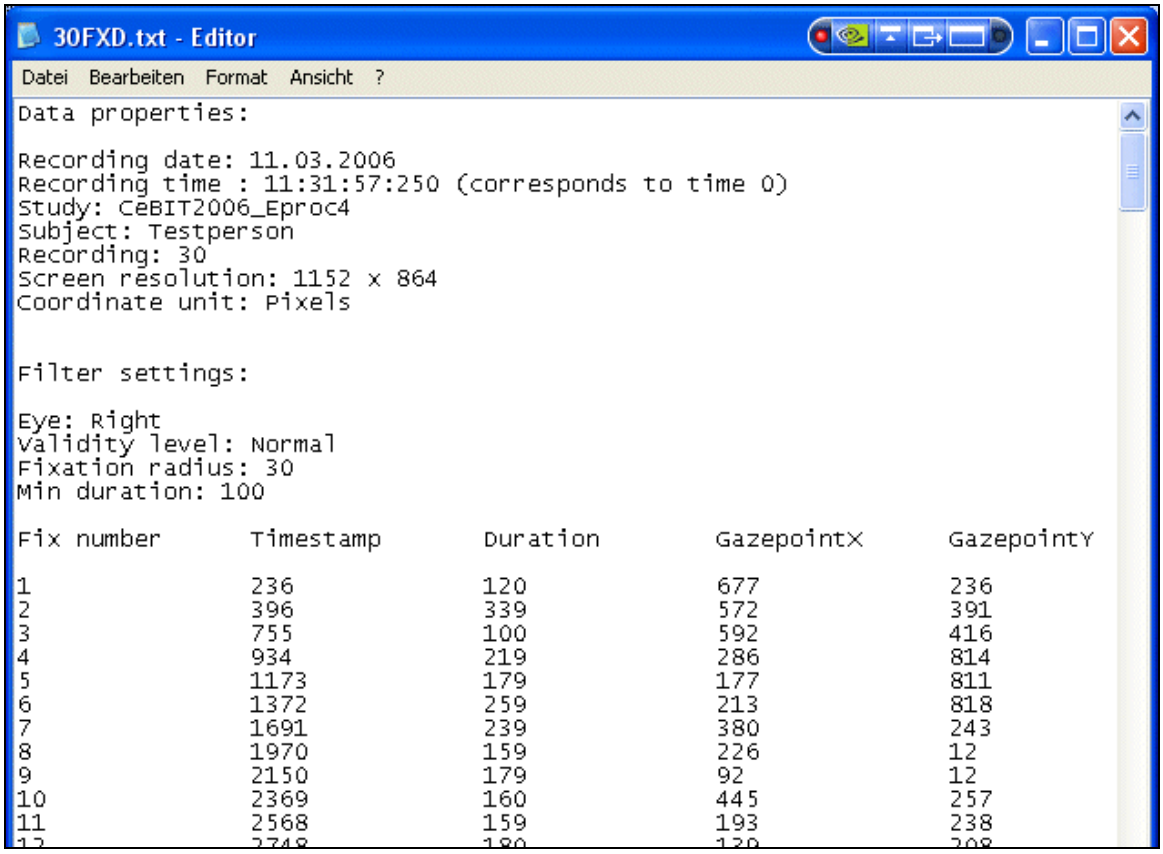

Abbildung 8: FXD-Datei Fixationen

Bedeutung der Spalten in Abbildung 8 von links nach rechts:

- **Fix number:** Sequentielle Nummer, vergeben für jeden Blickpunkt (Fixation)
- **Timestamp:** Zeitstempel [in Millisekunden], zeigt an, wann die Fixation begann, gemessen von Aufzeichnungsstart bis -ende
- **Duration:** Fixationsdauer [in Millisekunden]
- **GazepointX:** Horizontale Position des Fixationsmittelpunktes, gemessen in Pixeln vom linken Bildschirmrand
- **GazepointY:** Vertikale Position des Fixationsmittelpunktes, gemessen in Pixeln vom oberen Bildschirmrand

EVD-Datei: Event Daten

In diesem Export sind die Events, d. h. Ereignisse der Benutzerinteraktion, enthalten.

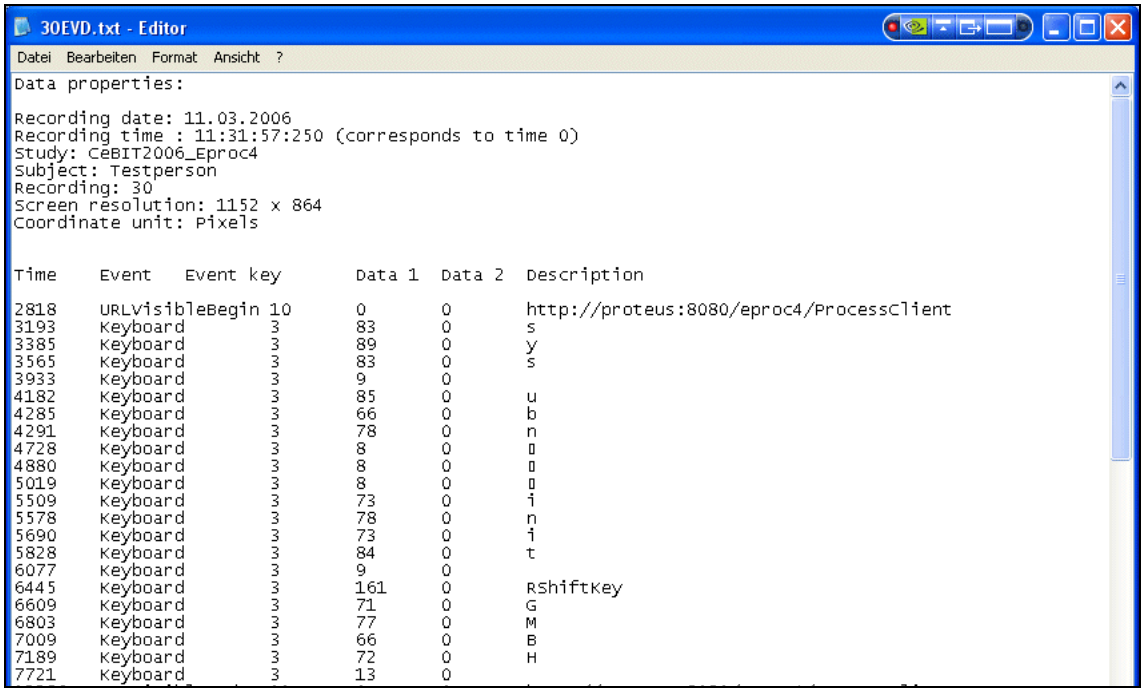

Abbildung 9: EVD-Datei Events

Bedeutung der Spalten in Abbildung 9 von links nach rechts:

- **Time:** Zeitstempel [in Millisekunden], zeigt an, wann der Event eintritt, gemessen von Aufzeichnungsstart bis -ende

- **Event:** Bezeichnung des Event-Typs. Es werden unterschieden: Showslide, Hideslide, Keyboard, LMouseButton, RMouseButton, URLVisibleBegin, URL-VisibleEnd etc.
- **Event key:** Eindeutiger Schlüssel des Event-Typs
- **Data 1:** Datenfeld für den Event, abhängig vom Event-Typ
    a. Keyboard: Ascii code der Taste
    b. LMouseButton: X Koordinate Mausklick links
    c. RMouseButton: X Koordinate Mausklick rechts
- **Data 2:** Wie Data 1.
    a. LMouseButton: Y Koordinate Mausklick links
    b. RMouseButton: Y Koordinate Mausklick rechts
- **Description:** Beschreibung des Events, abhängig vom Event-Typ
    a. Keyboard: Zeichen der Tastatureingabe

Von insgesamt 120 Aufzeichnungen wurden 100 Aufzeichnungen detailliert gesichtet. Die ersten 20 Aufzeichnungen wurden verworfen, da sie dazu dienten, den experimentellen Aufbau vor Ort zu optimieren. Zusätzlich zu den Exportdateien wurden die Blickverlaufs-Videos sowie die Gaze-Plots systematisch untersucht. Die Untersuchung wurde für jede Aufzeichnung einzeln durchgeführt.

Aufzeichnung vs. Testperson

Es ist zu beachten, dass die Anzahl Aufzeichnungen nicht gleichzusetzen ist mit der Anzahl Testpersonen. Eine Testperson konnte mehrere Aufzeichnungen durchführen. Das war der Fall, wenn es während einer Aufzeichnung zu einem Systemabsturz kam. In diesen Fällen wurde eine neue Aufzeichnung gestartet. Die Anzahl der Fehlversuche war nicht beschränkt, sie lag aber nicht höher als bei drei. Es kam vor, dass eine Testperson mehrere „Abbruchs“-Aufzeichnungen hatte, aber maximal eine erfolgreich zu Ende geführte Aufzeichnung. Die Ordnung erfolgte nach mehreren Gesichtspunkten:

- Jede Aufzeichnung erhielt eine eigene Nummer.
- Zu jeder Aufzeichnung wurden die einzelnen Softwaremasken ermittelt. Jede Maske wurde in einer eigenen Zeile festgeschrieben. Für jede Aufzeichnung

können unterschiedlich viele Masken vorliegen, da nicht jede Testperson den gleichen Weg in der Software ging und es auch bei gleichem Lösungsweg zu unterschiedlichen Interaktionen kam. Wenn eine Testperson in der Baumstruktur beispielsweise nicht sofort die richtigen Einträge anklickte, ergab sich pro Baumzustand eine neue Maske. Gerade die Katalog-Zustände variierten von Minimum drei bis zu sechs Masken. Eine Maske entspricht einem Gaze-Plot.

- Als nächstes wurde die Reliabilität des Tests bewertet. Dazu zählt die Genauigkeit je Plot. Ungenaue Plots wurden mit „Versatz“ gekennzeichnet. Wenn es zu einem Softwareabbruch kam, wurde die letzte Maske vor dem Abbruch mit „Abbruch“ gekennzeichnet.
- Sonstige Auffälligkeiten in den Aufzeichnungen wurden ebenfalls notiert. Beispielsweise gab es Plots, die keine oder nur sehr wenige Fixationen enthielten. Dies liegt an dem Fixations-Filter, d. h. die Fixationen waren kürzer als 100 ms im Radius von 30 Pixeln. Solche Aufzeichnungen wurden verworfen, da man nur die Videos hätte auswerten können. Davon wurde jedoch Abstand genommen, da die quantitative Auswertung von dynamischem Bildmaterial einen hohen Aufwand bedeutet.

Es lässt sich keine pauschale Aussage treffen, wie viele Aufzeichnungen insgesamt verwendet werden konnten. Je ausgewerteter Parameter variiert die Anzahl der verwertbaren Aufzeichnungen. Auch die Aufzeichnungen, die teilweise Versatz (d. h. visualisierte Blickpunkte weichen von den tatsächlichen ab) aufweisen oder zu einem Softwareabsturz führten, konnten in Teilen für die Auswertungen berücksichtigt werden. Weiterhin hing es immer von dem auszuwertenden Parameter ab, welche Aufzeichnungen und welche Plots darin berücksichtigt werden konnten.

Wie die Sichtung und Ordnung zeigt, ist die Auswertung von Eye-Tracking-Studien zur Usability von Software sehr komplex. Es gibt keine normierten Betrachtungszeiten pro Maske und die Benutzer bestimmten die Reihenfolge der Masken durch den individuellen Lösungsweg selbst. Zudem wird jede Softwaremaske unterschiedlich bedient, d. h. nicht alle Benutzer füllen optionale Felder aus. Oder Benutzer klappen den Katalogbaum an unterschiedlichen Stellen auf. Jede individuelle Interaktion erzeugt einen eigenen Gaze-Plot. Ziel der Sichtung aller Gaze-Plots ist es, vergleichbare Gaze-Plots aus

der Gesamtzahl der Plots herauszufinden und gegenüberzustellen. Durch die Komplexität in der Auswertung sind maschinelle Auswertungen weitgehend unmöglich.

2.4.2 Bewertung der Reliabilität der Tests

Um die Messgüte des Eye-Tracking-Tests zu bewerten, wurde zum einen die Genauigkeit der Messung geprüft. Auf die Genauigkeit der Messung hat vor allem die Qualität der eingesetzten Hardware und Software des Eye-Tracking-Systems einen Einfluss. Zum anderen soll die Stabilität der getesteten Anwendungssoftware bewertet werden. Da es sich um eine Software im Entwicklungsstadium handelte, gab es softwaretechnisch bedingt mehrere Abbrüche.

Fehleranalyse:

Die Sichtung der Ergebnisse ergab, dass die Aufzeichnungen unterschiedlich genau waren. Unter Genauigkeit wird hier verstanden, wie exakt die Blickpunkte auf den tatsächlich betrachteten Oberflächenelementen liegen. Die Genauigkeit lässt sich über eine Sichtprobe auf dem Gaze-Plot überprüfen. Gaze-Plots sind statische Darstellungen der Blickpunkte je Aufzeichnung und je Maske. Als Maßstab dafür, wie genau die Messung ist, wurde folgendes Regelwerk definiert. Ein Beispiel dazu zeigt Abbildung 10.

- Wenn die Blickpunkte (Fixationen) direkt über den Oberflächenelementen der Software liegen, ist die Messung genau. Beispiel A in Abbildung 10: Die Fixationen liegen direkt über den Buttons „Zurück“ und „Zum Warenkorb“. Ebenso liegt der über ein Kreuz visualisierte Mausklick auf einem Button.

- Wenn die Blickpunkte (Fixationen) versetzt neben den Oberflächenelementen der Software liegen, ist die Messung ungenau. Der Versatz kann dabei größer 10 mm sein. Ein Indiz für ungenaue Messung ist auch, dass die Stelle des Mausklicks nicht mit Fixationen bedeckt ist. Beispiel B in Abbildung 10: Die Fixationen liegen oberhalb der drei Buttons, wo sich Weißraum befindet. Es kann davon ausgegangen werden, dass die Testperson nicht die leeren Stellen, sondern die Buttons betrachtet hat.

- Die dritte Regel bezieht sich auf einen Sonderfall. Dieser liegt vor, wenn die Blickpunkte (Fixationen) zwar direkt auf Oberflächenelementen liegen, aber nicht diese, sondern benachbarte Elemente tatsächlich betrachtet wurden. Welche Elemente betrachtet wurden und welche nicht, lässt sich in der Regel aus der

Sequenz der Fixationen und den Übergangs-Sakkaden zwischen mehreren Blicken ermitteln. Beispiel C in Abbildung 10: Die Testperson las die einzelnen Wörter der Fehlermeldung mittels mehrerer Fixationen von links nach rechts. Die Fixationen haben einen Versatz, indem sie oberhalb der Fehlermeldung liegen. In der Plot-Darstellung liegen die Fixationen jedoch direkt auf den Buttons. Dass diese Darstellung nicht der realen Betrachtung entspricht, lässt sich aus der Kette der Fixationen 13-20 erkennen.

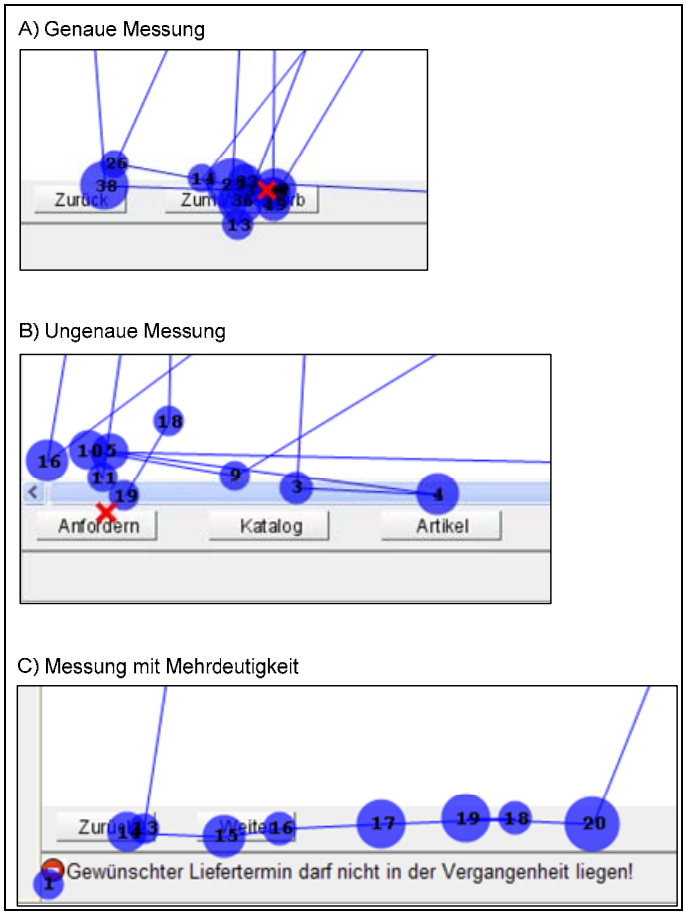

Abbildung 10: Genauigkeitsstufen der Messung

Die Genauigkeit variiert innerhalb einer Aufzeichnung von Maske zu Maske. Eine oder mehrere Masken einer Aufzeichnung können dabei sehr genaue Blickpunkte aufweisen, bei anderen Masken gibt es einen Versatz. Es kommt auch vor, dass nur einzelne Blickpunkte auf derselben Maske ungenau sind. Bei der Sichtung der Daten wurde beobachtet, dass die Blickpunkte im oberen Bildschirmbereich (z. B. Menüs) häufig genauer sind als die Blickpunkte im unteren Bildschirmbereich (z. B. Buttons oder Fehlermeldungen).

Die Genauigkeit wurde der Einfachheit halber in drei Stufen unterteilt. Sehr genaue Messungen (siehe Fall A oben) und ungenaue Messungen (siehe Fall B und C oben). Die ungenauen Messungen werden hier unterteilt in teilweise oder systematisch ungenau. „Teilweise ungenau“ ist eine Aufzeichnung, wenn nur einzelne Masken ungenaue Blickpunkte aufweisen. Diese Aufzeichnungen können dennoch ausgewertet werden, indem nur die ungenauen Masken nicht berücksichtigt werden, aber nicht die gesamte Aufzeichnung verworfen werden muss. Bei einem „systematischen“ Ungenauigkeitsfehler wurde die gesamte Aufzeichnung nicht in den Auswertungen berücksichtigt.

Tabelle 6: Häufigkeitsverteilung bezogen auf die Genauigkeitsstufen

Einstufung	**Anzahl Aufzeichnungen**	**Prozent**	**Kumulierte Prozent**
Genau	52	63,4 %	63,4 %
Teilweise ungenau	24	29,3 %	92,7 %
Systematisch ungenau	6	7,3 %	100 %
	Summe: 82	Summe: 100 %	

Die Zusammenfassung zeigt, dass knapp zwei Drittel der Aufzeichnungen (63,4 %) durchgängig genaue Messwerte aufweisen. Zusammen mit den teilweise ungenauen Messungen ergibt sich ein Anteil von 92,7 %. Die Unbrauchbarkeit aufgrund der Genauigkeit beschränkt sich auf weniger als 7,3 % und ist damit relativ niedrig.
Die Folgen der Messungenauigkeit wirken sich auf verschiedene Aspekte aus. Zum einen mussten die Aufzeichnungen mit größter Sorgfalt untersucht werden, um die teilweisen oder mehrdeutigen Blickpunkte aufzuspüren und separat zu behandeln. Der manuelle Auswertungsaufwand steigt dadurch sehr an.

Eine andere Auswirkung der Messungenauigkeit ist, dass sich sehr kleine Elemente nicht gezielt untersuchen ließen. So war es aufgrund der Genauigkeit nicht möglich, die beiden Symbole vor den Baumeinträgen differenziert zu untersuchen. Vor jedem Baumeintrag stehen zwei Symbole, ein grauer ausgefüllter Rechts-Pfeil und ein gelbes Ordner-Symbol. Um die Intuitivität der Software zu prüfen, wäre es interessant gewesen, die Fixationen auf diesen metaphorischen Piktogrammen zu untersuchen. Als relevanter Eye-Tracking-Parameter wäre die (Einzel-)Fixation und dabei speziell die Fixationsdauer zu messen. Je intuitiver ein Symbol verstanden wird, umso geringer der mentale Verarbeitungsaufwand. Der geringe Aufwand sollte sich theoretisch in einer extrem kurzen Fixationsdauer niederschlagen. Die Fixationen auf den beiden extrem kleinen

Symbolen konnten jedoch nicht eindeutig voneinander unterschieden werden aufgrund der für diesen Fall ungenügenden Genauigkeit der Messung. Um die intuitive Wirkung der beiden Symbole zu untersuchen, wäre ein anderer Versuchsaufbau notwendig gewesen. Die Symbole müssten isoliert werden und in einer größeren Auflösung dargestellt werden. An dieser Stelle sei auf die Forschungsarbeit zur Blickbewegungsanalyse zur Bewertung piktografischer Darstellungen verwiesen (Mussgnug u. Stowasser 2003).

Die Ergebnisse des Experimentes der vorliegenden Arbeit legen nahe, dass vor allem die kombinierten und sequenzabhängigen Eye-Tracking-Parameter aufschlussreich für interaktive Softwarebedienung sind. Dadurch fällt die Messungenauigkeit der einzelnen Parameter nicht sehr ins Gewicht und kann als statistische Messungenauigkeit nahezu vernachlässig werden.

Systemabbrüche in der E-Procurement-Software

Zum anderen soll in diesem Kapitel die Stabilität der getesteten Anwendungssoftware (zu unterscheiden von der Eye-Tracking-Software) erläutert werden. Da die Software für sich gesehen nicht stabil lief, kam es während einer Aufzeichnung zu einem plötzlichen Abbruch. Der Abbruch trat immer in Folge einer Interaktion ein, also eines Mausklicks oder einer Tasteneingabe. Bei einem Abbruch gab der Microsoft Internet Explorer eine Meldung aus: „*Die Internet Seite kann nicht geöffnet werden. Vorgang abgebrochen*.“ Die Meldung musste dann mit OK bestätigt werden.

In diesen Fällen wurde die Testperson gefragt, ob Sie bereit wäre, eine neue Aufzeichnung zu starten. Die vor dem Abbruch aufgezeichneten Masken konnten in den Analysen und Auswertungen berücksichtigt werden.

3 Ergebnisse

Das Kapitel 3 stellt die einzelnen Ergebnisse der Studie dar. Am Ende des Kapitels befindet sich eine Zusammenfassung aller Ergebnisse.

3.1 Leistungsdaten zur Effizienzmessung

3.1.1 Bearbeitungsdauer bzw. Blickpfaddauer

Die Blickpfaddauer ist die Zeitdauer, die Benutzer für die gesamte Aufgabe benötigt haben. Berücksichtigt werden dabei nur die erfolgreich beendeten Aufgaben.

Die gesamte Bearbeitungsdauer kann grundsätzlich auch über herkömmliche Logging-Methoden gewonnen werden. Dabei wird der Aufruf der Software bis zum Beenden der Software von der Systemuhr des Rechners aufgezeichnet. Die Bearbeitungsdauer entspricht bei Eye-Tracking der Blickpfaddauer. Dadurch erübrigt sich ein separates Logging. Bei der ClearView-Software wird die Blickpfaddauer nicht explizit ausgegeben. Sie kann aus den Zeitstempeln für die Fixationen (FXD-Datei) erschlossen werden. Im Experiment wurde der Zeitraum zwischen der ersten Fixation bis zur letzten Fixation vor Beenden der Aufzeichnung ermittelt. Dazwischen liegen eine Vielzahl von Klicks und Wechsel von Masken.

Die Datenauswertung ergab, dass die Aufgabe bei 70 Aufzeichnungen am Stück erfolgreich bearbeitet und beendet wurde. Gleichzeitig waren die Daten messtechnisch reliabel, d. h. bei diesen Aufzeichnungen traten keine kritischen Messfehler auf. Die Blickpfaddauer liegt in einem Wertebereich von 75 bis 250 Sekunden. Es wurden sieben Zeitdauer-Klassen von 25 Sekunden-Intervallen gebildet, um nicht jeden einzelnen Zeitwert unterscheiden zu müssen (Tabelle 7). Circa 70 % der Aufzeichnungen liegen im Bereich von 101 bis 175 Sekunden. Der Mittelwert beträgt 148,53 Sekunden.

Tabelle 7: Häufigkeitsverteilung der Blickpfaddauern

Zeitdauer [Sekunden]	Anzahl Aufzeichnungen	Prozent [%]	Prozent kumuliert [%]
75 bis 100	4	5,71	5,71
101 bis 125	18	25,71	31,43
126 bis 150	18	25,71	57,14
151 bis 175	13	18,57	75,71
176 bis 200	10	14,29	90,00
201 bis 225	6	8,57	98,57
226 bis 250	1	1,43	100
Summen	70	100	

Das Effizienz-Kriterium fordert eine möglichst kurze Bearbeitungsdauer einer Aufgabe. Im vorliegenden Beispiel liegt jedoch keine Skala zu einem Vergleich vor. Es kann daher keine Aussage getroffen werden, ob und wie effizient zwei Minuten Bearbeitungsdauer sind. Ein Maßstab ließe sich durch eine Testwiederholung an einer optimierten Fassung gewinnen. Dabei darf jedoch nur eine veränderliche Variable vorkommen, in diesem Fall die verbesserte Oberfläche. Die Software-Funktionen müssten gleich bleiben. Der Benutzertest könnte dann mit der zweiten Variante dagegen getestet werden und würde eine Werteskala liefern. Bei einem zweiten Test wäre zu beachten, dass für den Vergleichstest neue Testpersonen eingesetzt werden müssten. Die Testpersonen aus dem ersten Test sind bereits mit der Software vertraut und haben bestimmte Erwartungshaltungen.

Vergleich durch Wiederholungsaufzeichnungen aufgrund von Abbrüchen

Wie geschildert, kann die gemessene Bearbeitungszeit der gesamten Aufgabe wegen des fehlenden Maßstabes nicht bewertet werden. Aufgrund der Abbrüche liegt jedoch ein Umstand vor, der einen Vergleich ermöglicht.

Die Vergleichs-Tests erstreckten sich dabei jedoch nicht über die gesamte Bearbeitungszeit, sondern für die Zeit während einer Teilaufgabe. Die vergleichbare Zeit beschränkte sich auf die erste Teilaufgabe, in der das richtige Menü aufgerufen werden musste. Dabei war es egal, ob nach dem Hauptmenü „Anforderung“ das Untermenü „Katalog“ oder „Produktsuche“ gewählt wurde. Das Zeitfenster, in dem die Erst- und die Wiederholungsaufzeichnungen verglichen werden konnten, lag also zwischen Aufruf der Startseite mit den sichtbaren Menüs und Auswahl eines Untermenüs.

Zum Vergleich bietet sich die Erst-Aufzeichnung gegen eine wiederholte Aufzeichnung nach einem Abbruch an. Der Software-Neustart nach einem Abbruch stellt dabei ein Optimum dar, da die Bedienung bereits bekannt war. Ingesamt wurden 17 Wiederholungs-Aufzeichnungen betrachtet. Der Mittelwert der Blickpfaddauer bei der erwähnten Teilaufgabe liegt bei 9,2 Sekunden. Dem gegenüber stehen 55 Erst-Aufzeichnungen, ca. 70 Aufzeichnungen abzüglich der 17 Wiederholungs-Aufzeichnungen. Die Erst-Aufzeichnungen ergaben einen Mittelwert der Blickpfaddauer von 16 Sekunden. Das folgende Diagramm von Box & Whisker (Abbildung 11) zeigt die Streuung. Auf der Ordinate ist die Blickpfaddauer [Sekunden] aufgeführt, auf der Abszisse sind die beiden Varianten Wiederholungs- und Erstaufzeichnung gegenübergestellt.

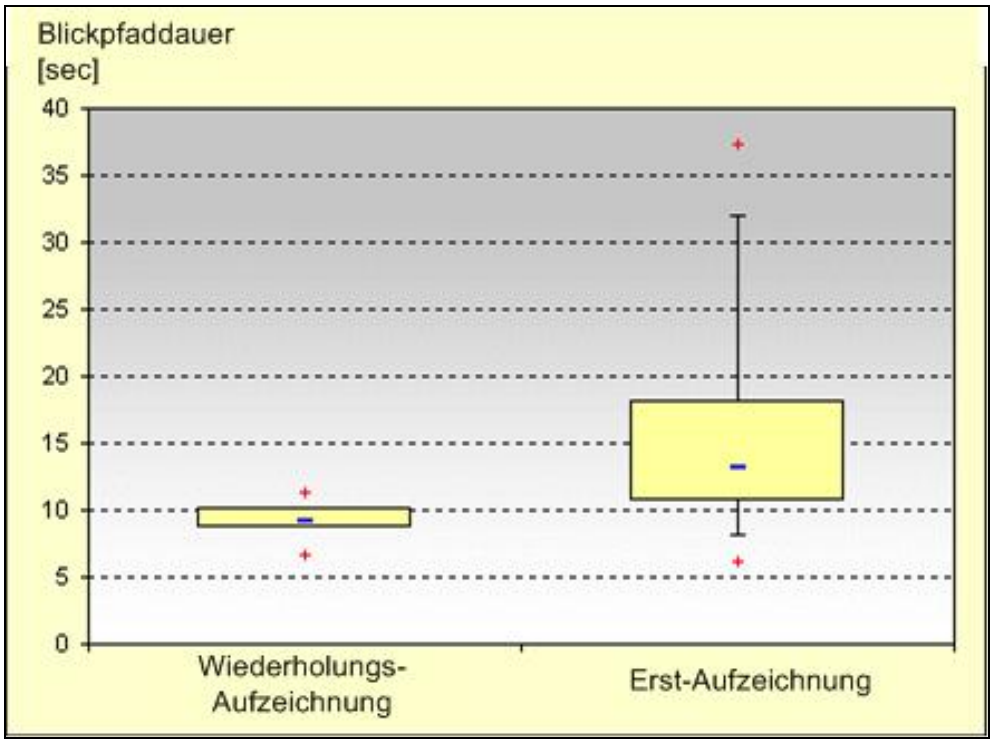

Abbildung 11: Vergleichs-Aufzeichnungen Box & Whisker Diagramm

Ergebnis der Vergleichs-Aufzeichnungen

Mit diesen Vergleichs-Aufzeichnungen lässt sich in jedem Fall der Parameter der Blickpfaddauer für die Effizienz bestätigen. Die Unterschiede in der Zeitdauer sind signifikant. Je kürzer die Blickpfaddauer ist, umso optimaler ist die Oberfläche gestaltet. Im Beispiel zeigte sich bei der Wiederholung eine fast halb so kurze Bearbeitungsdauer (bzw. Blickpfaddauer) wie bei der ersten Aufzeichnung.

3.1.2 Weitere Leistungsdaten zur Effizienzmessung

Neben der benötigten Bearbeitungszeit zur Effizienzmessung gibt es weitere Leistungsdaten. Dazu zählen die mit Erfolg beendeten Aufzeichnungen, die Klickanalyse sowie die Fehlerrate bei der Bearbeitung. Es soll ausdrücklich erwähnt werden, dass die Ergebnisse der Leistungsdaten nicht im Vordergrund dieser Arbeit stehen. Sie können auch mit anderen Evaluationsmethoden, wie etwa rechnergestützte Zeitmessung beim Öffnen und Schließen der Software oder bei Mausklicks oder Tasteneingaben, ermittelt werden. Es soll hier nur zur Vollständigkeit eines Usability-Tests erwähnt werden. Gleichzeitig ist eine Bewertung der Leistungsdaten aufgrund eines fehlenden Maßstabes nicht möglich. Für einen Vergleich der Leistungsdaten müsste der Test beispielsweise mit einer optimierten Oberfläche wiederholt werden.

Erfolgreich beendete Aufzeichnungen

In 70 von 100 gesichteten Aufzeichnungen haben die Testpersonen die Aufgabe am Stück und bis zum erfolgreichen Ende bearbeitet. Das Ende war erfolgreich, wenn der Wilken-Kugelschreiber in der Bedarfsanforderungs-Liste lag und abschließend mit dem Button „Anfordern“ bestellt wurde. Die erfolgreiche Beendigung der Gesamtaufgabe impliziert nicht, dass die Testpersonen während der Bearbeitung keine Bedienfehler machten oder sich für umständliche Lösungswege entschieden.

Die 30 Aufzeichnungen, die nicht erfolgreich zu Ende gebracht wurden, hatten ausschließlich Fehler in der Software zur Ursache. Diese Fehler sind damit nicht auf Bedienfehler zurückzuführen.

Zusammengefasst kann festgestellt werden, dass 100 % der Testpersonen die Aufgabe erfolgreich lösten. Dieser hohe Erfolgsgrad muss jedoch mit einem systematischen Vorbehalt betrachtet werden. Da das Experiment das Ziel hatte, möglichst viele Blickbewegungen zu messen, half die Studienleiterin in seltenen Fällen weiter, wenn die Testperson nach längerer Zeit einen Hinweis benötigte. Durch die gegebene Unterstützung konnte die Aufzeichnung fortgesetzt und die Blickverläufe auf den restlichen Masken bewertet werden. Ein Abbruch der Aufzeichnung bewirkt jedoch immer einen Neustart der Software und ermöglicht kein Fortsetzen an der letzten Stelle.

Klickanalyse

Die Klickanalyse gibt Aufschluss über den Bearbeitungsaufwand. Je mehr Klicks zur Lösung notwendig sind, als umso aufwändiger gilt die Bearbeitung allgemein. Dabei ist zu unterscheiden, wie viele Klicks im optimalen Fall die Mindestanzahl sind und wie viele Klicks tatsächlich stattfanden. Die Häufigkeit von Klicks ist damit ein Indikator für die Effizienz der Software.

Die Untersuchung der Mausklicks ergab, dass über die 70 erfolgreichen Aufzeichnungen hinweg insgesamt 1.161 Mausklicks ausgeführt wurden. Der Mittelwert der Klicks pro Aufzeichnung liegt demnach bei 16,6 Klicks. Das Minimum liegt bei neun Klicks (Lösungsweg über Produktsuche) bzw. elf Klicks (Lösungsweg über Katalogsuche). Dieses Minimum stimmt zufällig mit dem Optimum überein, d. h. die minimalen Klicks waren auch die geringste Anzahl der notwendigen Klicks. Der Mittelwert (16,6 Klicks) liegt deutlich über dem Optimum (neun bzw. elf Klicks). Die höchste Anzahl Klicks betrug 26 Klicks.

Die Klickanalyse ergibt damit, dass die Bedienung gemessen an der Klickhäufigkeit aufwändiger als notwendig gewesen war. Das wiederum lässt den Rückschluss auf eine noch nicht optimale Oberflächengestaltung zu.

Fehlerrate

Das System zeigte insgesamt eine niedrige Robustheit und Fehlertoleranz. Bei der kurzen Analyse hier sollen zwei Fehlerarten je nach Softwarereaktion unterschieden werden: Fehler, die eine Fehlermeldung erzeugten und Fehler, die keine Fehlermeldung ausgaben. Für den Benutzer stellte das einen schwerwiegenden Unterschied dar.

Die Bedienfehler, die keine explizite Fehlermeldung erzeugten, verursachten eine große Unsicherheit. Als Beispiel dafür sei genannt, dass nur eines der beiden Symbole vor den Verzeichnis-Baumeinträgen klickbar war. Nur das graue ausgefüllte Pfeilsymbol war sensitiv, nicht das gelbe Ordnersymbol. Als Folge wurde der Baumeintrag rot eingefärbt und nicht aufgeklappt.

Die andere Fehlerart erzeugte eine in der Statuszeile ausgegebene Fehlermeldung. Das war die Reaktion auf Eingabefehler, wie zum Beispiel das falsche Datum im Lieferdatums-Feld. Benutzer sind es heute gewohnt, in Volltextsuchfeldern nur Teilbegriffe einzugeben. Das war in der Testsoftware nicht möglich. Hier mussten ganze Wörter erfasst werden, eine generische Suche war nicht möglich. Die Ausgabe einer aussagekräftigen Fehlermeldung war insgesamt hilfreich. Allerdings stellte die Positionierung der Fehlermeldung eine Schwierigkeit dar. Die Meldung wurde nicht sofort gefunden oder musste relativ aufwändig mit einer unter Umständen weit entfernt liegenden fehlerhaften Eingabe verglichen werden.
Dieses Verhalten wurde detailliert in den Eye-Tracking-Daten untersucht und dazu der Parameter der kumulierten Übergangszeiten ermittelt.

3.2 Experten-Bewertung der Problemstellen

3.2.1 Teilaufgaben

Im Folgenden wird das Ergebnis der empirischen Analyse potenzieller Problemstellen dargestellt. Dafür werden die Teilaufgaben der Testaufgabe (siehe Tabelle 3, Seite 39) wiederholt und durch eine vierte Spalte ergänzt, in der die einzelnen Problemstellen aufgeführt sind (Tabelle 8).

Tabelle 8: Teilaufgaben der Testaufgabe mit Problemstellen

Teilaufgabe	Benutzer	System	Problemstellen
1. Artikelrecherche starten	**B1-1**	**S1-1**	- Grafik zeigt Einkaufskorb mit verfremdetem Software-Menü. Verwechslungsgefahr mit realem Menü. - Beschriftungen der Hauptmenü-Einträge sind mehrdeutig (Katalogverwaltung, Anforderung, Einkauf). - Anzeige nicht-relevanter Menüs für normale Anforderer. - Wenn Maus zu lange auf Menüpunkt gedrückt gehalten wird, färbt sich der ganze Menübereich ein. Dann: Klicken an beliebiger Stelle notwendig.
	B1-2	**S1-2**	- Cursor-Symbol ist invertiert: „Mouse-Over“ über einem Menüpunkt zeigt den Einfüge-Cursor, an anderer Stelle den Umriss-Pfeil.
2a. Artikelauswahl über Katalog	**B2a-1**	**S2a-1**	- Aufklappen des Kataloges ist nur über Pfeilsymbol möglich, nicht über gelbes Ordnersymbol, nicht über Bezeichnung. Das Pfeilsymbol ist zudem sehr klein. - Falls fälschlich der Ordner oder die Bezeichnung angeklickt wird, wird die Seite neu aufgebaut. Der Inhalt bleibt gleich, nur der angeklickte Eintrag wird rot eingefärbt. - Versuch, Werte in der leeren Tabellenzeile einzugeben, dort in eingabefähiger Zelle „Bezeichnung“.
	B2a-2	**S2a-2**	
	B2a-3	**S2a-3**	- Kein Ordnungssystem der Gruppen erkennbar (alphabetisch, logisch). - Die Gruppen überschneiden sich inhaltlich.
	B2a-4	**S2a-4**	
	B2a-5	**S2a-5**	Kein Pfeilsymbol mehr vorhanden.
2b. Artikelauswahl über Produktsuche	**B2b-1**	**S2b-1**	
	B2b-2	**S2b-2**	- Fehlerhafte Eingaben in Formularfeldern. - Feld „Bezeichnung“ muss kompletten Namen, nicht Teilwort, enthalten. - Kontrollkästchen muss aktiviert werden. - Fehlermeldung in Statuszeile wird nicht gefunden.

Teilaufgabe	Benutzer	System	Problemstellen
3. Artikel aus Artikelliste auswählen	**B3-1**	**S3-1**	- Es wird nicht sofort wahrgenommen, dass die Artikelliste nun gefüllt ist, da vorher auch bereits eine Zeile (leer) vorhanden war. - Bilder in Daumenabdruck-Größe sind zu klein und zu wenig unterscheidbar. - Spalte „ID“, dargestellt als unterstrichener Hyperlink, wird nicht als klickbares Element erkannt. - Spalte „Bezeichnung“ ist nicht klickbar.
4. Artikeldetail prüfen, Menge eingeben und Aufgabe fortsetzen	**B4-1**	**S4-1**	- Piktogramm hinter Mengeneinheit „Stück“ ist nur Grafik, nicht klickbar. - Feld „ID“ heißt jetzt abweichend von der Beschriftung in der Katalogrecherche „Artikelnummer“, ist jedoch dasselbe. - Liefertermin musste in einem bestimmten Format eingegeben werden, „heute“ war nicht möglich (Fehlermeldung). - Abkürzung „Gew. Liefertermin“ muss interpretiert werden (gewünschter?).
	B4-2	**S4-2**	- Button „Weiter“ wird rechts unten erwartet, nicht links unten.
5. Kontierung durchführen und Aufgabe fortsetzen	**B5-1**	**S5-1**	
	B5-2	**S5-2**	- Button „Zum Warenkorb“ unterscheidet sich von vorherigem Button „Weiter“.
6. Warenkorb prüfen, Aufgabe fortsetzen und beenden	**B6-1**	**---**	
	B6-2	**S6-2**	- Im Gegensatz zu den vorherigen Seiten ist der weiterführende Button „Anfordern“ jetzt der Button ganz links in der Buttonreihe, nicht ganz rechts. - Das System gibt keine Rückmeldung, dass der Anforderungsprozess abgeschlossen ist.
	B6-3	**S6-3**	

Weiterhin gab es folgende Problemstellen, die unabhängig von der Teilaufgabe waren, indem sie übergreifend vorhanden waren.

- Überschrift der Seite wird nicht wahrgenommen. Die Überschrift steht auf jeder Seite links oben, ist jedoch sehr dicht an den Feldern und hat die gleiche Schriftgröße wie Blocküberschriften innerhalb des zentralen Darstellungsbereiches.

- Um dem Benutzer anzuzeigen, dass das System gerade arbeitet, wird die Fortschrittsanzeige des Browsers aktualisiert. Die Fortschrittsanzeige befindet sich rechts unten im Browser. Sie war nur versierten Anwendern bekannt und wurde nur von diesen angesehen und beobachtet. Andere Benutzer waren verunsichert, weil sie sich fragten, ob das System noch arbeitet oder auf Eingaben wartet.
- Fehlermeldungen werden in der Statuszeile, d. h. die letzte Zeile unten, ausgegeben. Das fehlerhafte Feld wird zusätzlich rot eingefärbt (auffällig). Falls nur wenige Felder auf dem zentralen Darstellungsbereich sichtbar waren, und das rote Feld zudem hervorgehoben war, wurde die Meldung aufgrund des dazwischen liegenden Weißraumes nicht gesehen.
- Langsame Rechner-Arbeitsleistung, vermutlich aufgrund des Netzwerkes auf dem Messestand. Während Neuaufbau einer Seite (weiß), waren die Benutzer verwirrt.

3.2.2 Problemstellen je Oberflächen-Komponente

Die vorab ermittelten potenziellen Problemstellen in der Softwarebedienung werden im Folgenden einzelnen Oberflächen-Komponenten zugeordnet. Diese Zuordnung ist zweckmäßig, um Unterschiede der Komponenten hinsichtlich Problemhäufigkeiten nachzuweisen. Die problematischen Komponenten sind für Fehlbedienungen und Abbrüche entscheidend. Im nächsten Schritt werden die als wichtig ermittelten Komponenten den Eye-Tracking-Designdomänen zugeordnet.

Menüpunkte

Die Menüpunkte in der Menüleiste der E-Procurement-Software sind Drop-Down-Menüs. Bei Anklicken eines Menüpunktes werden die Menüeinträge nach unten aufgeklappt. Drop-Down-Menüs sind motorisch anspruchsvoll, da der Menüpunkt nur kurz angeklickt werden darf, keinesfalls darf die Maus ziehende Bewegungen ausführen. Das Menü dient zum Navigieren durch die Inhaltsstruktur.

Folgende Problemfelder können „Menü“ zugeordnet werden:

(1) Anzeige nicht-relevanter Menüs für normale Anforderer.

(1) Wenn Maus zu lange auf Menüpunkt gedrückt gehalten wird, färbt sich der ganze

Menübereich ein. Dann: Klicken an beliebiger Stelle notwendig. (Softwarefehler)

(1) Die Beschriftungen der Hauptmenü-Einträge sind mehrdeutig (Katalogverwaltung, Anforderung, Einkauf).

(1) Cursor-Symbol ist invertiert: Mouse-Over Effekt über einem Menüpunkt zeigt den Einfüge-Cursor, an anderer Stelle den Umriss-Pfeil. (Softwarefehler)

Baum und Baumeinträge

Der Katalog bildet eine Baumstruktur, bestehend aus Kataloghauptgruppen und Untergruppen. Bäume sind gängige Darstellungsformen für hierarchische Beziehungen von Inhalten. Bekannt sind Bäume für die Dateiverwaltung, wie z. B. der Datei-Explorer in Windows. Die Ebenen von Bäumen, z. B. Ordner im Explorer, können interaktiv auf- und zugeklappt werden. Die Blätter des Baumes befinden sich auf der untersten Ebene und sie enthalten die verwalteten Elemente, z. B. die Dateien. Die grafische Gestaltung von Bäumen muss unmissverständlich sein. Dazu zählen auch die klickbaren Symbole zum Aufklappen einer Baumhierarchie.

Folgende Problemfelder können „Baum“ zugeordnet werden:

(2a) Die Kataloggruppen überschneiden sich inhaltlich.

(2a) Kein Ordnungssystem der Gruppen erkennbar (alphabetisch, logisch).

(2a) Kein Pfeilsymbol mehr vorhanden.

(2a) Aufklappen des Kataloges ist nur über Pfeilsymbol möglich, nicht über gelbes Ordnersymbol, auch nicht über Bezeichnung. Das Pfeilsymbol ist zudem sehr klein.

(2a) Falls fälschlich der Ordner oder die Bezeichnung angeklickt wurde, wurde die Seite neu aufgebaut. Der Inhalt blieb der gleiche, nur der angeklickte Eintrag war rot eingefärbt.

(2a) Zwei Symbole zum Expandieren vorhanden! Ordnersymbol und Pfeilsymbol.

Buttons

Die Buttons in E-Procurement sind aktionsauslösende Buttons. Bei Klicken lösen sie direkt eine Aktion aus. Nachdem ein Formular ausgefüllt wurde, wird über den Button „Weiter“ in die nächste Seite entlang des Bestellprozesses verzweigt. Buttons sollten durch ihre gute Beschriftung verständlich sein. Gleiche Buttons sollten immer an der gleichen Position stehen, z. B. „Weiter“ rechts von „Zurück“.

Folgende Problemfelder können „Buttons“ zugeordnet werden:

(4) Button „Weiter“ wird rechts unten in der Seite erwartet, nicht links unten.

(5) Button „Zum Warenkorb“ unterscheidet sich von vorherigem Button „Weiter“.

(6) Im Gegensatz zu den vorherigen Seiten ist der weiterführende Button „Anfordern“ jetzt der Button ganz links in der Buttonreihe, nicht ganz rechts.

Tabelle mit Hyperlinks

Hyperlinks ermöglichen Verknüpfungen zu anderen Seiten oder Stellen innerhalb der aktiven Seite. Hyperlinks treten in E-Procurement als Link-Listen auf. In einer Artikeltabelle verweisen Hyperlinks auf die einzelnen Artikel. Hyperlinks lösen keine Funktionen aus. Als Hyperlinks können Texte oder Grafiken dienen. Für die Gestaltung ist es wichtig, dass der Benutzer die Links als solche erkennen und von dem normalen Text oder nicht sensitiven Grafiken unterscheiden kann. Links sollten konsistent gestaltet sein. Zudem sollte der Benutzer erkennen können, wohin die Links führen.

Folgende Problemfelder können „Tabelle mit Hyperlinks“ zugeordnet werden:

(3) ID, dargestellt als unterstrichener Hyperlink, wird nicht als klickbares Element erkannt.

(3) Bezeichnung wird als Link erwartet, ist aber nicht klickbar.

(2a) Fehleingaben in der leeren Tabellenzeile im Eingabe-Feld „Bezeichnung“.

(3) Es wird nicht sofort wahrgenommen, dass die Artikelliste jetzt gefüllt ist, da vorher auch bereits eine Zeile (leer) vorhanden war.

(3) Bilder in Daumenabdruck-Größe sind zu klein und zu wenig unterscheidbar.

Eingabefelder

Felder ermöglichen die Eingabe von Daten. Sie können entweder freie Texteingabefelder oder Felder mit Auswahlmöglichkeiten sein. Auswahlen sind dann sinnvoll, wenn große Wertebereiche erfasst werden sollen, die womöglich zu ungültigen Eingaben führen. Bei freien Texteingaben sollte die Länge von Textfeldern der erwarteten Eingabe entsprechen.

Folgende Problemfelder können „Eingabefeldern“ zugeordnet werden:

(2b) Fehlerhafte Eingaben in den Formularfeldern.

(2b) Feld „Bezeichnung“ muss kompletten Namen enthalten, nicht nur ein Teilwort.

(2b) Das Kontrollkästchen muss aktiviert werden.

(4) ID heißt jetzt abweichend von der Beschriftung in der Katalogrecherche „Artikelnummer“, ist jedoch dasselbe.

(4) Liefertermin musste in einem bestimmten Format eingegeben werden, „heute“ war nicht möglich (Fehlermeldung).

(4) Abkürzung „Gew. Liefertermin“ muss interpretiert werden. (gewünschter?)

(4) Piktogramm hinter Mengeneinheit „Stück“ ist nur Grafik, nicht klickbar.

Systemreaktionen

Meldungen vom System dienen dazu, dem Benutzer ein Feedback über Erfolg oder Nichterfolg seiner Aktionen zu geben.

Folgende Problemfelder können „Systemreaktionen“ zugeordnet werden:

(6) Das System gibt keine Rückmeldung, dass der Anforderungsprozess abgeschlossen ist.

(2b) Fehlermeldung in Statuszeile wird nicht gefunden.

Um dem Benutzer anzuzeigen, dass das System gerade arbeitet, wird die Fortschrittsanzeige des Browsers aktualisiert. Die Fortschrittsanzeige befindet sich rechts unten im Browser. Sie war nur versierten Anwendern bekannt und wurde von diesen auch tatsächlich angesehen und beobachtet. Andere Benutzer waren verunsichert und fragten sich, ob das System noch arbeitet oder auf Eingaben wartet.

Fehlermeldungen werden in der Statuszeile, letzte Zeile unten, ausgegeben. Das fehlerhafte Feld wird zusätzlich rot eingefärbt (auffällig). Falls nur wenige Felder auf dem zentralen Darstellungsbereich sichtbar waren, und das rote Feld zudem hervorgehoben war, wurde die Meldung aufgrund des dazwischen liegenden Weißraumes nicht gesehen.

Überschriften

Seitenüberschriften identifizieren die Seite und helfen dem Benutzer sich zu orientieren.

Folgende Problemfelder können „Überschriften“ zugeordnet werden:

Die Überschrift der Seite wird nicht wahrgenommen. Die Überschrift steht auf jeder Seite links oben, ist jedoch sehr dicht an den Feldern und hat die gleiche Schriftgröße wie Blocküberschriften innerhalb des zentralen Darstellungsbereiches.

Startseite

Die Startseite ist vergleichbar mit dem Deckblatt eines Buches. Es soll in die Thematik einführen. Meist geschieht dies mit Hilfe eines Bildes.

Folgende Problemfelder können „Startseite“ zugeordnet werden:

(1) Grafik zeigt einen Einkaufskorb mit einem verfremdeten Software- Menü. Letzteres kann mit dem realen Menü verwechselt werden.

3.2.3 Gewichtung der Problemstellen je Oberflächen-Komponente

Die ermittelten Problemstellen je Oberflächen-Komponente sollen nun nach der Wichtigkeit eingestuft werden. Zu einigen Komponenten offenbarte der Benutzertest Probleme. Manche Probleme wirken sich schwerwiegend aus, indem sie den Benutzer in die Irre führen oder Datenverlust bewirken. Andere Schwachstellen verursachen lediglich Irritationen oder führen zu unnötigen Verzögerungen in der Handhabung. Hier soll ein einfaches dreistufiges Gewichtungssystem genügen:

- Stufe „Dringlich“ (Anzahl der Probleme mit Faktor 3 multiplizieren)
- Stufe „Wichtig“ (Anzahl der Probleme mit Faktor 2 multiplizieren)
- Stufe „Optimierbar“ (Anzahl der Probleme ohne Multiplikation, Faktor 1)

Im Folgenden soll nun die dreistufige Gewichtung für die Probleme an den jeweiligen Oberflächen-Komponenten begründet werden.
Das Menü stellte die Eingangshürde in den Bestellprozess dar. Wenn an dieser anfänglichen und entscheidenden Stelle Benutzer bereits Fehler machten, war der ganze weitere Fortgang und die Lösung der Aufgabe insgesamt gefährdet. Dass die Menüs eine Hürde darstellten, lag vorwiegend in der mehrdeutigen Beschriftung der Menüs. Das Menü erhält damit die höchste Gewichtungsstufe „Dringlich“ (Faktor 3).
Der Katalogbaum und seine Bedienbarkeit sind sehr wichtig, da über den Katalog der gesuchte Artikel gefunden werden musste. Gerade an der Baumstruktur gab es viele Bedienprobleme, da der Baum kein klares Ordnungssystem aufwies und auch die einzelnen Einträge unübersichtlich waren. Der Baum erhält damit die höchste Gewichtungsstufe „Dringlich“ (Faktor 3).

Die Buttons stellen die wichtigsten Bedienelemente dar, um Aktionen in der Software auszulösen. Die nicht eindeutige Beschriftung und die Positionsänderung widersprechen der Erwartungskonformität und führten zu Umwegen oder Abbrüchen. Die Buttons erhalten damit die höchste Gewichtungsstufe „Dringlich“ (Faktor 3).

In der Tabelle mit Hyperlinks war entgegen der Erwartung der Benutzer weder das Bild noch die Bezeichnung anklickbar. Ausschließlich die ID war sensitiv. Da die ID unterstrichen dargestellt war, wurde diese als Link erkannt und die Aufgabe konnte bewältigt werden. Die Hyperlinks erhalten die mittlere Gewichtungsstufe „Wichtig“ (Faktor 2).

Die Eingabefelder hatten keine wesentliche Bedeutung, da nur sehr wenige Eingaben getätigt werden mussten. Die meisten Funktionen konnten über Mausklicks und nicht über Dateneingaben ausgeführt werden. Den Personen, die die Produktsuche über Freitexteingabe nutzten, gelang es fast nie, auf das erste Mal eine einwandfreie Eingabe zu tätigen. Durch „Versuch und Irrtum“ gelangen den Benutzern schließlich vom System akzeptierte Eingaben. Da fehlerhafte Eingaben zu bestimmten Systemreaktionen führen, ist ein Zusammenhang mit diesen herzustellen. Insgesamt erhalten die Eingabefelder die mittlere Gewichtungsstufe „Wichtig“ (Faktor 2).

Grundsätzlich sind Systemreaktionen für Benutzer von größter Wichtigkeit, da das System dem Benutzer damit das Feedback gibt. In dem Experiment sind zwei Systemreaktionen festzustellen, die für die Benutzer unzulänglich waren. Zum einen handelt es sich um die bloße Einfärbung bei den Baumeinträgen, wenn das falsche der beiden Symbole zum Aufklappen des Baumes angeklickt wurde. Es wurde keine zusätzliche Text-Fehlermeldung ausgegeben, sodass die Benutzer allein gelassen wurden bei der Fehlerbehebung. Die andere Hürde war es, die Fehlermeldung auf Eingabefehler aufzufinden. Die Meldung wird in der Statuszeile angezeigt und ist damit in vielen Situationen räumlich weit von der Fehlerquelle entfernt. Systemreaktionen erhalten die höchste Gewichtungsstufe „Dringlich“ (Faktor 3).

Überschriften besitzen grundsätzlich eine Orientierungsfunktion. Im Beispiel konnten sie aufgrund der unauffälligen Gestaltung nicht gut erkannt werden. Dieser Mangel führte jedoch zu keinen Problemen, da die Benutzer aufgrund der auf der Maske dargestellten Informationen ihre aktuelle Stelle erkennen konnten. Überschriften erhalten damit die Gewichtungsstufe „Optimierbar“ (Faktor 1).

Auf der Startseite bildete eine große Grafik einen optischen Reizstimulus. Da auf dem Stimulus Menüpunkte in künstlerisch verfremdeter Art dargestellt waren, führte dies zu Irritationen mit den realen Menüpunkten. Die Folge waren eine verlängerte Betrachtungsdauer und Fehlklicks auf die Grafik. Dieser gestalterische Mangel erhält die Gewichtungsstufe „Optimierbar“ (Faktor 1).

Die Problemstellen lassen sich je Oberflächen-Komponente gewichten. Die gewichteten Problemstellen haben einen Wertebereich von 1 bis 18. Die mit der Gewichtung „Dring-

lich“ bedachten Oberflächen-Komponenten werden mit Eye-Tracking detaillierter ausgewertet (Tabelle 9).

Tabelle 9: Gewichtung der Problemstellen

Oberflächen-Komponente	Anzahl Probleme	Gewichtungsfaktor	Probleme gewichtet
Menü	4	3	12
Baum, Baumeinträge	6	3	18
Buttons	3	3	9
Systemreaktionen	4	3	12
Tabelle mit Hyperlinks	5	2	10
Eingabefelder	7	2	14
Überschriften	1	1	1
Startseite	1	1	1

3.2.4 Oberflächen-Komponenten korreliert mit Designdomänen

Wie im vorherigen Kapitel dargelegt, existieren Problemstellen an verschiedenen Oberflächen-Komponenten. Die Problemstellen konnten gewichtet werden. Dringend zu behebende Probleme zeigten sich bei Menü, Baum, Buttons und Systemreaktionen. Diese Komponenten sollen nun den Eye-Tracking-Designdomänen zugeordnet werden. Zur kurzen Wiederholung sollen die drei Designdomänen nochmals erwähnt werden. Unter „Struktur“ lassen sich alle über- und untergeordneten Ordnungselemente untersuchen. „Interaktion“ fasst alle Bedienelemente zusammen. „Präsentation“ ist die Gestaltungsqualität der Oberflächen-Komponenten.

Die als dringlich eingestufte Komponente „Menüs“ ist ein Bedienelement. Mit Hilfe von Menüs werden Bereiche ausgewählt, in denen die Aufgabe gelöst werden kann. Menüs und deren Handhabung zählen zu der Designdomäne „Interaktion“.

Die Oberflächen-Komponente „Baum“ wies im Wesentlichen zwei Mängel auf. Der Baum hatte kein klares Ordnungssystem inne und zusätzlich waren die einzelnen Einträge unübersichtlich angeordnet. Die unklare Ordnung und die Länge des Baumes sind in der Struktur behaftet. Der Baum als Gesamtes wird damit der Designdomäne „Struktur“ zugeordnet. Dagegen ist bei einem einzelnen Eintrag, gleichgültig auf welcher Ebene er sich befindet, die Textgestaltung von Bedeutung. Textlänge und inhaltliche Klarheit in der Beschriftung sind dem Designbereich „Präsentation“ zuzuordnen. Damit

wird der bisher als eine einzige Komponente geführte Baum an dieser Stelle aufgespaltet in „Baum gesamt“ und „Baumeintrag“.

„Buttons“ sind ebenso wie „Menüs“ die klassischen Bedienelemente, über die Benutzer mit der Software interagieren. Buttons sind damit der Designdomäne „Interaktion“ zugeordnet.

„Systemreaktionen“ werden an verschiedenen Positionen auf einer Seite angezeigt. Die Systemreaktion „Fehlermeldung“ stellt daher ein Gestaltungsproblem dar, das sich auf das Seitenlayout zurückführen lässt. Damit soll hier die Komponente „Seitenlayout“ eingeführt werden. Das hat zudem den Vorteil, dass weitere Suchprobleme, wie z. B. die Suche nach den Buttons, über diese Komponente untersucht werden können. Das Seitenlayout lässt sich der Designdomäne „Präsentation“ zuordnen.

Oberflächen-Komponenten und Eye-Tracking-Designdomänen hängen zusammen. Menüs und Buttons gehören der Designdomäne „Interaktion“ an. Baumeinträge und das Seitenlayout sind der Domäne „Präsentation“ zuzuordnen. Der Baum ist der Designdomäne „Struktur“ zugehörig (Tabelle 10).

Tabelle 10: Oberflächen-Komponenten und ihre Designdomänen

Domäne / **Komponente**	Struktur	Interaktion	Präsentation
Menüs		**X**	
Baum gesamt	**X**		
Baumeintrag			**X**
Buttons		**X**	
Seitenlayout			**X**

Soll geklärt werden, wie gut die Usability bezüglich der Interaktion ist, kann man dies konkret an Menüs und Buttons untersuchen. Soll die Güte der Präsentation beurteilt werden, kann das Seitenlayout und die Textgestaltung der Baumeinträge betrachtet werden. Und um die Strukturqualität zu ermitteln, kann dazu der gesamte Baum evaluiert werden. Eine andere Richtung wäre, man evaluiert die einzelnen Komponenten und schlussfolgert, in welcher Designdomäne die meisten Plus- oder Minuspunkte liegen.

3.3 Ergebnisse pro Parameter

3.3.1 Suchzeit

Die Suchzeit ist hier definiert als die Zeit vom Start einer Suchaufgabe bis zum ersten Mausklick auf dem Zielobjekt.

Beispiel

Das gewählte Beispiel für die Suchzeit ist der Katalogbaum. Der Katalog dient dazu, dass Besteller den gewünschten Artikel in einer Kataloghierarchie finden können. Es handelt sich um einen aufklappbaren Baum. Über die Suchzeit kann die Struktur des Baumes analysiert werden. Zu diesem Zweck werden die Suchzeiten im Baum zu zwei verschiedenen Zuständen verglichen.

- A) Baum mit 2 Ebenen, insgesamt 3 Einträge sichtbar
- B) Baum mit 3 Ebenen, insgesamt 14 Einträge sichtbar

A) Baum mit 2 Ebenen, insgesamt 3 Einträge sichtbar (Abbildung 12)

Die richtige Auswahl im Katalogbaum ist wie folgt:

- 1. Ebene: Guilbert Katalog 2001
- 2. Ebene: Bürobedarf (Mausklick!)

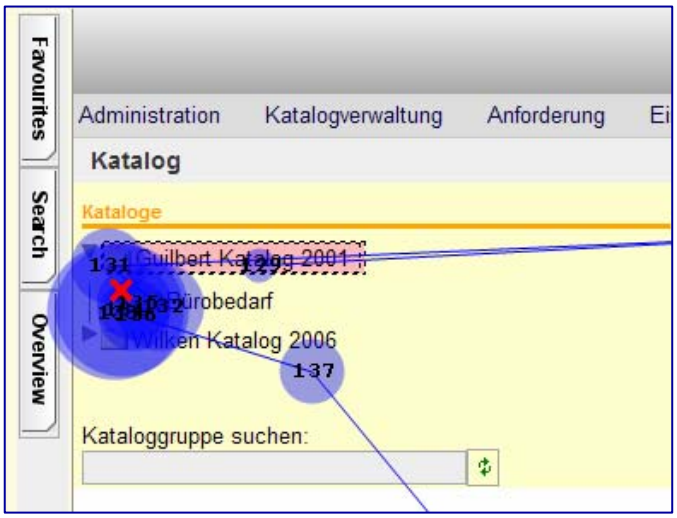

Abbildung 12: Beispielmaske Suchzeit Baum mit zwei Ebenen

B) Baum mit 3 Ebenen, insgesamt 14 Einträge sichtbar (Abbildung 13)

Die richtige Auswahl im Katalogbaum ist wie folgt:

- 1. Ebene ist bereits aufgeklappt: Guilbert Katalog 2001

- 2. Ebene ist bereits aufgeklappt: Bürobedarf
- 3. Ebene: Schreiben und Korrigieren (Mausklick!)

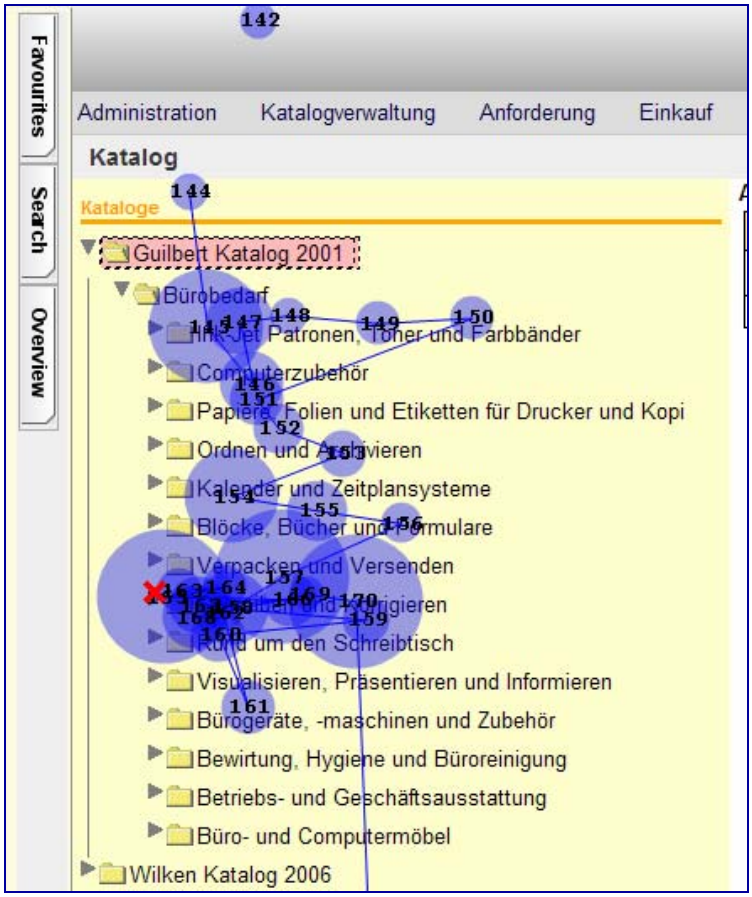

Abbildung 13: Beispielmaske Suchzeit Baum mit drei Ebenen

Berechnung des Parameters

In der ClearView-Software wird vom Studienbeginn an bis zum Ende der Blickverlauf im Abstand von 20 Millisekunden ermittelt. In dieser Zeit liegen Fixationen, Sakkaden und sonstige Körperbewegungen sowie Messartefakte. Neben den Blickdaten werden auch die Ereignisdaten protokolliert, die sich aus der Interaktion des Benutzers mit der Software ergeben. Ein typischer Event ist ein Mausklick oder eine Tasteneingabe. Die Ereignisdaten können in ClearView über eine Exportdatei (event data, EVD-Datei) ausgelesen werden. Aus den Ereignisdaten lassen sich die zwei entscheidenden Zeitpunkte bestimmen. Der Startzeitpunkt entspricht dem Öffnen der Maske. Das Ende der Suchzeit ist der Zeitpunkt des Mausklicks auf dem Zielobjekt.

Weitere Merkmale bei der Berechnung des Parameters:

- Es konnten nur die Aufzeichnungen berücksichtigt werden, die den Artikel über die Katalogsuche auswählten. Der alternative Weg über die Produktsuche führte auf eine andere Maske und ist damit nicht vergleichbar.
- Weiterhin konnten nur die Aufzeichnungen verwendet werden, die beide Vergleichswerte A) und B) für die Suchzeiten lieferten. Die erste Suchzeit wurde auf dem Baum mit zwei Ebenen untersucht, die zweite Suchzeit auf dem Baum mit drei Ebenen. Die beiden Teilschritte folgten im Idealfall direkt aufeinander. A) Einträge in der zweiten Ebene betrachtet und richtigen Eintrag geklickt, dann B) Einträge in der dritten Ebene betrachtet und richtigen Eintrag geklickt. Einige Testpersonen klickten in der zweiten Ebene nicht den richtigen Eintrag im Baum an. Über diesen oder mehrere Umwege kamen sie aber schließlich doch noch zu dem richtig aufgeklappten Baum auf der dritten Ebene. Die Betrachtung der richtigen dritten Ebene konnte in der Auswertung berücksichtigt werden, die Zwischen-Baumzustände nicht. Nicht vergleichbar waren die Vergleichswerte, wenn dazwischen ein Systemabsturz lag.
- Die beiden Daten in Baum A) und Baum B) wurden für jede Aufzeichnung zunächst getrennt ermittelt. Kumuliert wurden dann alle Aufzeichnungen für Baum A) und dann alle Aufzeichnungen für Baum B).

Ergebnisse

Insgesamt konnten 53 Aufzeichnungen für relevant befunden und ausgewertet werden. Die 53 Aufzeichnungen verfügten über jeweils beide Plots mit den zwei verschiedenen Baumstrukturen Baum A und Baum B (Tabelle 11).

Tabelle 11: Ergebnis Suchzeiten in zwei Baumstrukturen

Struktur	**Anzahl Baum-einträge**	**Mini-mum [s]**	**Maxi-mum [s]**	**Mittel-wert [s]**	**Standard-Abweichung**
Baum A (2. Ebene)	3 Einträge	0,2	6,3	3,0	1,6
Baum B (3. Ebene)	14 Einträge	3,1	28,3	12,2	6,1

Das Ergebnis zeigt, dass drei Einträge im Mittelwert nach 3,0 Sekunden eine Entscheidung für einen der drei Einträge ermöglichen. Im Vergleich dazu benötigen 14 Einträge

12,2 Sekunden. Je mehr Einträge in einem Baum sichtbar sind und durchsucht werden müssen, umso länger ist auch die Suchzeit. Längere Suchzeiten deuten demnach auf komplexere Strukturen hin. Das Verhältnis zwischen der Anzahl der Einträge und der benötigten Zeitdauer je Anzahl Einträge ist damit proportional. Da jedoch nur zwei Punkte bekannt sind, ist der genaue Verlauf unbekannt.

3.3.2 Kumulierte Übergangszeit

Die kumulierte Übergangszeit wurde hier ermittelt als ein Parameter, der Aufschluss über die Gestaltungsqualität gibt. Bei gut strukturierter Seitengestaltung ist eine möglichst kurze Übergangszeit anzustreben. Je größer der Anteil der Blickzeiten ist, umso weniger Zeit musste für Übergänge zwischen den Objekten verwendet werden. Es wurde auch ermittelt, dass die prozentualen Anteile der Blickzeiten bzw. der Übergangszeiten an der Gesamtdauer besonders interessant für die Beurteilung der Gestaltung sind.

Beispiel

Das Beispiel zur Operationalisierung der kumulierten Übergangszeiten stammt aus der Bearbeitungsmaske zum Warenkorb (Abbildung 14). Im Warenkorb prüft der Besteller den/die Artikel vom Besteller auf die Richtigkeit (Artikel, Menge, Preis). Wenn die Angaben für richtig befunden werden, bestellt er die Warenkorb-Artikel über den Button „Anfordern“.

Im Beispiel werden für diese Untersuchung zwei Blickobjekte identifiziert:

- **Blickobjekt 1: Warenkorb-Liste.** Die Liste ist im oberen Bereich, direkt unterhalb des Titels angeordnet. Sie besteht aus mehreren Spalten wie Position, Artikel-Nr., Bild, Bezeichnung, Nettowert, Menge etc. Die Liste erstreckt sich horizontal über die gesamte Breite der Maske.
- **Blickobjekt 2: Buttonleiste.** Die Buttonleiste befindet sich am unteren Rand der Maske. Der Button „Anfordern“ löst die Bestellung aus.

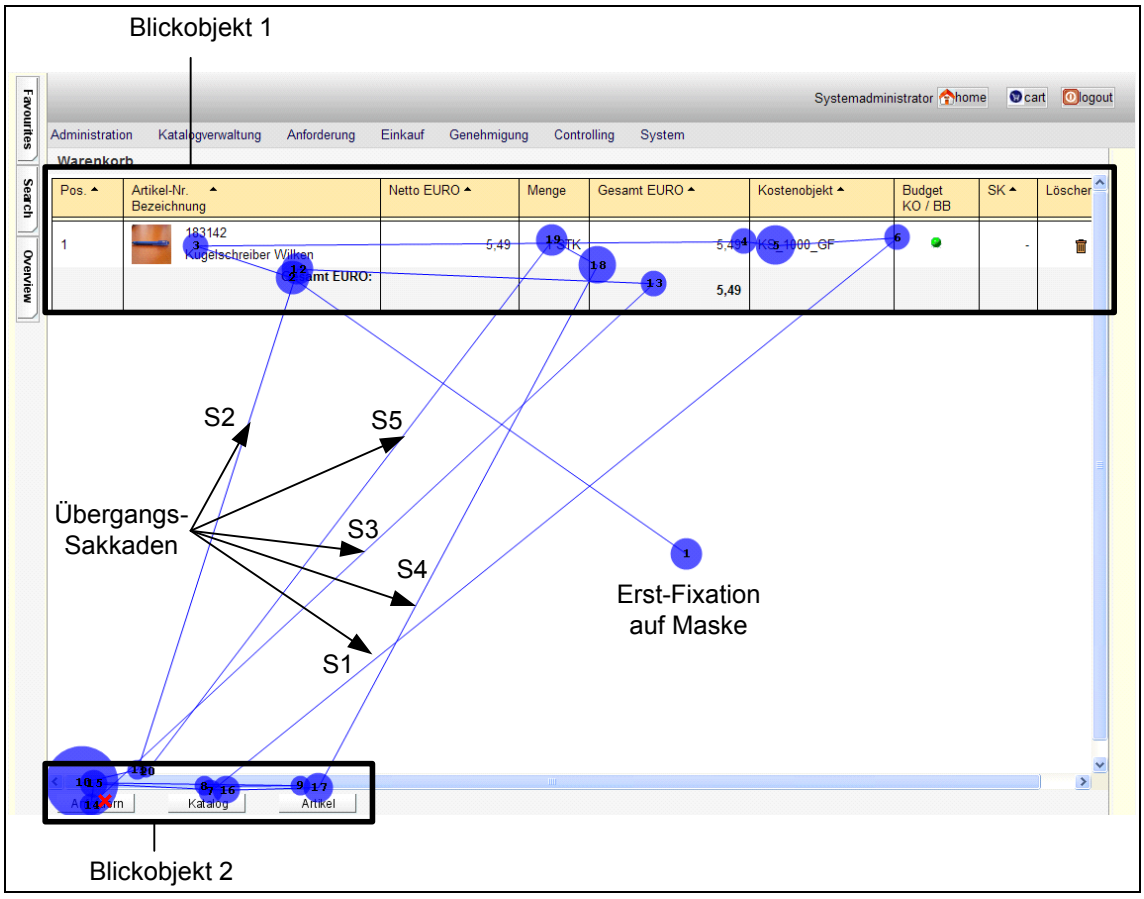

Abbildung 14: Beispielmaske kumulierte Übergangszeiten

Um die Artikeldaten des im Warenkorb befindlichen Artikels zu verifizieren, wandert der Blick wiederholt zwischen den beiden Blickobjekten hin und her. Für dieses kontrollierende Verhalten ist die Gestaltung von großer Bedeutung. Da hier die beiden Objekte (zu betrachtendes und interaktives Objekt) an weit voneinander entfernten Positionen liegen, entstehen große Sakkadenweiten bzw. lange Sakkadenzeiten. Hinzu kommt, dass der prüfende Blickpfad nicht nur einmal, sondern mehrmals zwischen den beiden Blickobjekten hin und her verläuft. In diesem Beispiel finden sogar fünf Übergangs-Sakkaden statt. Die erste Sakkade, das ist die Sakkade zwischen der Erst-Fixation auf der Maske und der ersten Fixation in einem der Blickobjekte, wird nicht mitgerechnet. Die Erst-Fixation ist von dem Endpunkt auf der vorherigen Maske abhängig und damit für die aktuelle Auswertung irrelevant.

Berechnung des Parameters

In der ClearView-Software wird die Sakkadendauer nicht ausgegeben. Diese Angabe lässt sich prinzipiell aus dem zeitlichen Abstand zwischen zwei Fixationen abzüglich der Fixationsdauer der ersten Fixation errechnen. Diese Angabe kann entweder aus den

statischen Plot-Darstellungen ermittelt werden oder aus der Fixations-Exportdatei. Erstere erwies sich als effizienter, da zunächst die relevanten Fixationen ermittelt werden mussten. Diese konnten durch manuelles Abfahren des Blickverlaufs leicht ermittelt werden. Im Beispiel: Zwischen den Fixationen 6 und 7 liegt die erste Übergangs-Sakkade S1, zwischen den Fixationen 11 und 12 die zweite Sakkade S2 und so fort. Anschließend konnten pro Erst-Fixation der Endezeitpunkt und pro Folge-Fixation der Startzeitpunkt ermittelt werden, deren Differenz die Sakkadendauer der Übergangs-Sakkade ist.
Berücksichtigt wird die Blickpfaddauer bis zum Mausklick. Da die Software mit leichter Zeitverzögerung reagierte, fanden auch nach dem Mausklick aus Gründen der Verunsicherung, ob die Software bereits arbeitet oder nicht, mehrfach wiederholte Übergangs-Sakkaden statt.
Die Blickobjekte, zwischen denen die im Fokus stehenden langen Übergangs-Sakkaden stattfinden, entsprechen den Interessengebieten. Die kumulierten Übergangszeiten werden für jede Testperson ermittelt, d. h. für jede einzelne Aufzeichnung. Eine Summierung über alle Aufzeichnungen findet nicht statt. Der Grund liegt zum einen in dem sehr hohen manuellen Auswertungsaufwand, zum anderen beeinflusst es die Aussage zur Gestaltungsqualität kaum, wie Stichproben zeigten.

Weiterhin wurde bei der Ermittlung der Daten von folgenden Annahmen ausgegangen:

- Aufzeichnungen ausschließen, die großen Versatz zeigen.
- Nicht interpretiert wurden die Aufzeichnungen, die keine Übergänge enthielten. Das trifft jedoch nur bei ca. 5 % der Aufzeichnungen zu. Diese Personen wollten möglichst bald den Vorgang abschließen, ohne den Wareneingang nochmals zu kontrollieren.
- Es wurden auch diejenigen Aufzeichnungen berücksichtigt, bei denen der Klick nicht auf den „Anfordern“-Button, sondern auf einen der beiden anderen Buttons erfolgte. Die Wahl des Buttons war für die Analyse der Verarbeitungsanforderungen auf dieser Maske nicht ausschlaggebend.
- Nicht als Übergangs-Sakkade wurde gewertet, wenn der Benutzer von der Buttonleiste in gleicher Höhe den Blick auf die rechte untere Bildschirmecke wandte. Diese Benutzer erwarteten dort entweder weitere Buttons oder wollten den Scrollbalken bewegen. Dort befand sich jedoch kein Objekt, auch musste nicht

gescrollt werden. Es handelte sich hierbei zwar um Sakkaden, jedoch führten sie zu keinem Blickobjekt, weshalb sie nicht gewertet wurden.

Ergebnisse

Der Parameter der kumulierten Übergangszeiten fällt in dem Beispiel mit dem Warenkorb besonders auf. Zum einen erstrecken sich die Sakkadenweiten über die gesamte vertikale Breite. Dabei entstehen vertikale und diagonale Übergangs-Sakkaden. Zum anderen kumuliert sich pro Aufzeichnung die Übergangszeit aufgrund der mehrfach auftretenden Sakkaden (hier: fünf Sakkaden). Tabelle 12 zeigt zunächst die einzelnen Sakkaden und deren Dauern.

Tabelle 12: Übergangszeiten beim Warenkorb

Übergangszeiten der Übergangs-Sakkaden	**Sakkade 1**	**Sakkade 2**	**Sakkade 3**	**Sakkade 4**	**Sakkade 5**
Ende-Zeitstempel der Fixation n	164.682 ms	166.835 ms	167.473 ms	168.429 ms	169.167 ms
Start-Zeitstempel der Fixation n+1	165.180 ms	167.074 ms	167.772 ms	168.768 ms	169.705 ms
Sakkadendauer	498 ms	239 ms	299 ms	339 ms	538 ms

Bei den ermittelten Sakkadendauern beträgt das Minimum 239 Millisekunden bis 538 Millisekunden im Maximum. Die kumulierte Übergangszeit beträgt 1913 Millisekunden. Im Vergleich zur Mindestdauer einer Sakkade, größer als 30 Millisekunden, haben die im Beispiel ermittelten Sakkaden eine extrem lange Dauer.

In der nächsten Ergebnistabelle werden die Zeiten summiert und die Anteile berechnet (Tabelle 13). Die kumulierte Übergangszeit der Übergangs-Sakkaden beträgt 1913 Millisekunden. Diese hat einen Anteil von knapp 30 % an der Gesamtzeit. Ein Drittel der gesamten Betrachtungszeit dient demnach nicht zur Informationsaufnahme, sondern muss für die visuelle Suche und Orientierung verwendet werden. Analog dazu werden zwei Drittel der Zeit (71,1 %) für Blickzeiten mit bewusster Informationsaufnahme verwendet. Ein Teil davon sind jedoch auch Sakkadenzeiten. Dadurch verringert sich die tatsächliche Fixationsdauer und Informationsaufnahme noch mehr und liegt bei weniger als zwei Drittel.

Tabelle 13: Anteile der Übergangszeiten und Blickzeiten beim Warenkorb

Zeitdauern	**Zeiten [ms]**	**Anteil [%]**
Kumulierte Übergangszeit der großen Sakkaden	1913	28,9 %
Gesamtzeit *)	6611	100 %
Blickzeiten **)	4698	71,1 %

*) Gesamtzeit = Fixationszeiten + Übergangs-Sakkadenzeiten + Blick-Sakkadenzeiten

**) Blickzeiten = Fixationszeiten + Blick-Sakkadenzeiten

Zwei weitere Ergebnisse sind erwähnenswert und sollen kurz erläutert werden. Wie eingangs erwähnt, wurde keine Summierung der Übergangszeiten über alle Testpersonen pro Warenkorb-Maske vorgenommen. Der einzig veränderliche Parameter bei der Einbeziehung aller Recordings ist die Anzahl der Übergangs-Sakkaden. Es wurde jedoch untersucht, welche Bandbreite diese Anzahl aufweist. Die Anzahl variiert von ein bis maximal 13 Übergänge. 30 % der Testpersonen benötigten nur ein oder zwei Übergänge, den Großteil der Aufzeichnungen bilden 3 bis 5 Übergänge mit einem Anteil von 63 %. Ein geringer Anteil von 7 % benötigte 6 und mehr Übergänge. Das ausführlich vorgestellte und berechnete Beispiel (5 Übergänge) liegt demnach in dem Hauptbereich der Anzahl Übergänge.

Das zweite Ergebnis ist, dass aufgrund der großen Sakkadenweite ca. 10 % der Testpersonen „Stütz-Fixationen“ benötigen. Diese Fixationen unterbrechen die lange Sakkade an einer Stelle, meist in der räumlichen Mitte.

3.3.3 Relative Häufigkeit der Fixationen von Objekten

Die relative Häufigkeit der Fixationen von Objekten gibt Aufschluss darüber, welche Objekte häufiger als andere betrachtet wurden. Es lässt sich dabei auch ermitteln, welche Bereiche keine Aufmerksamkeit erhalten. Die Maske wird in Blickgebiete eingeteilt. Ein Blickgebiet kann der Navigationsbereich, der Titelbereich oder auch nur eine Überschrift im Text sein. Anschließend werden für diese Bereiche die Fixationen bestimmt. Die relative Häufigkeit der Fixationen auf den verschiedenen Blickgebieten lässt sich so berechnen. Üblicherweise wird die Anzahl der Fixationen eines Objektes im Verhältnis zur Gesamtzahl der Fixationen während eines bestimmten Zeitraumes angegeben.

Beispiel

Die relative Häufigkeit der Fixationen ließe sich an mehreren Beispielen darstellen. Es ist fast für jede Maske von besonderem Interesse, welche Bereiche prozentual wie häufig betrachtet werden. Die Aufgabenstellung nimmt grundsätzlich eine herausragende Rolle bei dem Blickverhalten ein. Dieser Häufigkeits-Parameter ist besonders aussagekräftig bei sehr konkreten Aufgaben wie zum Beispiel bei einer Suchaufgabe. Der Grund liegt darin, dass der Fokus der Wahrnehmung bei diesen Aufgaben sehr konzentriert und von der Kognition her gesteuert ist und nur die Blicke stattfinden, die zwingend für die visuelle Suche notwendig sind. Für die Gestaltung der Elemente und deren Positionen in den Blickbereichen lassen sich daraus Optimierungspotenziale ableiten. Es können Gestaltungslösungen gefunden werden, die effizienter sind und bei denen Informationen mit weniger Fixationen aufgenommen werden können.

Bei anderen Aufgaben, wie der Datenerfassung in einem Formular, ist es oft erwünscht, dass der Blick fast den gesamten Stimulusbereich systematisch exploriert. Für diesen Fall ist der Häufigkeitsparameter weniger aussagekräftig, da er gleichmäßige Verteilungen über die einzelnen Blickbereiche liefert und damit keine Gewichtung zulässt.

Um die relative Häufigkeit der Fixationen zu operationalisieren, wurde hier der Katalogbaum ausgewählt (Abbildung 15). Der Baum ist auf der zweiten Ebene aufgeklappt und zeigt 13 Einträge, von denen nur ein Eintrag richtig ist. Mittels dieses Parameters kann das Suchverhalten beim Leseverhalten von Bäumen und Baumeinträgen detailliert untersucht werden. Baumstrukturen verfügen generell über eine bestimmte Tiefe und eine bestimmte Breite. Im Idealfall ist ein Optimum zu finden zwischen Tiefe und Breite eines Baumes. In diesem Beispiel interessiert insbesondere, wie die einzelnen Einträge einer Hierarchieebene vom Benutzer gelesen werden. Dabei wurde der Baum derartig aufgeklappt gezeigt, dass möglichst viele Einträge sichtbar sind (14 Einträge).

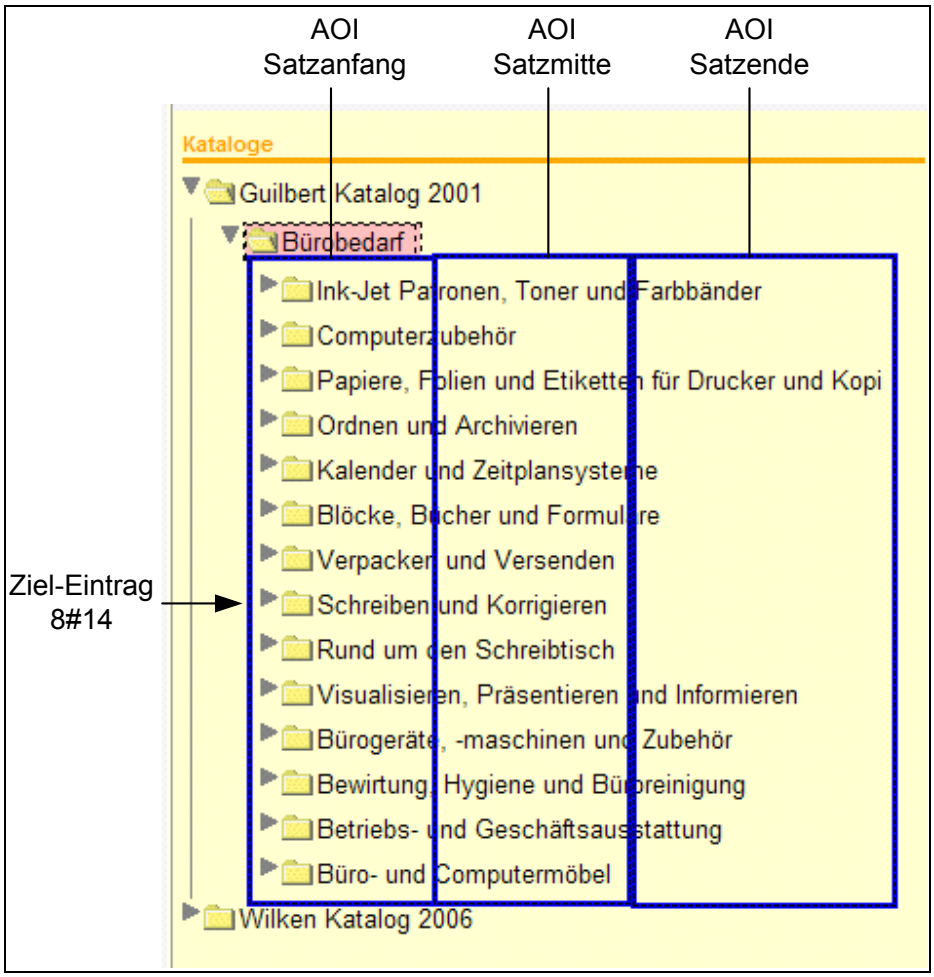

Abbildung 15: Beispielmaske Baum mit AOIs

Der Häufigkeitsparameter macht dabei zwei Aspekte deutlich. Zum einen veranschaulicht er, mit welcher Häufigkeit die 14 Einträge im vertikalen Verlauf (oben vs. unten) betrachtet werden. Das entspricht der Breite des Baumes, d. h. mit anderen Worten, wie viele Einträge auf einer Ebene vorhanden sind. Dieser Aspekt ist unter der Prämisse der Aufgabenstellung zu analysieren. Demnach ist der richtige Ziel-Eintrag der achte von vierzehn Einträgen. Die Testpersonen, die den achten Eintrag beim ersten Fixieren als Ziel-Eintrag interpretiert und angeklickt haben, betrachteten den Baum nicht weiter nach unten. Die Häufigkeitsverteilung über alle Einträge ist damit nicht aussagekräftig, da die Einträge gemäß Aufgabenlösung nicht gleichartig relevant waren. Interessanter für die Usability-Bewertung ist bei der Untersuchung der Anzahl der Einträge die Fixations-Reihenfolge. Werden die Einträge von oben nach unten sukzessive betrachtet oder finden dazwischen Sprünge statt? Dies lässt sich mit den Parametern veranschaulichen, die Übergänge und Richtungswechsel verdeutlichen.

Zum anderen lässt sich am Beispiel des Katalogbaumes untersuchen, wie sich die Aufmerksamkeit beim Lesen eines Eintrages im horizontalen Verlauf (links vs. rechts) verhält. Die Einträge sind unterschiedlich lang. Die Länge variiert zwischen einem einzigen Wortkompositum („Computerzubehör“) bis hin zu acht Wörtern („Papiere, Folien und Etiketten für Drucker und Kopi…“). Der Großteil der Einträge besteht aus zwei bis drei Wörtern („Schreiben und Korrigieren“, „Büro- und Computermöbel“ etc.). Es lassen sich damit drei Abstufungen – kurz, mittel, lang – festlegen. Zum Zweck der Längenanalyse wurden über die Einträge darum drei „Areas of Interest“ (AOIs) gelegt: AOI Satzanfang, AOI Satzmitte und AOI Satzende. Der Satzanfang umfasst das erste Substantiv, die Satzmitte das zweite Substantiv mit Präposition und das Satzende weitere Wörter, sofern vorhanden.

Berechnung des Parameters

Die Daten für die beiden Untersuchungsgegenstände – vertikale und horizontale Blickhäufigkeiten – werden auf zwei verschiedene Arten ermittelt.

Für die vertikale Häufigkeitsverteilung der Blicke wurde die visuelle Darstellungsform des Hotspots betrachtet. Die Auswertungssoftware wurden so eingestellt, dass die Verteilungen mit Farben, von grün über gelb und orange nach rot, sichtbar gemacht werden. Rot eingefärbte Bereiche wurden mindestens 20-mal fixiert. Graue Bereiche sind ohne Fixationen. Eine quantitative Auswertung für die vertikale Verteilung findet nicht statt. Diese wäre möglich, wenn man den Baum der Länge nach in mehrere AOIs unterteilen würde. In diesem Fall wurde darauf verzichtet, da das Ergebnis wie oben erläutert nicht ergiebig ist.

Um das Leseverhalten pro Eintrag zu analysieren, wurde die AOI-Analyse eingesetzt. Mit der AOI-Analyse der ClearView-Software können die Fixationen innerhalb eines AOIs ermittelt werden und daraus die relative Häufigkeit der Fixationen errechnet werden. Mittels des AOI-Tools konnten drei AOI-Rechtecke für Satzanfang, Satzmitte und Satzende eingezeichnet werden. Anschließend wurden Aufzeichnungen und AOIs nach Excel exportiert und dort mit Hilfe eines Templates und mit Makros ausgewertet. Dabei mussten für den direkten Excel-Export einige Einschränkungen getroffen werden (Tobii 2006). So konnten nur ca. 30 Aufzeichnungen mit einem einzigen Export exportiert werden. Um alle Aufzeichnungen auszuwerten, müsste der Export wiederholt und anschließend die Ergebnisse summiert werden. In diesem Fall wurde festgestellt, dass sich

die vier Exports, die jeweils 30 Aufzeichnungen umfassten, von der Häufigkeitsverteilung her sehr ähnlich waren und darum wurde auf die Summierung verzichtet. Ein weiteres Problem stellt die Frame-Bildung von ClearView dar. Jeder Maskenzustand, auch bei gleichen Inhalten, erzeugt einen eigenen Frame. Eine Maske, bei der das Pfeilsymbol anstelle des Ordnersymbols des Oberbegriffes angeklickt wurde, ergibt zwei verschiedene Frames. Insgesamt gibt es 12 Frames, die den aufgeklappten Baum zeigen. AOIs hingegen können nur je Frame definiert werden. Die AOIs mussten also für jeden Frame dupliziert werden. Trotz dieser Schwierigkeiten konnte eine aussagefähige Auswertung über 13 Aufzeichnungen auf dem gleichen Frame getroffen werden.

Für die Berechnung dieses Parameters waren die Mausklicks auf Baumeinträgen nicht relevant. Der Grund liegt darin, dass die relative Häufigkeit der Fixationen sich auf mehrere Aufzeichnungen bezieht und in jeder Aufzeichnung der Mausklick zu einem unterschiedlichen Zeitpunkt stattfand. Das Interaktionsverhalten und seine zeitliche Eingrenzung auf Entscheidungen, d. h. Mausklicks, lassen sich mit dem Häufigkeitsparameter der Fixationen nicht bestimmen.

Ergebnisse

Das Ergebnis zeigt den Hotspot über die Aufzeichnungen, bei denen die Benutzer den Ziel-Eintrag „Schreiben und Korrigieren“ im Katalogbaum suchten (Abbildung 16). Es ließen sich dabei als Stichprobe insgesamt 20 vergleichbare Aufzeichnungen zugrunde legen. Nur bei diesen Aufzeichnungen lag das gleiche Bildschirmmaterial vor, das heißt der ab „Bürobedarf“ aufgeklappte Guilbert-Katalogbaum bei gleichzeitig geschlossenem Wilken-Katalog. Die geringe Stichprobe von 20 Aufzeichnungen liegt darin begründet, dass nur ein Teil der Benutzer die Katalogsuche durchführte, andere bedienten sich der Volltextsuche. Da der Baum von den Testpersonen an unterschiedlichen Stellen aufgeklappt wurde, ergab sich daher eine weitere Einschränkung.

Der Hotspot lässt erkennen, dass die häufig betrachteten Stellen zu Beginn der Textabschnitte liegen und im Verlauf der Zeilen nach rechts abnehmen. Dieses vertikale Blickverhalten wird durch die zweite Auswertung noch detaillierter untersucht.
Weiterhin ergibt der Hotspot, dass die oberen ca. zehn Einträge sehr häufig betrachtet werden, die letzten vier dagegen weniger häufig bis selten. Der letzte Eintrag in der Liste ist nach links ausgerückt, da dies ein übergeordneter Punkt ist. Dieser wird nicht

mehr betrachtet. Die Ursache für das nach unten nach zehn Einträgen abnehmende Blickverhalten liegt darin, dass der achte Eintrag der gesuchte Eintrag ist und die visuelle Suche für viele Benutzer damit beendet war.

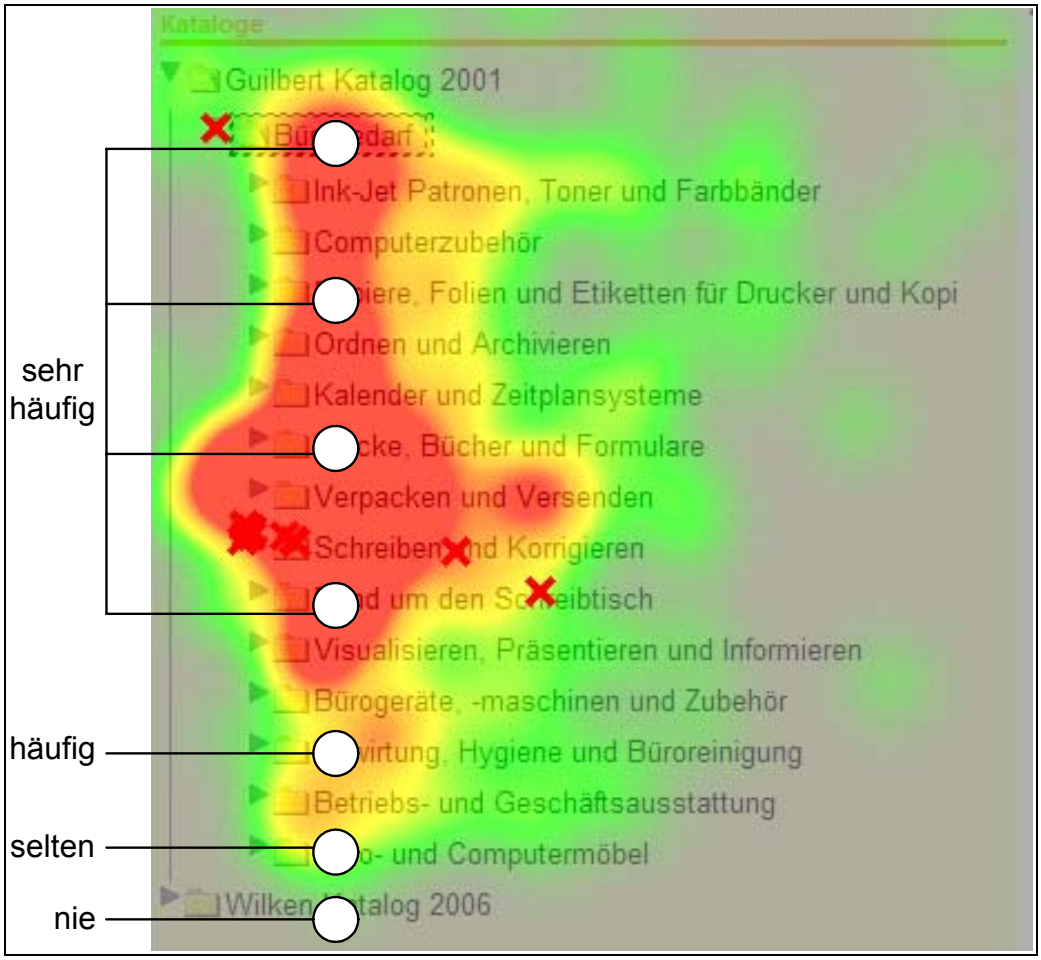

Abbildung 16: Hotspot-Katalogbaum

Das Ergebnis des Leseverhaltens pro Eintrag zeigt die AOI-Analyse noch detaillierter. Tabelle 14 beschreibt die deskriptive Statistik der Anzahl der Fixationen je Objekt. In den ersten Spalten werden Minimum, Maximum, Summe, Mean-Wert und Median-Wert aufgeführt. Die letzte Spalte fasst die relative Häufigkeit der Fixationen pro Objekt zusammen.

Das zentrale Ergebnis dabei ist, dass der Satzanfang, also das erste Wort, mit 70 % relativer Häufigkeit das am häufigsten betrachtete Objekt ist. Auf die Satzmitte fallen 26 % Häufigkeit und zum Satzende hin nimmt der Fixationsanteil deutlich ab (4 %).

Tabelle 14: Ergebnistabelle Relative Häufigkeit von Fixationen

Fixationen	Anzahl Fixationen					Relative Häufigkeit
	Min [1]	Max [2]	Sum [3]	Mean [4]	Median [5]	Prozent(%)[6]
Satzanfang	3	61	299	23	20	70 %
Satzmitte	1	39	113	8,7	6	26 %
Satzende	0	5	17	1,3	0	4 %

1) Min: Minimale Anzahl der Fixationen einer Aufzeichnung (von 13 Aufzeichnungen)

2) Max: Maximale Anzahl der Fixationen einer Aufzeichnung (von 13 Aufzeichnungen)

3) Sum: Summe aller Fixationen von allen 13 Aufzeichnungen

4) Mean: Mittelwert Anzahl Fixationen aller 13 Aufzeichnungen

5) Median: Medianwert Anzahl Fixationen aller 13 Aufzeichnungen

6) Relative Häufigkeit der Fixationen pro Objekt

Zum Vergleich werden die Blickzeiten während der Fixationen detailliert untersucht (Tabelle 15). Die zeitlichen Ergebnisse korrelieren mit den anzahlbezogenen Ergebnissen. Die Verteilung ist noch deutlicher ausgeprägt, d. h. 70 % der relativen Fixationen nahmen 76,5 % der relativen Blickdauer in Anspruch. Da sich die Verteilung der Betrachtungsintensität je AOI nicht nur mit der Anzahl der Fixationen belegen lässt, sondern auch die zeitliche Auswertung zum gleichen Ergebnis kommt, kann dieses Ergebnis als valide betrachtet werden.

Tabelle 15: Ergebnistabelle Relative Blickzeitdauern

Blickdaten	Blickdauer [ms]					Relative Blickdauer
	Min [1]	Max [2]	Sum [3]	Mean [4]	Median [5]	Prozent(%)[6]
Satzanfang	1853	25494	102592	7891,69	6756,0	76,5
Satzmitte	219	10161	26827	2063,62	1215,0	20
Satzende	0	1156	4623	355,62	0,0	3,5

1) Min: Minimale Blick-Zeitdauer einer Aufzeichnung (von 13 Aufzeichnungen)

2) Max: Maximale Blick-Zeitdauer einer Aufzeichnung (von 13 Aufzeichnungen)

3) Sum: Summe aller Blick-Zeitdauern (von 13 Aufzeichnungen)

4) Mean: Mittelwert der Blick-Zeitdauern (von 13 Aufzeichnungen)

5) Median: Medianwert Blick-Zeitdauern (von 13 Aufzeichnungen)

6) Relative Blickdauer pro Objekt

3.3.4 Richtungsänderungen des Blickpfades

Die Definition von Richtungsänderungen des Blickpfades sieht vor, die Richtung direkt aufeinanderfolgender Sakkaden zu operationalisieren. Der Winkel α liegt zwischen zwei aufeinanderfolgenden Sakkaden. Falls α kleiner 90° beträgt, zählt die zweite der beiden Sakkaden als rückläufige Bewegung, auch „backtrack“ genannt. Falls α größer 90° ist, zählt die zweite Sakkade als Vorwärts-Bewegung, auch „forwardtrack“ genannt. In Abbildung 17 symbolisiert jeder Kreis eine Fixation. Zwischen zwei Fixationen liegt eine Sakkade, deren Richtung durch einen Pfeil angedeutet ist. Die Sakkade zwischen Fixation n und Fixation n+1 weicht von der vorherigen Sakkade in einem Winkel kleiner 90° ab. Man spricht dabei von einer Richtungsänderung. In diesem Fall führt die Fixationsfolge einen „backtrack“ aus. Im Gegensatz dazu sollen hier Sakkaden, die einen Winkel größer 90° beschreiben, mit „forwardtrack“ bezeichnet werden. Diese Sakkaden setzen die Richtung der jeweils vorhergehenden Sakkaden fort.

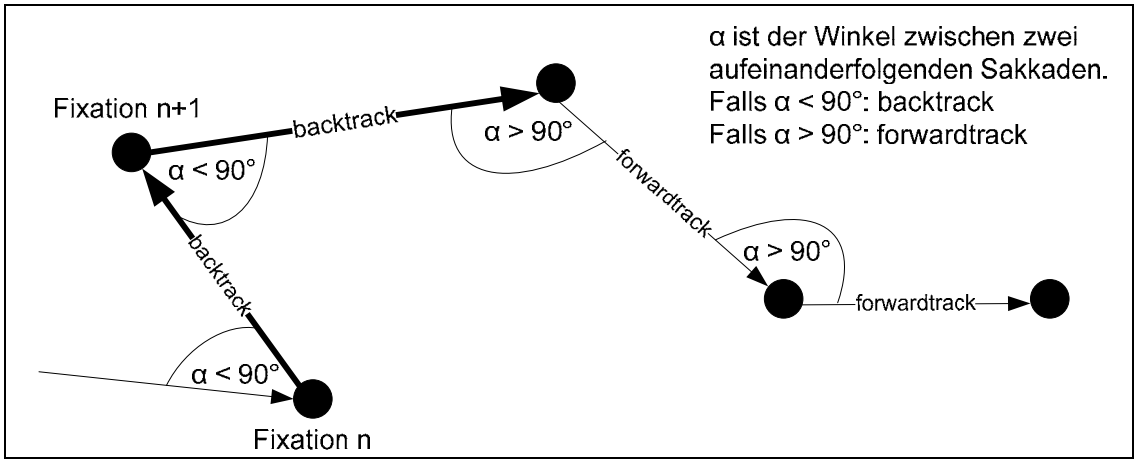

Abbildung 17: Richtungsänderungen (backtrack, forwardtrack)

Variation der Definition

Bei der explorativen Durchsicht der Experimentdaten fielen speziell bei den Blickpfaden auf den Buttons deutliche Richtungsänderungen auf. In der dritten von drei Buttonleisten, verteilt auf je eine Maske, wurde die Reihenfolge der Buttons gravierend geändert und dies war der Auslöser für eine Vielzahl von Richtungsänderungen.

Aufgrund der näheren Erforschung der Richtungsänderungen bei den Buttons, ist es jedoch notwendig, die oben genannte Definition weiter zu fassen. Zum einen liegen nicht ausschließlich Einzelfixationen zwischen den richtungsändernden Sakkaden. Der Blick setzt sich gegebenenfalls aus mehreren Fixationen zusammen, bevor eine größere Sakkade die Richtungsänderung auslöst. Zu diesem Zweck verwendet diese Auswer-

tung den Begriff „Blick“, der weiter oben bereits von „Fixation“ unterschieden wurde. Ein Blick ist die Sequenz mehrerer Fixationen auf demselben Objekt, hier: auf einem Button. Abbildung 18 veranschaulicht diese Variation.

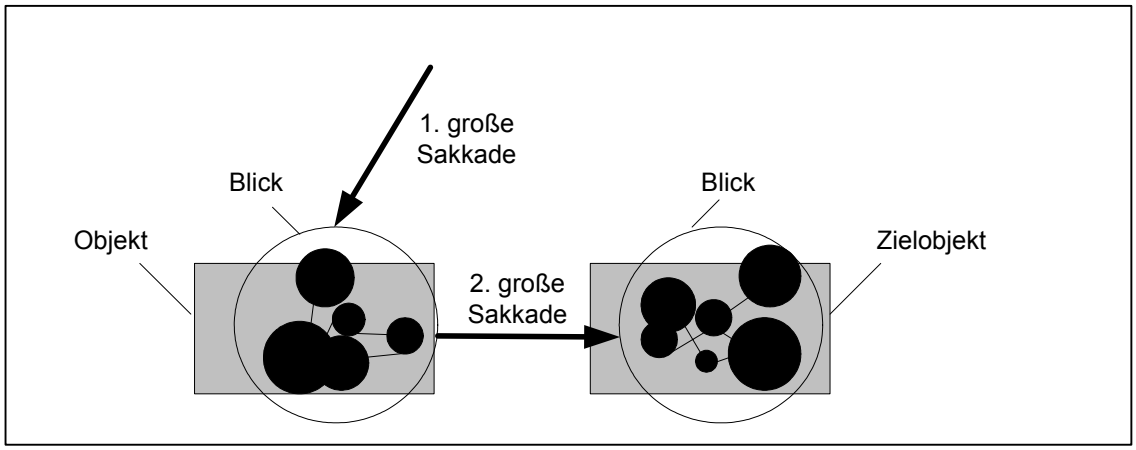

Abbildung 18: Richtungsänderung Variation

Eine weitere Variation besteht darin, dass die Richtungsänderungen nicht ausschließlich zwischen aufeinanderfolgenden Blicken stattfinden müssen. Richtungsänderungen können auch in einem größeren Zusammenhang zutage treten, wenn Benutzer etwa spezielle Strategien anwenden. Die Richtungsänderungen folgen dann nicht unmittelbar aufeinander, sondern werden beispielsweise in verschiedenen Masken bei gleicher Problemlösungssituation angewendet. Eine Strategie könnte lauten: alle Buttons gemäß der Leserichtung von links nach rechts betrachten. Diese Strategie verfolgt einen Algorithmus, der mit Sicherheit zum Ziel führt. Wenngleich er nicht immer die schnellste Lösung darstellen muss. Eine andere Strategie, basierend auf Erfahrung, könnte lauten: Wenn in anderen Masken oder Software meist der rechte Button weiterführt und der linke zurückführt, wird gleich der rechte Button betrachtet.

Unabhängig von der angewandten Strategie ist immer der Strategiewechsel von Bedeutung. Wechsel in der Strategie treten auf, wenn die Software beispielsweise inkonsistent und nicht erwartungsgemäß gestaltet ist oder unerwartete Änderungen enthält. Es wird darum vermutet, dass sich auch Strategiewechsel durch Richtungsänderungen andeuten. Sowohl direkte als auch strategische Richtungsänderungen sollen in diesem Abschnitt analysiert werden.

Beispiel

Das Beispiel, das die Richtungsänderungen veranschaulicht, bezieht sich auf die Navigation über die Buttons. Die Buttons dienen dazu, Funktionen auszulösen. Nach dem Ausfüllen aller notwendigen Daten auf der aktuellen Maske kann mittels eines Buttons eine Verarbeitung initiiert werden und in den meisten Fällen wird eine neue Maske angezeigt. Die Buttons befinden sich erwartungsgemäß im unteren Softwarebereich. Die Buttons sind mit Text beschriftet, z. B. „Zurück“, „Weiter“, „Zum Warenkorb“.

Die Richtungsänderung soll mittels drei aufeinanderfolgenden Masken analysiert werden.

- **Maske 1 „Anforderung“** (Abbildung 19)**:** Die Buttons lauten von links nach rechts: „Zurück“ und „Weiter“. Gemäß der Leserichtung von links nach rechts betrachten die **meisten Anwender** zuerst kurz den linken Button, stellen aber rasch fest, dass dieser nicht der gesuchte Button ist. Der Blick wandert mit einer großen Sakkade weiter nach rechts zum nächsten Button. Dieser ist der gesuchte Button. Er wird angeklickt. Die Blickrichtung zwischen linkem und rechtem Button ist eine mehr oder weniger horizontale Gerade.

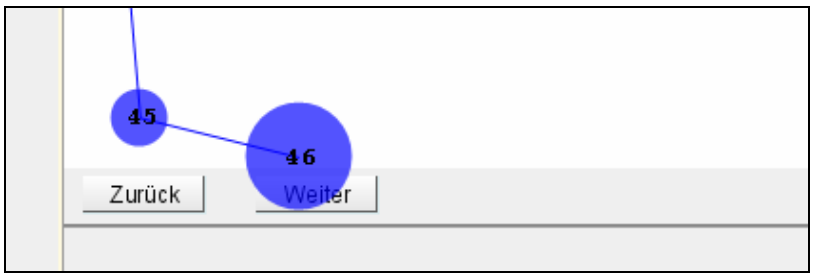

Abbildung 19: Beispielmaske Buttons forward-Blickfolge

- **Maske 2 „Kontierung“** (Abbildung 20)**:** Die Maske „Kontierung“ folgt direkt auf die vorherige Maske „Anforderung“. Die Buttons lauten von links nach rechts: „Zurück“ und „Zum Warenkorb“. Der richtige Button ist der rechte Button „Zum Warenkorb“. Die **Mehrzahl der Anwender** wenden ihren Blick nun, im Gegensatz zur vorherigen Maske, sofort auf den rechten der beiden Buttons. Zwischen den Buttons ist in diesem Fall keine große Sakkade mehr notwendig. Der rechte Button wird angeklickt.

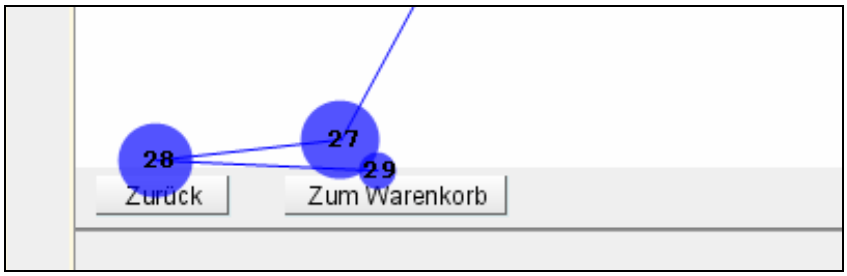

Abbildung 20: Beispielmaske 1 Buttons backtrack-Blickfolge

- **Maske 3 „Warenkorb“** (Abbildung 21)**:** Die Maske „Warenkorb“ folgt wiederum direkt auf die vorherige Maske „Kontierung“. Die Buttons lauten von links nach rechts: „Anfordern“, „Katalog“ und „Artikel“. Der linke Button „Anfordern“ ist der Button, um den Bestellvorgang fortzusetzen. Die beiden anderen Buttons führen zurück zur Katalog- bzw. Artikelauswahl. Die **meisten Anwender** beginnen ihre Blickfolge mit einem Blick auf den rechten Button. Aufgrund der unerwartet geänderten Button-Reihenfolge ist der weiterführende Button aber nicht mehr der jeweils rechte Button. Der Blick springt zurück zu dem links davon liegenden, mittleren Button. Da auch dieser falsch ist, erfolgt nochmals ein Sakkadensprung nach links bis zum linken, richtigen Button. Die beiden Sakkadensprünge haben eine Richtung, die der Blickrichtung in der ersten Maske um 180° bzw. 0° entgegengesetzt ist. Der linke Button „Anfordern“ wird angeklickt.

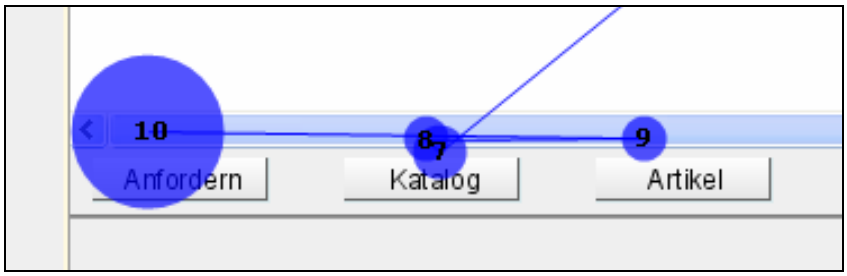

Abbildung 21: Beispielmaske 2 Buttons backtrack-Blickfolge

Berechnung des Parameters

Die Ermittlung der Richtungsänderungen von großen Sakkaden zwischen Blicken (bestehend aus mehreren Fixationen) auf „Areas of Interest“ stellt die Software ClearView nicht zur Verfügung. Dieser Parameter muss sehr sorgfältig aus den Fixationsdaten (FXD-Datei) und „Areas of Interest“ manuell erstellt werden.

Das Vorgehen war im Einzelnen wie folgt:

- Für jede Testperson und für jede der drei Masken die Reihenfolge der Blicke pro Button ermitteln. Die Einzelfixationen pro Blick wurden nicht ausgewertet. Das geschah über die statische Anzeige des Gaze-Plot. Die Fixationen in FXD-Datei waren den AOIs nicht eindeutig zuzuordnen, da die Buttons relativ nahe beieinander lagen.
- Aufzeichnungen ausschließen, die großen Versatz zeigen.
- Es wurden auch die Aufzeichnungen berücksichtigt, bei denen der Klick auf den falschen Button erfolgte. Eine Fehlerrate ist damit vorhanden.
- Aufzeichnungen, bei denen Klick auf einen Button ohne Blick erfolgte, wurden einbezogen.
- Wenn der Blick nochmals den unteren Buttonbereich verließ, dann aber wieder zu den Buttons zurückkehrte, wurde die Aufzeichnung dennoch gewertet. Die Zwischenblicke blieben unberücksichtigt.

Ergebnisse

Im Folgenden werden die drei Masken und darin jeweils die sukzessiven Blicke innerhalb der Buttons dokumentiert. Richtungsänderungen sind dabei ausgewiesen.

Maske 1#3 „Anforderung“: Zwei Buttons (Tabelle 16)

- 1: „Zurück“ links
- 2: „Weiter“ rechts (Zielobjekt)

Tabelle 16: Ergebnistabelle 1 Richtungsänderungen

Sequenz der Blicke	Start	Anzahl Blicke	Anzahl Testpersonen
1	1	1	2
1-2	1	2	39
1-2-1-2 *)	1	4	5
Zwischensumme Start 1	Varianten: 3	Mittelwert: 2,3	Summe: 46 (66 %)
2	2	1	9
2-1-2 *)	2	3	14
2-1-2-1-2 *)	2	5	1
Zwischensumme Start 2	Varianten: 3	Mittelwert: 3	Summe: 24 (34 %)
Summen	Varianten: 6	Mittelwert: 2,65	70

*) Blickfolgen mit Richtungsänderungen (backtracks)

In der ersten Maske, in der die Benutzer einen Button bedienen mussten, starteten ca. zwei Drittel der Testpersonen (66 %) mit dem Button ganz links (1). Das entspricht der Leserichtung von links nach rechts. Ein Großteil davon wiederum (39 von 46) benötigten nur einen weiteren Blick zum Button rechts, um diesen als Zielbutton zu erkennen. Die restlichen Probanden vollzogen nochmals einen prüfenden Blick zurück zum ersten und dann endgültig auf den zweiten Button.

Ein Drittel der Testpersonen (34 %) betrachtete als ersten Button den richtigen Button rechts. Davon vollzogen jedoch zwei Drittel der Testpersonen einen Korrekturblick, indem sie nochmals Button 1 und dann wieder Button 2 anblickten.
Insgesamt gab es bei dieser Buttonleiste sechs Varianten von Blickfolgen. Die Hälfte davon enthält backtracks. Der Mittelwert der Blicke pro Blickfolge beträgt weniger als drei Blicke.

Maske 2#3 „Kontierung“: Zwei Buttons (Tabelle 17)

- 1: „Zurück“ links
- 2: „Zum Warenkorb“ rechts (Zielobjekt)

Tabelle 17: Ergebnistabelle 2 Richtungsänderungen

Sequenz der Blicke	**Start**	**Anzahl Blicke**	**Anzahl Testpersonen**
1	1	1	2
1-2	1	2	11
Zwischensumme Start 1	Varianten: 2	Mittelwert: 1,5	Summe: 13 (20 %)
2	2	1	25
2-1-2 *)	2	3	25
2-1-2-1-2 *)	2	5	1
Zwischensumme Start 2	Varianten: 3	Mittelwert: 3	Summe: 51 (80 %)
Summen	Varianten: 5	Mittelwert: 2,25	64

*) Blickfolgen mit Richtungsänderungen (backtracks)

Von den Benutzern begannen in der zweiten Maske nur 20 % ihre Blickfolge mit dem ersten Button links und 80 % mit dem rechten (richtigen) Button. Das steht im deutlichen Gegensatz zur vorherigen Maske, in der zwei Drittel mit dem ersten Button starteten – gemäß der Leserichtung von links nach rechts. Eine Erklärung dafür kann sein, dass die Benutzer aus der ersten Maske gelernt haben, dass der weiterführende Button

rechts platziert ist. Da sie bei dieser Teilaufgabe auch im Prozess vorwärts arbeiten wollten und nicht zurück, fiel der erste Blick auf den rechten Button.
Allerdings gab es bei fast der Hälfte der Aufzeichnungen, bei denen zuerst der rechte Button betrachtet wurde, eine Korrektur-Blickfolge zu den anderen Buttons.
Insgesamt gab es bei dieser Buttonleiste fünf Varianten von Blickfolgen, das heißt eine weniger als in der vorhergehenden Maske. 40 % davon enthalten backtracks. Der Mittelwert der Blicke pro Blickfolge beträgt nur etwas mehr als zwei Blicke.
Da die Anzahl der Buttons bei der zweiten Maske konstant blieb, ebenso die Positionierung der Vor- und Zurück-Buttons, zeigt sich eine Verbesserung in der Bedienung.

Maske 3#3 „Warenkorb“: Drei Buttons (Tabelle 18)

- 1: „Anfordern“ links (Zielobjekt)
- 2: „Katalog“ Mitte
- 3: „Artikel“ rechts

Tabelle 18: Ergebnistabelle 3 Richtungsänderungen

Sequenz der Blicke	Start	Anzahl Blicke	Anzahl Testpersonen
1	1	1	12
1-2	1	2	2
1-2-1 *)	1	3	1
1-2-3	1	3	1
1-2-3-1 *)	1	4	2
1-3-1 *)	1	3	1
Zwischensumme Start 1	Varianten: 6	Mittelwert: 2,2	Summe: 19 (28 %)
2-1	2	2	7
2-1-2-3-1 *)	2	4	1
2-1-3 *)	2	3	3
2-1-3-1 *)	2	4	2
2-1-3-2-1 *)	2	5	1
2-1-3-2-3-1-2-1 *)	2	8	1
2-3-1 *)	2	3	4
2-3-1-2-1 *)	2	4	1
2-3-1-2-1-2 *)	2	6	1
2-3-1-2-3-2-1 *)	2	7	1
2-3-1-3-1 *)	2	5	2
2-3-2-3-1 *)	2	5	1

Sequenz der Blicke	Start	Anzahl Blicke	Anzahl Testpersonen
2-3-2-3-1-2-1 *)	2	7	1
2-3-2-3-1-2-3-1 *)	2	8	1
Zwischensumme Start 2	Varianten: 14	Mittelwert: 5	Summe: 27 (40 %)
3-2-1	3	3	6
3-1-2 *)	3	3	4
3-1	3	2	3
3-2-3-1-2-1 *)	3	6	1
3-2-1-2-1 *)	3	5	1
3-1-2-1 *)	3	4	1
3-1-3-1 *)	3	4	2
3-1-3-2-1 *)	3	5	1
3	3	1	1
3-2-3-2 *)	3	4	1
3-2-1-3-1 *)	3	5	1
Zwischensumme Start 3	Varianten: 11	Mittelwert: 3,8	Summe: 22 (32 %)
Summen	Varianten: 31	Mittelwert: 3,7	68

*) Blickfolgen mit Richtungsänderungen (backtracks)

Die Buttonleiste in der dritten Maske weist gegenüber den beiden vorhergehenden Masken zwei Unterschiede auf. Zum einen besteht die Buttonleiste aus drei anstelle von zwei Buttons, zum anderen wurde die Position des weiterführenden Buttons von rechts nach links verschoben. Diese Maske kann nicht wie zuvor über den rechten Button verlassen werden, sondern nur über den linken Button. Weiterhin wurden die Beschriftungen geändert, sodass drei neue Begriffe vorkommen. Auf die Unterschiede in der Begrifflichkeit wird an dieser Stelle nicht detailliert eingegangen. Es ist aber anzunehmen, dass die an keine Erinnerung anknüpfenden Begriffe die Blickfolgen negativ beeinflussten.

Die Erweiterung und Positionsänderung der Buttons verursachte Unsicherheiten bei den Testpersonen, die sich im Blickverhalten niederschlagen. Anstelle von 5 bzw. 6 Varianten bei den vorherigen Masken sind diese hier um das ca. Fünffache auf 31 Varianten gestiegen. Der Mittelwert der Blicke innerhalb einer Blickfolge hat ebenfalls zugenommen, d. h. auf knapp 4 Blicke. Die Streuung beträgt 1 bis 8 Blicke pro Folge.

72 % der Testpersonen starteten die Blicke mit Button 2 (40 %) und Button 3 (32 %). Hierbei kommt das in den Masken 1 und 2 erlernte Verhalten zum Ausdruck. Der wei-

terführende Button wird rechts erwartet. Da es nun jedoch zwei anstatt drei Buttons gibt, richteten 40 % den Blick zuerst auf den mittleren Button (entspräche Button 2 in den vorherigen Masken). 32 % wiederum nahmen (peripher) wahr, dass es einen noch weiter rechts liegenden Button gibt und betrachteten diesen als erstes.
12 Personen (ca. 18 %) blickten sofort und ausschließlich nur den Zielbutton 1 links an. Bei der Prüfung dieser „Sonderfälle“ lässt sich feststellen, dass es sich dabei vorwiegend um Aufzeichnungen handelt, die einen Versatz von mindestens 20 mm außerhalb des Buttonumrisses haben. Diese Fälle wurden nicht ausgeschlossen.

3.3.5 Übergangshäufigkeiten

Übergangshäufigkeiten lassen sich prinzipiell folgendermaßen bestimmen:

- Einteilung des Stimulusmaterials in Blickgebiete (Objekte oder AOIs)
- Auszählen der Übergänge von einem Blickgebiet zum anderen für alle Blickgebiete. Hier muss darauf geachtet werden, dass der Wechsel von A nach B und der Wechsel von B nach A erfasst wird.
- Angabe der Übergangshäufigkeiten als absolute Häufigkeiten oder als relative Häufigkeiten (d. h. prozentualer Anteil der Wechsel im Verhältnis zu allen Wechseln).

Bei der Angabe der Übergangshäufigkeiten gibt es zwei Varianten.
Variante 1) Die Übergangshäufigkeiten können gerichtet angegeben werden, d. h. es werden die Übergangshäufigkeiten von A nach B getrennt von den Übergangshäufigkeiten von B nach A berichtet.
Variante 2) Die Übergangshäufigkeiten von A nach B und von B nach A werden zusammengefasst, d. h. sie werden ungerichtet angegeben.

Beispiel

Als Beispiel diente das Hauptmenü. Auffällig zeigte sich bei dem Experiment beim bloßen Beobachten der Bedienung, dass die Menüs sehr lange betrachtet wurden. Mit dem Parameter der Übergangshäufigkeit ließ sich schließlich dieses Verhalten detaillierter analysieren. Abbildung 22 zeigt einen Beispiel-Plot einer Aufzeichnung. Auf den ersten Menüpunkten liegen Fixationen.

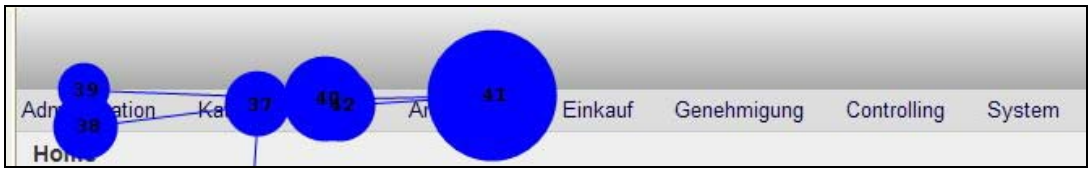

Abbildung 22: Beispielmaske Gaze-Plot Übergänge

Das richtige Menü ist das 3. Menü von links mit dem Namen „Anforderung“. Dieses Menü wurde mehrfach verwechselt. Die Aufgabe lautete, einen Artikel zu bestellen. Darum vermuteten die Benutzer, dass „Einkauf“ oder „Katalogverwaltung“ der richtige Einstieg sei. Neben der unklaren Beschriftung sorgte auch die Vielzahl von Menüs für Schwierigkeiten. Links, an der prominentesten Position, steht das Menü „Administration“. Dieses Menü können die Benutzer jedoch mangels Rechten nicht aufrufen. Das gleiche gilt für die Menüs „Genehmigung“ und „System“ etc.

Berechnung des Parameters

Die Übergangshäufigkeiten werden von ClearView nicht zur Verfügung gestellt. Die Übergänge könnten theoretisch über das AOI-Tool von ClearView und einen Excel-Export mit AOI-Template ermittelt werden. In der Praxis konnte das Tool jedoch aufgrund seiner Einschränkungen in der Anzahl der Aufzeichnungen (51) nicht genutzt werden.

Die Übergänge wurden daher manuell und rechnerisch wie folgt ermittelt:

- Die Menüleiste wurde in einzelne Bereiche eingeteilt. Jeder Menüpunkt stellt solch einen Bereich dar.
- Über die Gaze-Plots und die Fixationsdaten (FXD-Datei) wurden die Blickpunkte je Menüpunkt ermittelt. Es wurden nur Blicke und nicht einzelne Fixationen bewertet. Die Reihenfolge der betrachteten Menüpunkte wurde notiert, zum Beispiel 2-1-4-3. Es wurden alle Wechsel in alle Richtungen erfasst, also zum Beispiel auch 2-1-2.
- Anschließend wurden die Übergänge prozentual festgehalten und in einer Übergangsmatrix dargestellt. Es wurde also der prozentuale Anteil der Wechsel im Verhältnis zu allen Wechseln erfasst. Die Übergangshäufigkeiten werden ungerichtet angegeben.

Ergebnisse

Die einzelnen Ergebnisse sind in der folgenden Übergangsmatrix zusammengefasst (Tabelle 19). Die Matrix enthält im oberen Bereich die absoluten Werte. Die Summe der Übergänge beträgt 61. Im unteren Teil stehen die entsprechenden Prozentwerte.

Tabelle 19: Ergebnis Übergangsmatrix

	Menü 1	Menü 2	Menü 3	Menü 4	Menü 5	Menü 6	Menü 7	**Summe**
Menü 1		6	4				2	**12**
Menü 2	7		6	1	1	1		**16**
Menü 3	4	3		6				**13**
Menü 4		2	8					**10**
Menü 5				3		1		**4**
Menü 6		1	1		1			**3**
Menü 7		2			1			**3**
Summe	**11**	**14**	**19**	**10**	**3**	**2**	**2**	**61**
	Menü 1	Menü 2	Menü 3	Menü 4	Menü 5	Menü 6	Menü 7	**Summe**
Menü 1	0	9,84	6,56	0,00	0,00	0,00	3,28	**19,67**
Menü 2	11,48	0,00	9,84	1,64	1,64	1,64	0,00	**26,23**
Menü 3	6,56	4,92	0,00	9,84	0,00	0,00	0,00	**21,31**
Menü 4		3,28	13,12	0,00	0,00	0,00	0,00	**16,39**
Menü 5		0,00	0,00	4,92	0,00	1,64	0,00	**6,56**
Menü 6		1,64	1,64	0,00	1,64	0,00	0,00	**4,92**
Menü 7		3,28	0,00	0,00	1,64	0,00	0,00	**4,92**
Summe	**18,03**	**22,95**	**31,15**	**16,39**	**4,92**	**3,28**	**3,28**	**100,01**

Die häufigsten Übergänge finden bei den ersten (von links betrachtet) Menüpunkten statt. Je weniger Zellen in der Matrix gefüllt sind, desto besser ist das Menü gestaltet. Als problematische Menüpunkte lassen sich demnach die ersten Menüpunkte identifizieren.

Da an den Menüs Handlungen ausgeführt werden müssen, ist der Parameter der Designdomäne „Interaktion“ zugeordnet.

3.3.6 Übergangsdichte

Die Übergangsdichte ist der Anteil der mit mindestens einem Wert gefüllten Zellen in Prozent der möglichen Übergänge.

Beispiel

Zu dem vorangegangenen Menü-Beispiel mit den Übergangshäufigkeiten soll davon nun noch die Übergangsdichte abgeleitet werden.

Berechnung des Parameters

Die Übergangsdichte lässt sich aus den Übergangshäufigkeiten berechnen. Betrachtet werden die möglichen Übergänge bei sieben Menüs. Schließt man die Diagonale aus, sind dies insgesamt 42 mögliche Übergänge.

Aufzeichnungen ohne Übergänge, d. h. der richtige Menüpunkt wird sofort fixiert, fließen nicht in das Ergebnis ein. Das trifft bei wiederholten Aufzeichnungen aufgrund eines vorherigen Software-Abbruches zu. Den Testpersonen war das korrekte Menü bereits bekannt. Falls diese „Treffer“ berücksichtigt werden würden, wäre die Übergangsdichte wesentlich geringer, nämlich 9,0 %. Abbildung 23 zeigt einen Plot, bei dem nur das Menü „Bestellung“ betrachtet und angeklickt wurde.



Abbildung 23: Maske Gaze-Plot Menü ohne Übergänge

Ergebnis

Die Übergangsdichte beträgt 11,7 %.

3.4 Zusammenfassung der Ergebnisse

3.4.1 Designdomänen, Oberflächen-Komponenten und Parameter

Die Designdomänen, Oberflächen-Komponenten und Eye-Tracking-Parameter lassen sich wie folgt zusammenfassen (Tabelle 20). Zum einen werden die aus dem WebSCORE-Referenzmodell abgeleiteten Eye-Tracking-Designdomänen „Struktur“, „Interaktion“ und „Präsentation“ aufgeführt. Die Designdomänen sind in den Tabellenspalten dargestellt. Weiterhin wurden aus den Problemstellen der Software diejenigen Oberflächen-Komponenten ermittelt, an denen sich die Hürden konzentrierten. Dies waren die interaktionsreichen Komponenten „Menüs“, „Baum“ und „Baumeintrag“ sowie „Buttons“. Hinzugekommen ist die übergreifende Komponente „Seitenlayout“, um die gestalterische Qualität untersuchen zu können. Die Oberflächen-Komponenten sind in den Tabellenzeilen gelistet.

Die sechs Eye-Tracking-Parameter, mit deren Hilfe die jeweiligen Oberflächen-Elemente und damit indirekt die Designdomänen untersucht werden können, stehen in den entsprechenden Zellen der Ergebnistabelle.

Tabelle 20: Ergebnistabelle Designdomäne - Komponente - Parameter

Domäne **Komponente**	**Struktur**	**Interaktion**	**Präsentation**
Menüs		• Übergangshäufigkeit • Übergangsdichte	
Baum	Suchzeit		
Baumeintrag			Relative Häufigkeit der Fixationen von Objekten
Buttons		Richtungsänderungen	
Seitenlayout			Kumulierte Übergangszeit

Zusätzlicher Parameter Blickpfaddauer

Weiterhin wird der Parameter der Blickpfaddauer zur Beurteilung herangezogen. Er gibt jedoch nicht direkt Auskunft über eine der Designdomänen, sondern über die zeitliche Effizienz, mit der eine Aufgabe mit der Software bearbeitet werden kann.

3.4.2 Aussagekraft der Parameter

Insgesamt wurden sieben Eye-Tracking-Parameter ermittelt, deren Aussagekraft hier zusammengefasst ist (Tabelle 21).

Tabelle 21: Ergebnistabelle Aussagekraft je Parameter

Zeitbezogene Parameter	
Parameter	**Aussagekraft**
Blickpfaddauer (Leistungsdaten Bearbeitungszeit)	Je kürzer, umso effizienter ist die Bearbeitung einer Aufgabe.
Suchzeit (Baum – Struktur)	Je kürzer, umso einfacher und übersichtlicher ist eine Struktur.
Kumulierte Übergangszeit (Seitenlayout – Präsentation)	Je kürzer, umso besser ist die Softwareseite gestaltet.
Anzahlbezogene Parameter	
Parameter	**Aussagekraft**
Relative Häufigkeit der Fixationen von Objekten (Baumeintrag – Präsentation)	Je weniger Fixationen pro AOI, desto besser sind Listen für Suchprozesse gestaltet. (Achtung: Aussage ist nicht für Leseverhalten gültig)
Sequenzberücksichtigende Parameter	
Parameter	**Aussagekraft**
Übergangshäufigkeiten (Menü – Interaktion)	Je weniger Zellen in Matrix gefüllt sind, desto besser ist das Menü gestaltet.
Übergangsdichte (Menü – Interaktion)	Wenn die Übergangsdichte größer 11,3 % ist, ist die Gestaltung nicht effektiv.
Richtungsänderungen (Buttons – Interaktion)	Je weniger Richtungsänderungen, desto erwartungskonformer ist die Interaktion.

3.5 Ergebnisse des retrospektiven Protokolls mit Nutzerzitaten

Im Folgenden wird eine Auswahl an sinngemäßen Nutzerzitaten aus dem nachträglichen Interview bei gleichzeitigem Vorspielen des Blickverlaufs wiedergegeben. Die Stellungnahmen sind unterschieden nach „Kritik“ und „Verbesserungsvorschläge“.

Kritik an der Software-Usability

a) Startseite, Hauptmenü, Cursor-Symbol

Das falsche Cursor-Symbol ist irreführend!

b) Anforderung Kugelschreiber

Auf dieser Seite sind zwar viele Anzeigefelder. Aber ich möchte nichts falsch machen, darum lese ich gründlich.

c) Kommentar einer ca. 60-jährigen Testperson

Ich habe schon mit dem Öffnen eines Menüpunktes des Hauptmenüs große Probleme. Muss ich die Maus gedrückt halten? Außerdem ist im Katalog das Pfeilsymbol so winzig, dass ich es einfach nicht erwische, anzuklicken. Die Ergebnisliste habe ich gar nicht gesehen, weil ich nur links geschaut habe. Insgesamt war mir der Workflow nicht klar.

Beobachtung der Studienleiterin: Die Testperson hatte große haptische Probleme, die für einen ungeübten PC-Anwender typisch sind. Da die Software zudem besonders anfällig auf Bedienfehler reagierte, vor allem das Klicken von sensitiven Bereichen, verschärften sich die motorischen Probleme noch.

Verbesserungsvorschläge von den Testpersonen

a) Beschriftungen, Verständlichkeit

Im Hauptmenü auf der Startseite wäre das Menü „Bestellung“ viel klarer als „Anforderung“. Man bestellt doch etwas.

b) Erwartungshaltung

Der Baum mit den Kataloggruppen sollte wie im Windows Datei-Explorer funktionieren. Dort gibt es das Plus-/Minus-Zeichen, außerdem kann die Bezeichnung selbst auch angeklickt werden und es öffnet sich die Unterebene. Keinesfalls sollte es zwei Symbole (Pfeil und Ordner) geben, das ist verwirrend.

Mir fehlte der Überblick in dem Workflow. Alle Schritte der Bestellung sollten mit einem Fortschrittsbalken oder einer Übersichtsanzeige dargestellt werden, wobei jeweils der aktuelle Teilprozess hervorgehoben ist. So könnte man auch sehen, wie viele Schritte noch vor einem liegen. Ich konnte nicht feststellen, wann der Artikel nun endgültig angefordert ist. So ein Wizard wie bei ebay oder amazon wäre gut.

Stand der Technik ist doch die kombinierte Suche, also dass Volltext-Suche mit hierarchischer Suche kombiniert ist. Bei Volltext-Suche sollte bei Eingabe eines neuen Buchstabens dynamisch die verbleibenden Resultate angezeigt werden.

c) Komplexitätsreduktion

Die Seite „Kontierung“ mit seiner Verteilfunktion ist überflüssig, wenn keine Aufteilung auf verschiedene Kostenstellen vorgenommen werden soll/kann. In diesem Fall könnten in der vorherigen Seite „Artikel Kugelschreiber“ zwei Buttons bereitstehen, nämlich „Weiter mit Verteilung“ und „Weiter ohne Verteilung“. Letzterer würde die Seite „Kontierung“ überspringen und das System würde die Verteilung automatisch im Hintergrund vornehmen.

Warum sehe ich im Hauptmenü auf der Startseite überhaupt Menüpunkte, die für mich als normaler Besteller gar nicht relevant sind und die ich mit meiner Anmeldung gar nicht aufrufen kann? Blenden Sie diese doch einfach aus.

4 Diskussion und Ausblick

Das Kapitel 4 besteht aus der Diskussion der Ergebnisse. Sie gliedert sich nach den zwei Fragen aus der Fragestellung.

4.1 Diskussion

4.1.1 Auswahl der relevanten Parameter und deren Wertebereiche

Die erste Fragestellung in dieser Arbeit lautete:

Fragestellung 1: Frage nach der Auswahl relevanter Parameter
Mit welchen Eye-Tracking-Parametern lässt sich die Gebrauchstauglichkeit einer Software prüfen? Aus welchen Parametern lassen sich Kriterien zur Effizienz einer Software ermitteln? Welche Parameter geben Aufschluss über den Bedienaufwand, das heißt über das Aktivitäts-Niveau und die kognitiven Vorgänge bei der Softwarenutzung?

In dieser Arbeit konnten sieben Parameter ermittelt werden, die für eine Usability-Prüfung von Software aussagekräftig sind. Welchen Wertebereich die Parameter haben, hängt sehr von der Aufgabenstellung ab. Ein Beispiel soll das veranschaulichen: Beim Lesen von Text lässt sich die Anzahl der Fixationen pro Zeile identifizieren. Viele Fixationen weisen auf intensives Lesen, wenige Fixationen auf überfliegendes Lesen hin. Bei Suchprozessen dagegen ist eine geringe Anzahl von Fixationen wünschenswert, da sie auf ein schnell auffindbares Suchziel oder auf eine einfache Suchaufgabe hindeutet.
Aufgrund der Abhängigkeit vom Nutzungskontext wurden die Parameter und ihre Werte in dieser Arbeit immer in den Zusammenhang zu einer bestimmten Oberflächen-Komponente und einer Designdomäne gestellt.

Zusammengefasst geht aus der Wahl der gefundenen, aussagekräftigen Parameter hervor, dass die Summe der Fixationen ausschlaggebend ist: *„Bei der Auswertung von Fixationen nicht die einzelnen Parameterwerte die beschriebenen kognitiven Prozesse widerspiegeln, sondern eher deren Summe. Das heißt, nicht jede einzelne Fixation, sondern die Summe der Fixationen sollte betrachtet werden, damit es unerheblich wird, ob während einer Fixation das augenblicklich fixierte Objekt Gegenstand der kognitiven Verarbeitung ist oder ob dieses während einer vorhergegangenen Fixation bereits analysiert wurde.“* (Schimpfky 2003, S. 44)

In den Ergebnissen wurden nur die auffälligen Parameter den Designdomänen zugeordnet. Es kann davon nicht abgeleitet werden, dass nicht weitere Parameter Aussagen zur Usability ermöglichen können.

Die zweite Fragestellung war:

Fragestellung 2: Frage nach dem Wertebereich dieser relevanten Parameter

Wie müssen die Werte der relevanten Parameter ausfallen, um einen hohen oder niedrigen Grad von Usability zu indizieren?

Mit Ausnahme der Übergangsdichte, die konkrete Zahlenwerte vorgibt, haben die Parameter keine absoluten Zahlenwerte als Maßstab für den günstigen oder ungünstigen Wertebereich. Mehrere Studienergebnisse zeigen auf, dass die Usability-Qualität von untersuchten Software-Oberflächen nur relativ zueinander betrachtet werden kann. (Cowen 2001, Goldberg u. Kotval 1998). *„There is no absolute benchmark of what eye movement patterns a good or a bad interface produces."* (Cowen 2001, S. 18)

Die Wertebereiche werden vielmehr über die Relativität zueinander angegeben: je kürzer oder je seltener ein Wert ist, umso besser oder schlechter sind eine bestimmte Designdomäne oder bestimmte Leistungsdaten zu bewerten. Die kausale Bewertung setzt jedoch voraus, dass es Vergleichswerte gibt. Vergleiche werden in der Regel durch optimierte Varianten realisiert oder durch vergleichbare Stellen innerhalb der gleichen Software-Applikation. Die letztgenannte Möglichkeit war auch hier der methodische Weg zum Finden der aussagekräftigen Parameter und ihrer Wertebereiche. In der Testsoftware E-Procurement wurde nach Vergleichssituationen gesucht. Die Situationen wurden dann hinsichtlich ihrer auffälligen Parameter exploriert. In einem konkreten Usability-Test, der nicht wie im vorliegenden Fall nur methodischen Zwecken dient, könnten mittels der gefundenen Parameter Schwachstellen ermittelt werden. Die Schwachstellen sollten gemäß dem Usability-Engineering-Prozess durch ein Redesign behoben und anschließend erneut getestet werden.

Die folgenden Abschnitte diskutieren die Parameter und ihre Ergebnisse.

4.1.2 Blickpfaddauer

Die Blickpfaddauer entspricht der Bearbeitungszeit bis zur Lösung einer Aufgabe und zählt damit zu den Leistungsdaten. Um die Bearbeitungszeit zu ermitteln, ist der Einsatz der Methode Eye-Tracking nicht zwingend notwendig. PC Systemdaten, wie etwa sämtliche Texteingaben, Mausklicks, Mausbewegungen und Seitenaufrufe lassen sich mit handelsüblicher Aufzeichnungs-Software aufnehmen und darüber die Bearbeitungszeit ermitteln. Da Eye-Tracking jedoch Daten über die Blickpfaddauer bereitstellt, wurde diese als Bearbeitungszeit ausgewertet. Als Ergebnis dieses zeitlichen Parameters konnte festgehalten werden, dass die Blickpfaddauer ein Indikator für die Effizienz der Software ist. Je kürzer die Blickpfaddauer ist, umso effizienter ist die Bearbeitung einer Aufgabe. Sie lässt sich auf die Dauer der gesamten Aufgabe anwenden oder auf die Dauer von Teilaufgaben. Damit ist sie ein allgemeiner Parameter und dient nicht nur dazu, spezielle Software-Designprobleme aufzudecken.

Die Blickpfaddauer wird in diesem Sinne auch genutzt, um die visuelle Suche auf einem Interface zu beschreiben (Goldberg u. Kotval 1999). Sie zeigt hauptsächlich die Komplexität der Bearbeitung auf. Die Blickpfaddauer und die Anzahl der Rücksprünge zu bereits betrachteten Stellen haben ein lineares Verhältnis zu Usability-Bewertungen und sind die besten Vorhersagen für Usability. Als Beispiel nennen die Autoren, dass eine ungewöhnliche Oberfläche eine um 80 % längere Blickpfaddauer benötigt als eine sehr gute Oberfläche. Die gesamte Betrachtungszeit dient zudem als Beleg, um die Ergebnisse der anderen Eye-Tracking-Parameter zu überprüfen. In dem vorliegenden Experiment deckt sich beispielsweise die Übergangsdichte, die eine ungenügende Gestaltung der Software belegte, mit der langen Bearbeitungsdauer.

Die Gesamtdauer des Blickpfades kann insbesondere als Maß für den Bearbeitungsaufwand einer Schnittstellengestaltung verwendet werden, wenn verschiedene Varianten miteinander verglichen werden oder indem die gemessene Dauer mit einem normativen Optimum verglichen wird (Goldberg u. Kotval 1998). Der Vergleich mit einem Optimum und die Aussagekraft der Blickpfaddauer wurden auch in diesem Experiment durchgeführt und ließen sich durch die Ergebnisse bestätigen.

Als unzureichend beurteilt wird die Blickpfaddauer jedoch, wenn sie als einziger Parameter zur Evaluation zweier Gestaltungsvarianten herangezogen wird, die sich nur in

der Gruppierung der Oberflächenelemente unterscheiden (Cowen 2001). Sobald der Benutzer das gesuchte Ziel-Element gefunden hat, ist die Bearbeitung beendet und es zeigen sich keine Unterschiede in den Bearbeitungszeiten. Die Gesamtbetrachtungszeit ist demnach nicht sensitiv auf die Gruppierungsunterschiede.

Wissenschaftler vom MIT Media Lab Cambridge (Lieberman u. Wagner 2003) stellen dar, eine der größten Herausforderungen bestehe für E-Commerce-Software darin, dass elektronische Transaktionen zeitlich effizient sind und korrekt ablaufen. Verbraucher werden durch zu zahlreiche Menüs frustriert, das Lösen von auftretenden Problemen ist umständlich und verzögert häufig den Bestellabschluss. Das ist eine Motivation dafür, in Evaluationen die methodisch relative einfache Messung der Bearbeitungszeit durchzuführen und als Maßstab zu verwenden.

4.1.3 Suchzeit

Das Ergebnis der Studie war, dass die Suchzeit die Evaluierung der Designdomäne „Struktur“ ermöglicht. Je kürzer die Suchzeit ist, umso einfacher und übersichtlicher ist eine Struktur.

Suchzeiten sind Zeiten, in denen die Anwender ein Zielobjekt suchen. Die Zeit ist hier indirekt auch ein Maß für die Effizienz, aber sie geht darüber hinaus. Indem sich die Suche auf konkrete Oberflächen-Komponenten bezieht, können deren Strukturen bewertet werden. Kurze Suchzeiten deuten auf gut strukturierte, lange Suchzeiten auf mangelhaft strukturierte Elemente hin. Bei kurzen Suchzeiten wird das Ordnungssystem der Elemente rasch erkannt. Die Ergebnisse zeigten, dass Benutzer nur wenige Elemente auf einen Blick überblicken und merken können, ohne wiederholt die gleichen Elemente zu lesen.

Die Suchzeit lässt sich in Abhängigkeit von der Anzahl der Einträge darstellen. Der genaue Verlauf der Suchzeiten zwischen drei und 14 Einträgen ist nicht bekannt. Im Folgenden soll basierend auf den Erkenntnissen der Wahrnehmungs- und Gedächtnispsychologie der vermutete Verlauf aufgezeigt werden (Abbildung 24). Auf der Abszisse sind die Anzahl Einträge aufgetragen und dabei die durch das Ergebnis bekannten Einträge drei und 14. Auf der Ordinate sind die Mittelwerte der Suchzeiten zu den beiden verschiedenen Anzahl Einträgen angegeben. Die Kapazität des Arbeitszeitgedächtnisses

ist beschränkt, d. h. es können grundsätzlich nicht mehr als 5-9 Informationseinheiten – so genannte „chunks“ – behalten werden (Miller 1956). Ab einer größeren Anzahl von Elementen beginnt der Betrachter die Elemente zu gruppieren. Es ist also anzunehmen, dass sich der zeitliche Verlauf zwischen 3 und 14 Einträgen nicht linear verhält. Vielmehr wird vermutet, dass die benötigte Suchzeit bis zu sieben Einträgen nur unmerklich anwächst und zwischen sieben und 14 Einträgen exponentiell steigt.

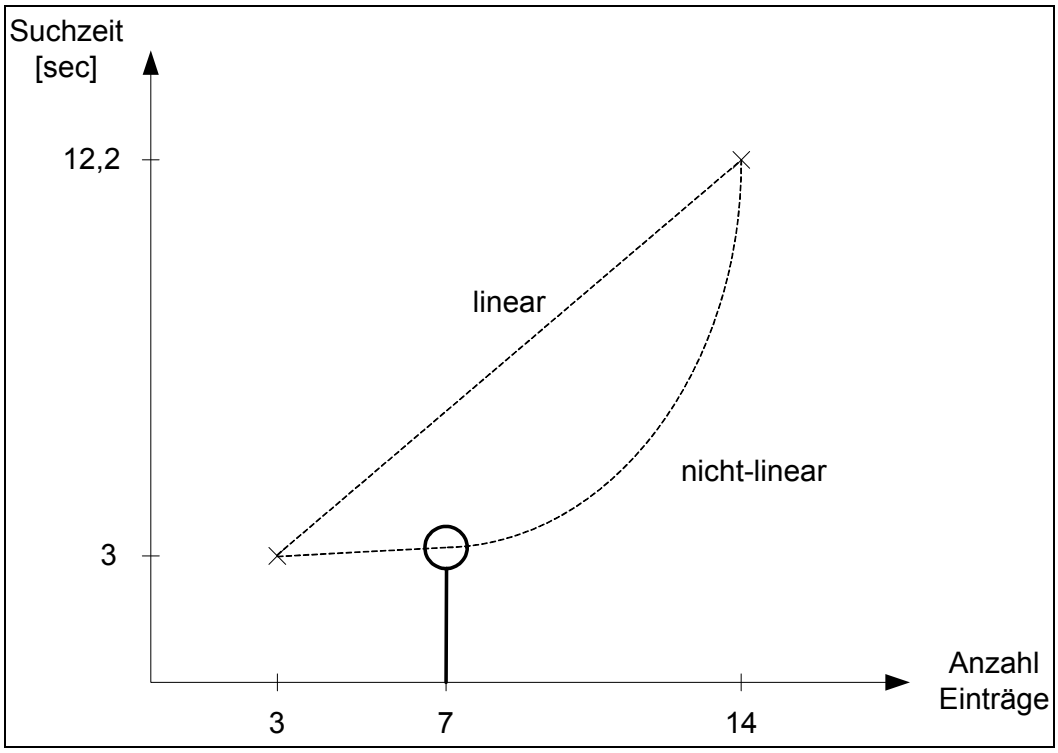

Abbildung 24: Diagramm Suchzeiten abhängig von der Anzahl Einträge

Mit dem Parameter der Suchzeit wurde der Zeitbedarf beim Erfassen von Verzeichnis-Strukturen ermittelt und davon die Komplexität abgeleitet. Bei der komplexeren Struktur, also 14 Einträge, wäre es zusätzlich von Interesse, welche Strategie die Anwender anwendeten. Die Blickbewegungen in Pull-Down-Menüs, Baumstrukturen und Miniaturbild-Strukturen sind hinreichend erforscht worden (Aaltonen et al. 1998, Byrne et al. 1999, Masuda et al. 2002). Es findet eine Vermischung der zwei prinzipiellen Suchstrategien statt. Bei der Reinform der algorithmischen Suche beginnen die Benutzer bei den obersten Items und der Blick wandert diszipliniert von einem Item zum nächsten nach unten. Diese Strategie führt mit Sicherheit zum Ziel, ist aber prinzipiell nicht die schnellste Methode. Die andere Suchstrategie in Reinform ist heuristisch, dabei werden Faustregeln angewendet, z. B. erster Eintrag, Sprung zum letzten Eintrag, dann zurück in die Mitte und jeweils iterativ die Items in den Zwischenräume betrachten. Die Blick-

folgen in Software-Applikationen weisen eine Mischform auf. Der erste Blick in Pull-Down-Menüs richtet sich demnach auf einen der obersten drei Einträge, anschließend werden die Einträge von oben nach unten gelesen. Während der Blick nach unten gleitet, werden einzelne Einträge übersprungen (Scheier u. Heinsen 2003).

Im Zusammenhang mit dem Erlernen und dem Anwenden von Software können folgende Schlussfolgerungen für die Gestaltung der Systeme gezogen werden.

- Damit Informationen bestmöglich rezipiert und erinnert werden können, muss die präsentierte Informationsmenge in der Software auf wenige, bedeutungsvolle und unterscheidbare Informationseinheiten reduziert werden. Überladene Software-Masken sind nicht lernförderlich.
- Gruppierung von zusammenhängenden Informationen kann besser im Gedächtnis behalten werden als isoliert dargestellte. Die Gruppierung erleichtert dem Benutzer, verschiedene Teile der Information besser zu verknüpfen.
- Damit der semantische Gehalt der Information im Arbeitsgedächtnis erschlossen werden kann, müssen im Langzeitgedächtnis entsprechende Verknüpfungen aktiviert werden. Deshalb ist es wichtig, die Wissensmodelle der Benutzer zu berücksichtigen. Die Gestaltung von Objekten und Informationen muss kompatibel zum Wissen und zur Erfahrung der Betrachter sein. Bekannte Begriffe und Standardisierungen, wie z. B. einheitliche Ordnersymbole oder optische Einrückungen zur Verdeutlichung von Baumhierarchien, ermöglichen die Bildung starker Assoziationen.

4.1.4 Kumulierte Übergangszeit

Das Ergebnis der Studie war, dass die kumulierte Übergangszeit die Evaluierung der Designdomäne „Seitenlayout – Präsentation“ ermöglicht. Je kürzer die Übergangszeit ist, umso besser ist die Software-Oberfläche gestaltet.

Die kumulierte Übergangszeit ist eine Zeitspanne, die nicht nur eine einzelne Sakkade, sondern die Summe aller Sakkaden in einem bestimmten Versuchsabschnitt umfasst. Übergangszeiten sind „tote Zeiten“, in denen keine Informationsaufnahme stattfindet. Diese sollten darum möglichst kurz sein. Da die Sakkadendauer im Wesentlichen von der Sakkadenweite abhängt, ist die Gestaltung von Informationen maßgeblich. Kurze

Wege und funktionell zusammengehörige Bereiche in örtlicher Nähe reduzieren die Übergangszeiten.

Ein typischer Anwendungsfall für zu weite sakkadische Blickbewegungen sind ungünstig positionierte Fehlermeldungen (Abbildung 25). Nach den Regeln der Ergonomie sollen Meldungen einen Bezug zwischen Meldung und Fehlerquelle herstellen. Man erkennt im abgebildeten Beispiel, dass der Benutzer eine weite Übergangs-Sakkade zwischen den Blickobjekten „Meldung“ und „Feld“ durchführen muss.

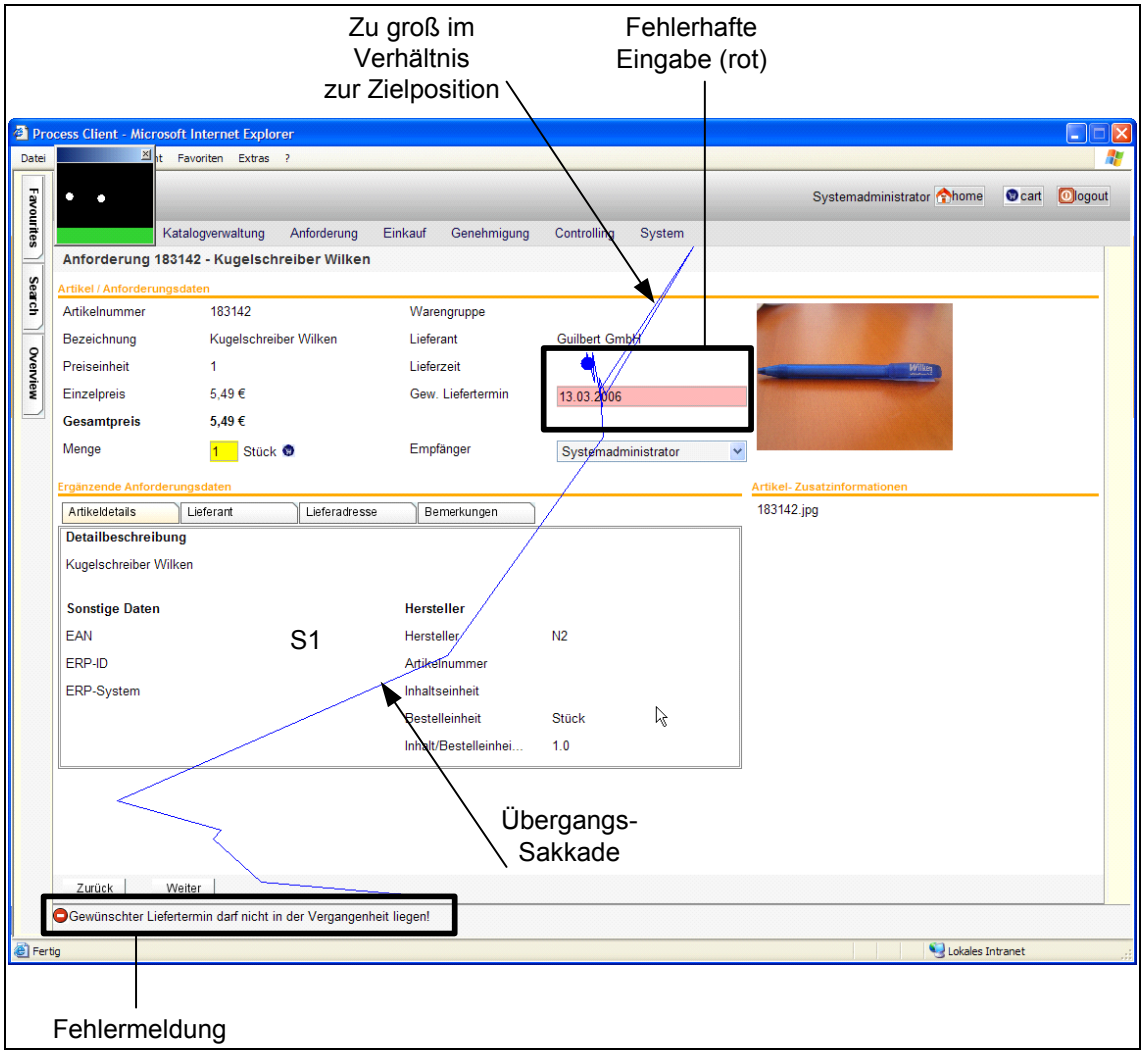

Abbildung 25: Beispiel Übergangszeiten Fehlermeldung

Das Beispiel macht eine weitere Auffälligkeit von sakkadischen Bewegungen deutlich. Die Sakkade ist in der Abbildung nicht zu einer Linie begradigt, wie in den Plot-Darstellungen, sondern ist in seiner realen Form dargestellt. Man kann daraus erkennen, dass die Sakkade über das Feld (Zielposition) hinausgeht. Die Ungenauigkeit der

Sprungweite von Sakkaden und die daraus resultierenden Korrektur-Sakkaden lassen sich physiologisch erklären: *„Das Steuersignal für den Agonisten besteht aus zwei Komponenten: einem Signalimpuls und einem Anschlusssignal. (…) Signalimpuls und Anschlusssignal können im Verhältnis zur Zielposition zu groß oder zu klein sein." (Rötting 2001, S. 75-76)*. Da die Sprungweite unter anderem die Dauer einer Sakkade beeinflusst, trägt eine mögliche Ungenauigkeit der Sprungweite auch zu einer verlängerten Übergangszeit bei.

Visuelle Aufmerksamkeit und Gestaltgesetze

Zur Vermeidung von langen kumulierten Übergangszeiten ist es empfehlenswert, die Erkenntnisse aus der Gestaltpsychologie zu berücksichtigen und damit die Präsentation der Software-Oberfläche zu optimieren. Die Gesetze der Gestaltpsychologie sind Mechanismen, die in die vorbewussten Prozesse des Verstehens greifen: Farben, Kontraste, Detail-Reichtum, Überraschungen (Emotionen) und Vorwissen. Die Gestaltgesetze (Wertheimer 1925) beschreiben räumlich-bildliche Strukturen, nach denen sich die Wahrnehmung und das Verstehen organisieren. Beobachtungen haben ergeben, dass unsere Wahrnehmung grafische Formen wie Kreise und Rechtecke in Gruppen sortiert und als Gestalten interpretiert. Insbesondere das Gesetz der räumlichen Nähe ist bei der Gestaltung relevant. Benachbarte Elemente werden in der Wahrnehmung zu einer Gruppe zusammengefasst.

4.1.5 Relative Häufigkeit der Fixationen von Objekten

Das Ergebnis der Studie war, dass die relative Häufigkeit der Fixationen von Objekten die Evaluierung der Designdomäne „Präsentation" ermöglicht. Je weniger Fixationen pro Interessensgebiet stattfinden, desto besser ist die Software-Oberfläche gestaltet.

Die relative Häufigkeit der Fixationen von Objekten gibt Aufschluss darüber, welche Bereiche im Vergleich zu anderen Bereichen häufiger oder seltener betrachtet wurden. Damit lässt sich die Aufmerksamkeitsverteilung auf einzelne Objekte ermitteln. Es stellt sich nun die Frage, welche Interpretation davon abgeleitet werden kann.

Die relative Anzahl der Fixationen ist ein in der Blickbewegungsanalyse häufig genutzter Parameter. Er wird im Allgemeinen so interpretiert, dass er die Wichtigkeit eines Elementes reflektiert. Demnach haben häufiger betrachtete Elemente für den Benutzer

eine höhere Wertigkeit oder Informationshaltigkeit als seltener betrachtete Elemente (Fitts et al. 1950).

Grundsätzlich sind die Aufgabenstellung und der Arbeitsprozess bei der Interpretation dieses Häufigkeitsparameters zu berücksichtigen. Es kann zwischen Lese- und Suchprozessen unterschieden werden. Die Verschiedenartigkeit dieser Prozesse führt auch zu unterschiedlichen Interpretationen dieses Parameters.
Bei Leseprozessen wird die Richtung und Form der Augenbewegungen vorgegeben. Viele Fixationen pro Textzeile, am besten pro Wort oder Wortteil, deuten demzufolge auf genaues Lesen hin. Wenige Fixationen lassen eher auf ein flüchtiges Lesen schließen. Eine höhere Anzahl von Fixationen pro Zeile kann hinsichtlich des Leseprozesses demnach positiv gewertet werden. Im Bereich der Diagnostik von Leseschwächen ergab eine Studie, dass sich die Untersuchungsgruppen (Kinder mit/ohne Legasthenie) in der Anzahl und Dauer der Fixationen pro Wort während des Lesevorganges signifikant unterschieden. Die Stichprobe der legasthenen Kinder machte während des Lesevorgangs deutlich mehr Fixationen pro Wort als die Kontrollgruppe (Dürrwächter 2003).
Bei Suchprozessen verhält es sich anders als beim Lesen, da die Augenbewegungen nicht vorgegeben sind und die dargebotenen Informationen frei erschlossen werden können. Während der Suchzeit, die vor allem unter Zeitdruck besonders kurz gehalten wird, werden nur die vom Benutzer als wichtig eingestuften Bereiche betrachtet. Für die Gestaltung von Software-Oberflächen bedeutet das, dass die wichtigsten oder am häufigsten genutzten Elemente an auffälliger Position angeordnet werden sollten. Bei Suchaufgaben sind demnach Gestaltungsvarianten zu bevorzugen, bei denen die gesuchte Information mit weniger Fixationen aufgenommen werden kann. Ein Anwendungsfall, bei dem mehrere Systemvarianten mit Hilfe der relativen Häufigkeit der Fixation der verschiedenen Objekte evaluiert werden, ist die Fahrernavigation (Fairclough et al. 1993). Um nicht von der primären Fahraufgabe abzulenken, sollten die wichtigen Informationen an zentralen Stellen auf dem Cockpit mit wenigen Fixationen erfasst werden können.

Die relative Anzahl der Fixationen wurde in der vorliegenden Studie am Beispiel des Verzeichnis-Katalogbaumes untersucht und stellt einen Suchprozess dar. Ein Ergebnis war die deutlich nachlassende Aufmerksamkeit je Eintrag vom Satzanfang über die Satzmitte bis zum Satzende hin und die damit verbundene nachlassende Fixationshäu-

figkeit dieser Satzabschnitte. Software-Designer werden aus dieser Erkenntnis heraus vor die Aufgabe gestellt, die sinntragenden Wörter gleich zu Beginn des Satzes anzuordnen. Keinesfalls sollte beispielsweise jeder Listen-Eintrag mit denselben Wörtern beginnen, denn das wäre für die Suche nachteilig. Die nach unten nachlassende Blickhäufigkeit ist allgemein, unabhängig von der Position des Zieles, in vielen Studien vermerkt worden. Die Konzentration lässt nach ca. sieben Elementen nach (Miller 1956).
Dieser Parameter eignet sich besonders dazu, direktes Optimierungspotenzial für die Präsentation von Oberflächenelementen abzuleiten.

4.1.6 Übergangshäufigkeiten und Übergangsdichte

Das Ergebnis der Studie war, dass sowohl die Übergangshäufigkeiten als auch die Übergangsdichte die Evaluierung der Designdomäne „Interaktion“ ermöglichen. Je weniger Zellen in der Übergangsmatrix gefüllt sind, desto besser ist eine Interaktions-Komponente gestaltet. Eine geringe Übergangsdichte deutet auf eine gute Interaktion hin.

Die Übergangshäufigkeiten können vom Blickpfad abgeleitet werden und können eine Aussage über die Effizienz der Anordnung von Elementen in Benutzeroberflächen geben (Jacob u. Karn 2003). Jacob und Karn fanden heraus, dass die Übergangshäufigkeit speziell in der visuellen Suche bei der Interaktion mit Menüs eine Messgröße darstellt. Die Blicksprünge von einem Menüpunkt zum anderen und wieder zurück waren besonders auffällig. Die Schlussfolgerung von Jacob & Karn soll hier jedoch leicht abgewandelt werden. Die Übergangshäufigkeit ist nicht nur ein Indiz für die Anordnungsgüte, das der Designdomäne „Präsentation“ entspräche. Sie ist auch ein Indiz, wie verständlich die Menübeschriftungen sind, wie viele Menüs vorkommen und ob beispielsweise unnötig viele Menüs sichtbar sind. Die Übergangshäufigkeit lässt auf komplexere Ursachen schließen als auf reine Anordnungsprobleme. Da an den Menüs Handlungen ausgeführt werden müssen, ist der Parameter der Designdomäne „Interaktion“ zugeordnet. In dieser Arbeit wurde lediglich die Übergangsmatrix 1. Ordnung ermittelt. Interessant wäre auch die Markov-Matrix 2. Ordnung. In dieser Matrix würde auch die letzte und vorletzte Fixation berücksichtigt.

Wenn man zwei Software-Oberflächen vergleicht, bedeutet eine Übergangsdichte kleiner <9,5 % eine gute Gestaltung der Oberfläche. Eine Dichte größer 10,2 % weist auf

eine schlechtere Gestaltung hin (Goldberg u. Kotval 1999). Der im Experiment ermittelte Wert lag bei 11,7 %. Damit wird zum einen bestätigt, dass die Prozentangabe wertmäßig nahe dem von Goldberg und Kotval gefundenen Bereich liegt. Da 11,7 % größer als 10,2 % ist, liegt ein langer, verteilter Blickpfad vor und lässt damit auf eine schlechte Gestaltung schließen. Das Ergebnis der Übergangsdichte bestätigt die beobachteten Probleme der Nutzer.

4.1.7 Richtungsänderungen

Das Ergebnis der Studie war, dass die Richtungsänderungen die Evaluierung der Designdomäne „Interaktion“ ermöglichen. Je weniger Richtungsänderungen stattfinden, desto erwartungskonformer ist die Interaktion.

Die Diskrepanz zwischen den Erwartungen der Person und der realen Situation, deren Folge sich in Richtungsänderungen im Blickverlauf niederschlägt, bestätigen auch Studienergebnisse zu interaktiver Software (Goldberg u. Kotval 1998, 1999). Wie das Beispiel in dieser Arbeit zeigte, verursachte die plötzliche Änderung der Button-Positionen eine Diskrepanz in der Erwartung und der vorgefundenen Situation beim Anwender. Je weniger Richtungsänderungen auftreten, umso besser können die Benutzer die eingeschlagene Blickrichtung fortsetzen und müssen sich nicht rückwärts bewegen.

Eine Richtungsänderung liegt vor, wenn *„eine Sakkade in der Richtung um mehr als ±90° von der Richtung der vorhergehenden Sakkade abweicht“* (Rötting 2001, S. 133-134). Diese Definition einer „Richtungsänderung“ wurde für die Berechnung des Parameters im Rahmen dieser Arbeit präzisiert (siehe Kapitel 3.3.4 Richtungsänderungen des Blickpfades, Seite 84). Das war einerseits notwendig, um Missverständnissen in der Berechnung von Richtungsänderungen vorzubeugen. Andererseits konnte über die Variation der Definition ermittelt werden, dass nicht nur die Richtungsänderungen zwischen einzelnen Sakkaden, sondern diejenigen, die zwischen Blickobjekten stattfinden, eine hohe Aussagekraft haben.

Allgemein zeigte das Beispiel im Experiment, dass Testpersonen verschiedene Strategien anwenden, um den richtigen Button auszuwählen. Die anerzogene Leserichtung von links nach rechts ist die erste Strategie, die bei erstmaliger Bedienung der Buttonleiste angewendet wurde. Bei der zweiten Buttonleiste wirkte sich die mit der ersten

Buttonleiste gemachte Erfahrung aus. Eine Erwartungshaltung war entstanden: der weiterführende Button ist rechts, der Back-Button ist links. Diese in der zweiten Buttonleiste bestätigte Erwartungshaltung wurde schließlich auf die dritte Maske angewendet. Hier war von den Software-Designern eine Umgestaltung vorgenommen worden, die für die Benutzer unerwartet war. Der Blickverlauf zeigt, dass die ersten Fixationen die mittlere oder rechte Position aufsuchten, dort jedoch nicht die erwarteten Buttons fanden. Die daraus resultierende Irritation äußert sich in langen Blickfolgen und einer Vielzahl von Regressionen. Bei Einstieg im mittleren Button, der offensichtlich nicht der gesuchte war, wenden die Testpersonen verschiedene Strategien an. Eine Gruppe begann diszipliniert wie zu Beginn von ganz links zu lesen. Die andere Gruppe hielt an der Vorerfahrung fest und ließ den Blick nach rechts schweifen, in der Hoffnung, dass es dennoch der rechte Button ist, dieses Mal nur um eine Position weiter nach rechts verschoben, da es insgesamt drei Buttons sind.

Das bei den Testpersonen beobachtete, rasche Einprägen der Button-Anordnung wird verstärkt durch die gängigen Styleguides der Softwarehersteller (Microsoft 1995, Sun-Microsystems 1999). Diese platzieren die Buttons in der Regel so, dass sich der Button „Abbrechen“ links vom Button „OK“ befindet. Die Testpersonen gaben in dem gebundenen Interview mehrfach an, dass ihnen eine bestimmte Button-Anordnung von Windows-Anwendungen geläufig sei.
Nutzungsprobleme, die Richtungsänderungen der Blicke verursachen, lassen sich auf mangelhafte Gestaltung zurückführen. Dies bestätigen mehrere Studien. So ermittelte beispielsweise Mackworth, dass die Anzahl der Richtungsänderungen fast linear mit der vertikalen Höhe des Stimulusmaterials steigt (Mackworth 1976). Demnach korreliert der prozentuale Anteil der Richtungsänderungen nicht nur mit der vertikalen Höhe des betrachteten Materials, sondern unterscheidet sich gleichzeitig zwischen dichten und weniger dicht angeordneten Elementen in dem Material. Je dichter die Elemente stehen, umso mehr Richtungsänderungen finden statt im Vergleich zu einer weniger dichten Anordnung der Elemente.

Stand der Forschung in der Leseforschung ist, dass Rücksprünge innerhalb einer Zeile ein Indikator für die jeweilige Lese-Erfahrung der Person sind. Je weniger Rücksprünge Leser ausführen, umso leseerfahrener zeigten sich die Probanden (Levy-Schoen u. O'-Regan 1979).

4.2 Kritische Bewertung der Gütekriterien

Im Folgenden sollen aus der Erfahrung, die in dem Experiment gemacht werden konnte, die Gütekriterien des Eye-Tracking-Tests beleuchtet werden. Die Gütekriterien wurden bereits allgemein erläutert (siehe Kapitel 1.2.5 Gütekriterien von Eye-Tracking-Tests, Seite 20) und werden hier um die Erfahrungen konkretisiert, die im Rahmen dieser Arbeit gesammelt werden konnten.

Objektivität

Die 120 Tests wurden nach einem standardisierten Testablauf durchgeführt. Die gewonnenen Daten wurden nach derselben Methode gesichtet, geordnet, analysiert und ausgewertet. Die Auswertung erfolgte mehr oder weniger automatisch, jedoch grundsätzlich softwareunterstützt. Alle Tests wurden von derselben Testleiterin durchgeführt, die eine Expertin für Usability-Tests mit Eye-Tracking ist. Zudem war während der Tests zu Beginn ein Mitarbeiter des Herstellers zugegen, um die Aufzeichnungen in der Software vorzubereiten. Die Testumgebung, das Stimulusmaterial, die Blickregistrierungs-Apparatur und die Testaufgabe waren für alle Teilnehmer identisch. Die Ergebnisse lassen sich somit miteinander vergleichen.

Reliabilität und Validität

Aufgrund mehrerer Faktoren wurden einige Aufzeichnungen ganz oder teilweise durch Messfehler beeinflusst. Gründe waren probandenbedingte Merkmale wie starke Seh- und Schielfehler sowie Abstürze der getesteten Software. Die mit Messfehlern behafteten Daten konnten jedoch identifiziert werden und von den Auswertungen ganz oder teilweise ausgeschlossen werden. Dies gelang beispielsweise, wenn die Augen einer Person nur zeitweise außerhalb des Winkels waren, den die Kamera erfassen kann. Dann konnten die restlichen genauen Daten dennoch verwendet werden. Ob und wie reliabel die Daten waren, konnte auf mehrere Arten kontrolliert werden: zum einen durch die Kalibrierungsergebnisse, zum anderen durch den Trackstatus der Augen (nur linkes Auge, nur rechtes Auge, binokular, kein Auge) in der Video-Wiedergabe.

Die Verdichtung der Daten wurde mit größtmöglicher Sorgfalt vorgenommen. Grundlage dafür war ein manuell erstelltes Protokoll über alle Personen und die Reihenfolge der jeweils besuchten Softwaremasken und der Blickbewegungen der Personen bis zum

Mausklick. Gleichzeitig ist es gerade Gegenstand dieser Arbeit, nur aussagekräftige Parameter auszuwählen und diese miteinander zu vergleichen.
Aufgrund der Reduktion auf messfehlerfreie Evaluationen und Verdichtung relevanter Blickverläufe kann die Güte der Zuverlässigkeit und Validität als ausreichend bezeichnet werden.

Praktikabilität

Der Gesamtaufwand für diese Evaluation lässt sich in mehrere Phasen unterteilen. In der Planungsphase war der Aufwand zum Erstellen der Testaufgabe und des Testablaufs deutlich niedriger als die Installation der zu testenden Software und das Einrichten der Daten (Katalog, Benutzer etc.) in der Software. Zudem stellte die technische Inbetriebnahme des Eye-Trackers, der Anschluss an den Computer und die Aufnahme-Funktionalität einen größeren Aufwand dar. Die Durchführung der Aufzeichnungen und Interviews auf der Computermesse nahm pro Person eine angemessen kurze Zeit in Anspruch. Vom Materialbedarf her war eine externe Festplatte für sechs Gigabyte Blickdaten (Videos, Kalibrationsbilder etc.) notwendig. Alle Testpersonen empfanden, laut eigenen Angaben, die in den Monitor integrierte Augenkamera und die natürliche Bearbeitungssituation als sehr angenehm.

In der Analyse- und Auswertungsphase musste im Vergleich zu traditionellen quantitativen Methoden (Fragebögen etc.) ein größerer zeitlicher Aufwand investiert werden. Einen wesentlichen Einfluss auf die Komplexität der Auswertung hat die Gebrauchstauglichkeit des Eye-Tracking-Systems. Im vorliegenden Fall wäre es wünschenswert gewesen, wenn die ClearView-Software grundsätzlich die Mausklicks mehr in die Blickdaten-Auswertungen integrieren würde. Im Rahmen dieser Arbeit mussten die Fixationen bis zum Klick-Event manuell berechnet werden. Weiterhin war der Direct Excel AOI-Export von ClearView aufgrund der Beschränkung auf maximal 51 Aufnahmen nicht nutzbar. Es lässt sich jedoch in der Vielzahl der neuen Programmversionen von ClearView ein Trend erkennen, dass sich die Gebrauchstauglichkeit zukünftig steigern wird. Eine gute Übersicht zu den aktuellen Anforderungen an Eye-Tracking-Systeme bietet ein Artikel von Schimpfy (Schimpfky 2004).

4.3 Ausblick

Die erste Fragestellung in dieser Arbeit war, eine geeignete Auswahl von potenziell aussagekräftigen Parametern zu treffen. In einem zweiten Schritt wurden die tatsächlich relevanten Parameter experimentell ermittelt und hinsichtlich ihrer Aussagekraft untersucht. Daran könnte sich nun noch eine systematische Untersuchung der Abhängigkeit der Parameter untereinander anschließen. Es wäre aufschlussreich zu sehen, ob sich speziell aus der Abhängigkeit von zwei oder mehreren Parametern spezifische Aussagemöglichkeiten ergeben.

Zukünftiger Forschungsbedarf lässt sich auch in der Analyse und der Auswertung des Pupillendurchmessers als Maß für den kognitiven Verarbeitungsaufwand ausmachen. Es wäre vielversprechend, gerade die Vorgänge bei Richtungsänderungen (backtracks) zu untersuchen. Bei den backtracks findet ein Strategiewechsel statt, der mit gedanklichem Umdenken verbunden ist. Der Pupillendurchmesser und seine Änderungen ließen sich anhand offensichtlicher Verarbeitungsaufwände validieren.

Auch die Automatisierung der Auswertung von entscheidungs- und handlungsrelevanten Blickverlaufs-Sequenzen durch Softwaretools wäre erstrebenswert. Die Entwicklung mathematischer Algorithmen zur Ermittlung weiterer abgeleiteter Parameter, wie z. B. die Übergangsdichte als Ableitung der Übergangshäufigkeit könnte eine weitere Forschungsarbeit darstellen. Damit verbunden werden könnte die Entwicklung anschaulicher softwaregestützter Darstellungsformen zur Präsentation von Interessengebieten und Häufigkeiten.

Auf der Tagung der Usability Professionals 2006, Gelsenkirchen, konnte ein Trend in Methoden und Tools beobachtet werden. Usability-Tests mit Blickbewegungsmessung sind mit einem hohen technischen und zeitlichen Aufwand verbunden und praktisch nur im Labor durchzuführen. Folglich werden alternative Methoden diskutiert, die mit weniger Nachteilen behaftet sind. Zum Beispiel wird untersucht, ob Covering-Methoden zur Erfassung der Aufmerksamkeitsverteilung bei der Interaktion mit Web-Seiten geeignet sind. Bei Site-Covering wird eine Webseite in Interessengebiete eingeteilt. Diese werden mit transparenten Flächen überdeckt, die die Information unleserlich machen. Die Benutzer können nun Feld für Feld mittels Klick aufdecken. Die Aufdeckvorgänge machen die Aufmerksamkeitssteuerung sichtbar.

5 Zusammenfassung

Die vorliegende Arbeit versucht, einen Beitrag zu leisten für die heute noch fehlende Standardisierung der Interpretation von Blickbewegungen bei Software-Usability-Tests. Zu diesem Zweck wurde eine experimentelle Studie zur Ermittlung von Korrelationen zwischen der Blickbewegungsanalyse und einem Referenzmodell für Experten-Bewertungen von Softwareanwendungen durchgeführt. Diese Arbeit soll die Blickbewegungsregistrierung als Erhebungsmethode validieren, um die Designdomänen einer Anwendung zu bewerten.

Die Expertenkriterien zur Evaluation von Software stammen aus dem WebSCORE Referenzmodell, entwickelt vom Fraunhofer-Institut für Arbeitswirtschaft und Organisation (IAO), Stuttgart. Die Domänen „Struktur“, „Interaktion“ und „Präsentation“ sind für die quantitative Evaluierungsmethode Eye-Tracking relevant und wurden deshalb aus dem Gesamtmodell extrahiert. In der Literatur werden je nach Anwendungskontext zahlreiche Eye-Tracking-Parameter verwendet. Es wurden davon die Parameter mit Relevanz für Usability-Evaluation ausgewählt.

Als Forschungsmethode wurde das explorative Verfahren angewendet. Die Offenheit des Verfahrens ist ein wichtiger Punkt, damit auch ganz neue Erkenntnisse aus dem Experiment gewonnen werden konnten. Das Experiment wurde am Beispiel einer webbasierten Software-Anwendung für elektronische Bedarfsanforderungen durchgeführt. Die Stichprobe bestand aus 120 Probanden. Die Testaufgabe, das Bestellen eines Artikels, ermöglichte verschiedene Umsetzungen in der Software und ließ damit nicht nur triviale Probleme erkennen, sondern auch komplexe Probleme in der Software. Die Testaufgabe wurde in Dialogschritte untergliedert.

Für die Analyse der gewonnenen Blickbewegungsdaten wurden zunächst aus den Rohdaten, die die Blicke pro Aufzeichnung in einer zeitlichen Auflösung von 20 ms beinhalten, die Fixationen gefiltert (mindestens 100 ms im Radius von 30 Pixeln). Weiterhin wurden die Ereignisse der Benutzerinteraktion wie Mausklick oder Tasteneingabe selektiert. Die Auswertung erfolgt sowohl individuell als auch summarisch für Testpersonen, die vergleichbare Aufnahmen aufweisen. Ein weiterer Schritt der Analyse war die Zuordnung der potenziell vorab ermittelten als auch der im Experiment real offenbarten Problemstellen zu den Designdomänen „Struktur“, „Interaktion“ und „Präsentation“.

Die Probleme mit der höchsten Gewichtung konnten dabei zentralen Oberflächen-Komponenten zugeordnet werden. Oberflächen-Komponenten sind zum Beispiel Bäume, Menüs oder Buttons.

Für die Ermittlung, welche Parameter Aussagen zu den Designdomänen zulassen, wurden Vergleichssituationen bei den Oberflächenelementen ausgewählt und deren Blickdaten untersucht. Beispielsweise wurden die Suchzeiten bei zwei Zuständen eines Verzeichnis-Katalogbaumes gegenübergestellt: Suchzeit bei drei sichtbaren Einträgen innerhalb eines Verzeichnisses vs. Suchzeit bei 14 sichtbaren Einträgen. Die Blickdaten auf diesen Objekten wurden dabei jeweils über alle relevanten Aufzeichnungen kumuliert. Diese Auswertungen, basierend auf dem Experiment, wurden dann wieder generalisiert und in die Systematik in Form von Kausalzusammenhängen festgehalten (je kürzer die Parameter der Suchzeit, umso besser ist die Designdomäne „Struktur“). Die Auswertung führt wiederum zu einer Tabelle, die die Designdomänen den jeweils aufschlussreichen Eye-Tracking-Parametern und deren Wertebereichen gegenüberstellt.

Das Experiment hat sieben Eye-Tracking-Parameter ergeben, die aussagekräftig für die Usability einer Software sind. Die Länge der Blickpfaddauer ist ein Maß der Effizienz der Aufgabenbearbeitung und damit ein Leistungsdatum. Mittels des zeitlichen Parameters der Suchzeit lässt sich eine Baumstruktur bewerten. Je kürzer die Suchzeit ist, desto mehr innere Ordnung besitzt eine Struktur. Wie ein einzelner Baumeintrag wahrgenommen wird, lässt sich über den Häufigkeitsparameter der Fixationen von Objekten bewerten. Für Interaktionselemente wie Menüs und Buttons sind ebenfalls die Aneinanderreihung mehrerer Fixationen und Sakkaden ein aussagekräftiges Maß. Mit Richtungsänderungen lassen sich erwartungskonform angeordnete Interaktionselemente, wie zum Beispiel Buttons, untersuchen. Je weniger Blickrichtungsänderungen der Benutzer vollziehen muss, desto eher entspricht die Software seinen Erwartungen. Wie mühelos Benutzer mit Menüs interagieren können, darüber gibt der Parameter der Übergangshäufigkeiten und der davon abgeleitete Parameter der Übergangsdichte Aufschluss. Das Seitenlayout lässt sich durch kumulierte Übergangszeiten ermitteln. Das sind Zeiten ohne bewusste Informationsverarbeitung. Die Übergangszeiten sollten möglichst kurz sein.

6 Literatur

Aaltonen A, Hyrskykari A, Räihä KJ: 101 Spots, or How Do Users Read Menus? Proceedings of CHT'98: 132-139 (1998)

Anderson JR: Kognitive Psychologie. Spektrum Akademischer Verlag, Heidelberg (2001)

BITKOM: Der elektronische Handel boomt. (2007) http://www.bitkom.de/de/presse/8477_43665.aspx, Zugriff: 22.04.2007

Blume M, Stokar D, Seewald F: Usability-Evaluation: Egal wer's macht? Ein schweizer Fallbeispiel. In: Hassenzahl M, Peissner M (Hrsg) Usability Professionals 2005, Fraunhofer IAO, Stuttgart, S. 140-143 (2005)

Byrne MD, Anderson JR, Douglass S, Matessa M: Eye Tracking the Visual Search of Click-Down Menus. Proceedings of CHT'99: 402-409 (1999)

Cowen L: An eye movement analysis of web-page usability. (2001) http://www.lauracowen.co.uk/download/mres_dissertation.pdf, Zugriff: 22.04.2007

DIN-EN-ISO: DIN EN ISO 9241-11: Ergonomische Anforderungen für Bürotätigkeiten und Bildschirmgeräten - Teil 11: Anforderungen an die Gebrauchstauglichkeit. Beuth, Berlin (1999)

Duchowski AT: Eye-Tracking Methodology: Theory and Practice. Springer, London, S. 12 (2003)

Dürrwächter U: Analyse der Blickbewegungen von Kindern mit einer Lese- und Rechtschreibstörung. (2003) http://w210.ub.uni-tuebingen.de/dbt/volltexte/2003/943/pdf/Dissertation_Duerrwaechter.pdf, Zugriff: 20-04-2007

Dzida W, Wiethoff M, Arnold AG: ERGOguide: The Quality Assessment Guide to Ergonomic Software. St. Augustin (1993)

Eger N, Ball LJ, Stevens R, Dodd J: Validating the Use of Eye-Movement Replay to Cue Retrospective Verbal Protocols in Online Usability Testing. Lancaster University, Lancaster (2005)

Fairclough SH, Ashby MC, Parkes AM: In-vehicle dislays, visual workload and usability evaluation. In: Gale AG, Brown ID, Haslegrave CM, Kruysse HW, Taylor SP (Hrsg) Vision in Vehicles - IV, Elsevier, Amsterdam, S. 245-254 (1993)

Fitts PM, Jones RE, Milton JL: Eye movements of aircraft pilots during instrument-landing approaches. Aeronautical Engineering Review 9: 24-29 (1950)

Geis T, Dzida W, Redtenbacher W: Specifying usability requirements and test criteria for interactive systems. Federal Institute for Occupational Safety and Health, Dortmund (2004)

Glaser C, Hurtienne J, Mohs C: Intuitiv und effizient - Thesen zur Intuitivität von User Interfaces. In: Hassenzahl M, Peissner M (Hrsg) Usability Professionals 2005, Fraunhofer Institut für Arbeitswirtschaft und Organisation, Stuttgart, S. 146-149 (2005)

Goldberg JH, Kotval XP: Eye movement-based evaluation of the computer interface. In: Kumar SK (Hrsg) Advances in Occupational Ergonomics and Safety, IOS Press, Amsterdam, S. 529-532 (1998)

Goldberg JH, Kotval XP: Eye Movement-Derived Measures of Interface Usability. Ergonomics (1999)

Goldberg JH, Wichansky AM: Eye Tracking in Usability Evaluation: A Practitioner's Guide. In: Hyona J, Radach R, Deubel H (Hrsg) The Mind's Eye, Elsevier Science BV, Amsterdam, S. 493-516 (2003)

Heidmann F: Auf dem Leuchttisch. Kommune21 1: 14-15 (2004)

Heidmann F, Ziegler J: WebSCORE - A structured method for evaluating web applications. Proceedings of the 6th International Conference WWDU 2002 (2002)

Hermann F: Qualitätsmerkmale erfolgreicher Web-Applikationen (Seminarunterlagen). Fraunhofer Institut für Arbeitswirtschaft und Organisation, Stuttgart (2003)

Heuer J: Expertenevaluation. In: Heinsen S, Vogt P (Hrsg) Usability praktisch einsetzen: Handbuch für Software, Web, Mobile Devices und andere interaktive Produkte, Hanser, München, S. 115-135 (2003)

Jacob RJK, Karn KS: Eye Tracking in Human-Computer Interaction and Usability Research: Ready to Deliver the Promises (Section Commentary). In: Hyona J, Radach R, Deubel H (Hrsg) The Mind's Eye: Cognitive and Applied Aspects of Eye Movements Research, Elsevier Science, Amsterdam, S. 573-605 (2003)

Krings H: Gebrauchstauglichkeit von Technischer Dokumentation: Vertrauen ist gut - Testen ist besser. technische kommunikation 5: 26-29 (2003)

Krömker H: Die Welt der Benutzungsfreundlichkeit. In: Hennig J, Tjarks-Sobhani M (Hrsg) Verständlichkeit und Nutzungsfreundlichkeit von technischer Dokumentation, Schmidt-Römhild, Lübeck, S. 22-32 (1999)

Levy-Schoen A, O'Regan K: The control of eye movements. Processing of visible language 1 (1979)

Lieberman H, Wagner EJ: End-User Debugging for E-Commerce. In: Johnson WL, André E, Domingue J (Hrsg) IUI 2003 International Conference on Intelligent User Interfaces, ACM Press, Miami, Florida, S. 257-259 (2003)

Lienert GA, Raatz U: Testaufbau und Testanalyse. Beltz, Weinheim, S. 7-12 (1994)

Mackworth NH: Stimulus density limits the useful field of view. In: Monty RA, Senders JW (Hrsg) Eye movement and psychological processes, Lawrence Erlbaum, Hillsdale, New Jersey, S. 307-321 (1976)

Masuda HM, Ichikawa KI, Go KG, Imamiya AI Measuring eye movements and mouse-pointing patterns using thumbnail images. (2002) http://www.noldus.com/events/mb2002/program/abstracts/masuda.html, Zugriff: 22.04.2007

MEYERS-Lexikon-online: Intuition. (2007) http://lexikon.meyers.de/meyers/Intuition, Zugriff: 22.04.2007

Microsoft: The Windows Interface Guidelines for Software Design: An Application Design Guide. Microsoft Press (1995)

Miller G: The Magical Number 7, Plus or Minus Two: Some Limits on Our Capacity for Processing Information. Psychological Review 63: 81-97 (1956)

Moll T: Über Methoden zur Analyse und Evaluation interaktiver Computersysteme. In: Fähnrich KP (Hrsg) Software-Ergonomie, Oldenbourg, München, S. 179-190 (1987)

Mussgnug J, Stowasser S: Blickbewegungsanalyse zur Bewertung piktographischer Visualisierungen. MMI-interaktiv Journal 6: 51-64 (2003)

Mussgnug J, Stowasser S: Ein Schema zur Auswahl geeigneter Usability-Methoden - dargestellt am Beispiel der Blickbewegungsanalyse. In: Hassenzahl M, Peissner M (Hrsg) Usability Professionals 2004, Fraunhofer IAO, Stuttgart, S. 28-30 (2004)

Nielsen J: Usability Engineering. Morgan Kaufmann, San Francisco (1993)

Nielsen J, Tahir M: Building Web Sites With Depth. (2001) http://www.webtechniques.com/archives/2001/02/nielsen/, Zugriff: 06.01.2007

Outing S, Ruel L: Eyetrack III. (2004) http://poynterextra.org/eyetrack2004/main.htm, Zugriff: 10.12.2006

Pashler HE: The Psychology of Attention. MIT Press, Cambridge, Massachusetts, S. 217-219 (1997)

Peissner M, Heidmann F, Wagner I: WebSCORE Expert Screening - a low-budget method for optimizing web applications. In: Jacko J, Stephanidis C (Hrsg) Human-Computer-Interaction - Theory and Practice Part I, S. 838 - 842 (2003)

Posner MI, Snyder CRR, Davidson BJ: Attention and the Detection of Signals. Experimental Psychology: General 109: 160-174 (1980)

Rosenthal R: Experimenter effects in behavioral research. Appleton Century Crofts, New York (1966)

Rötting M: Parametersystematik der Augen- und Blickbewegungen. Shaker, Aachen (2001)

Rubin J: Handbook of Usability Testing: How to plan, design and conduct effective usability tests Wiley technical communication library, Hoboken, S. 46 (1994)

Sarodnick F, Brau H: Methoden der Usability Evaluation, Wissenschaftliche Grundlagen und praktische Anwendung. Hans Huber, Bern (2006)

Scheier C, Heinsen S: Aufmerksamkeitsanalyse. In: Heinsen S, Vogt P (Hrsg) Usability praktisch einsetzen: Handbuch für Software, Web, Mobile Devices und andere interaktive Produkte, Hanser, München, S. 154-169 (2003)

Schiessl M, Duda S, Thölke A, Fischer R: Eye tracking and its application in usability and media research. (2003) http://useworld.net/servlet/handlearticle?obj_id=582&cat_id=40&vie=true, Zugriff: 22.04.2007

Schimpfky N: Aufklärung von Nutzungsproblemen mit interaktiven Systemen mittels Eye-Tracking: Ein methodischer Ansatz zur Evaluierung von Bildschirmanzeigen. Fraunhofer Institut für Autonome Intelligente Systeme, Bonn (2003)

Schimpfky N: Auf dem Wege zu gebrauchstauglichen Eye-Tracking-Systemen. i-com 3: 4-8 (2004)

Schweibenz W, Thissen F: Qualität im Web. Springer, Berlin, Heidelberg, S. 60-70 (2003)

Schweigert M: Fahrerblickverhalten und Nebenaufgaben. (2003) http://tumb1.biblio.tu-muenchen.de/publ/diss/mw/2003/schweigert.pdf, Zugriff: 22.04.2007

Sun-Microsystems: Java Look and Feel Design Guidelines. Addison Wesley (1999)

Tobii: User Manual Tobii Eye Tracker ClearView analysis software. Tobii Technology, Stockholm, S. 52-55 (2006)

Wertheimer M: Drei Abhandlungen zur Gestalttheorie. Verlag der Philosophischen Akademie, Erlangen (1925)

Wickens CD, Gordon E, Liu Y: An introduction to human factors engineering. Addison Wesley, New York (1998)

Wippermann P: Simplexity - Die neue effiziente Klugheit. (2006) http://www.trendbuero.de/index.php?f_articleId=1204, Zugriff: 06.01.2007

Zimbardo PG: Psychologie. 5. Aufl, Springer, Berlin (1992)

7 Danksagung

Abschließend möchte ich mich bei all denjenigen bedanken, die auf unterschiedliche Art und Weise zum Entstehen dieser Arbeit beigetragen haben.

Prof. Dr. Harald C. Traue gilt mein erster und besonderer Dank für seine Betreuung. Durch viele stimulierende Diskussionen und wertvolle Anregungen hat er es geschafft, mich fortwährend zu motivieren und für dieses Thema zu begeistern. Die Offenheit und das Vertrauen von Prof. Traue ermöglichte mir, mich dem Thema auf meine Art und Weise zu nähern.

Zur Aufnahme dieser Arbeit ermutigt haben mich Prof. Robert Schäflein-Armbruster von der Hochschule Furtwangen sowie mein langjähriger, sehr geschätzter Freund Herr Dr. Klaus-Dieter Häberle von der Hochschule Ulm.

Ich danke Dr. Petra Ritschi, die den Kontakt zu Prof. Dr. Traue und dem Forschungsbereich hergestellt hat und mir auch während der Promotion kompetenten Rat gab.

Unterstützt haben mich weiterhin Frau Dr. Claudia Schröck-Pauli, Frau Dr. Heike Koch-Beuttenmüller und Herr Dr. Volker Typke vom Kommunikations- und Informationszentrum, Universität Ulm, durch diverse Fachgespräche.

Der Wilken GmbH, Ulm, insbesondere dem Geschäftsführer Dr. Andreas Lied, danke ich für die Verfügbarkeit der Wilken Software für das Experiment. Vielen Dank auch an meine ehemaligen Kollegen für die Unterstützung auf der CeBIT 2006.

Mein Dank gilt auch Frau Susanne Schade für die zahlreichen Gespräche über Wissenschaft und unsere beruflichen Perspektiven. Herrn Michael Trögele möchte ich für seine unterstützenden Tätigkeiten bei der Sichtung der Blickdaten danken. Herrn Michael Ros danke ich für seine freundliche Unterstützung beim Lektorat.

Schließlich danke ich meinem lieben Ehemann Karlheinz, der mir stets den Rücken freihielt, damit ich diese Dissertation anfertigen konnte. Er hat mich in vielerlei Hinsicht unterstützt, half bei technischen Themen und war mir immer ein sehr wertvoller Diskussionspartner.

8 Anhang: Bildschirmseiten



Abbildung 26: Anmeldungs-Bildschirm

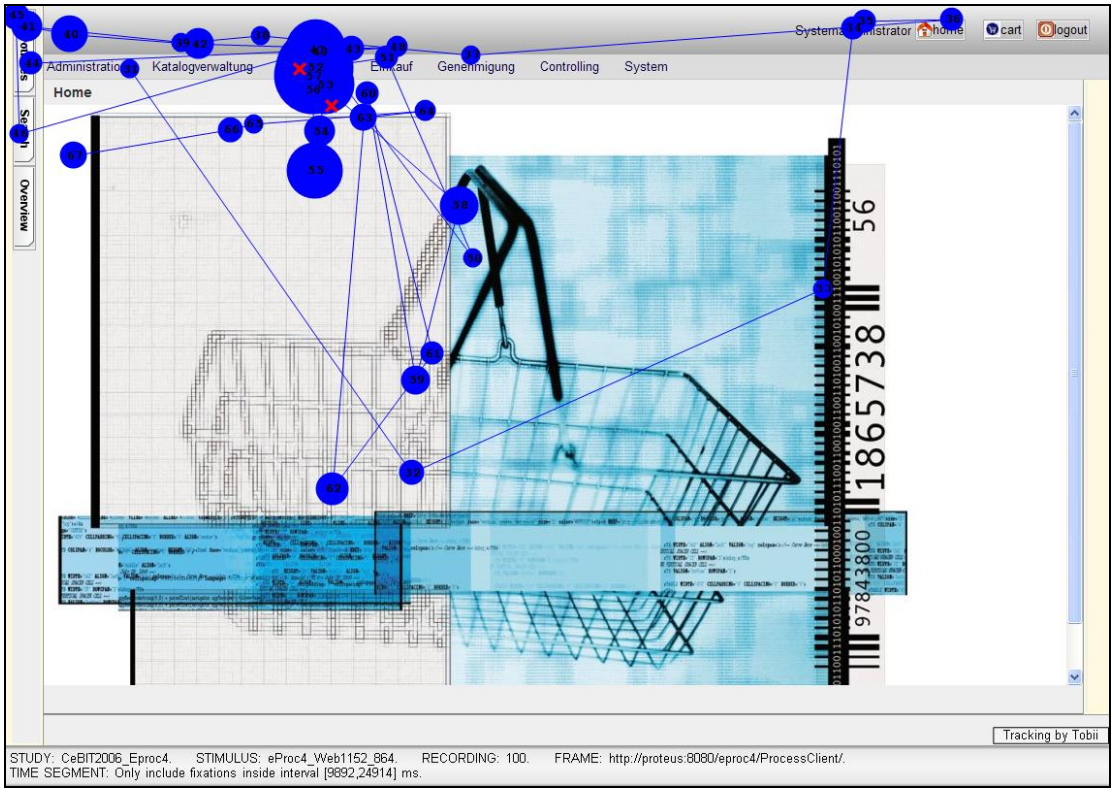

Abbildung 27: Startseite mit Hauptmenü

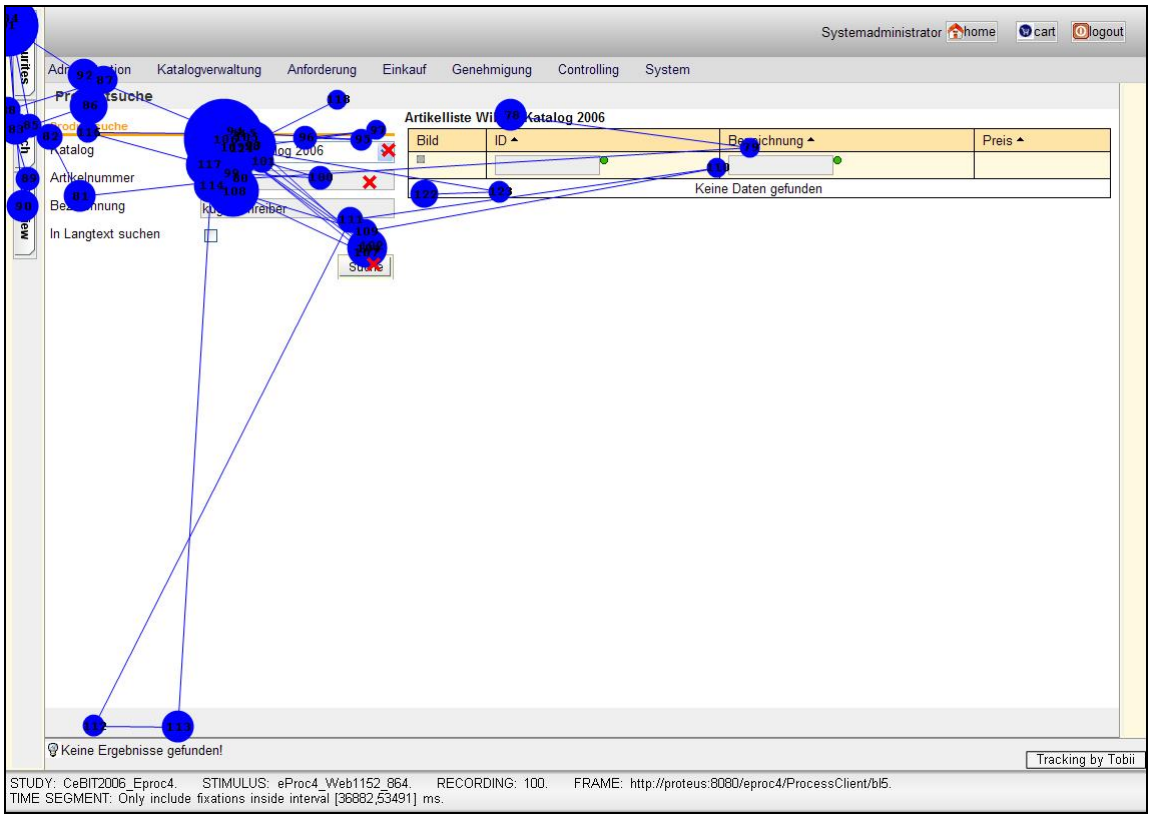

Abbildung 28: Seite Produktsuche

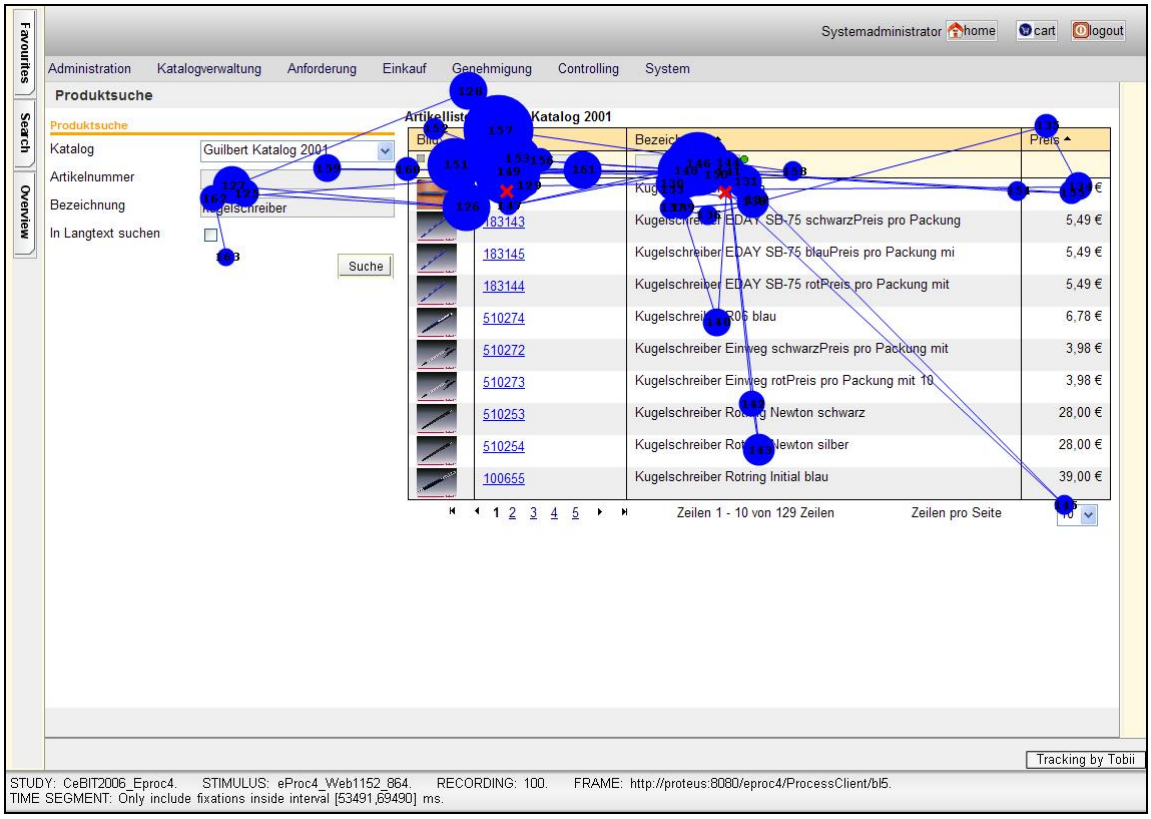

Abbildung 29: Seite Produktsuche mit Artikelliste

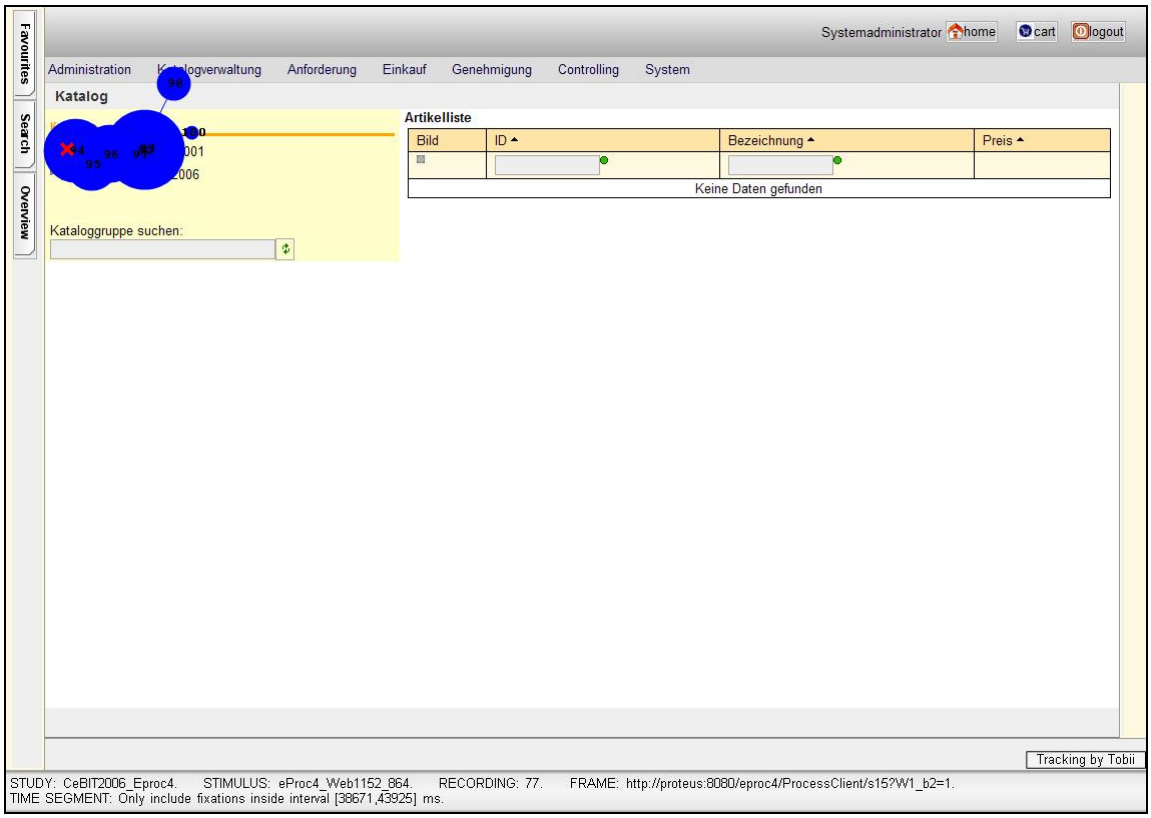

Abbildung 30: Seite Katalog, geschlossener Baum

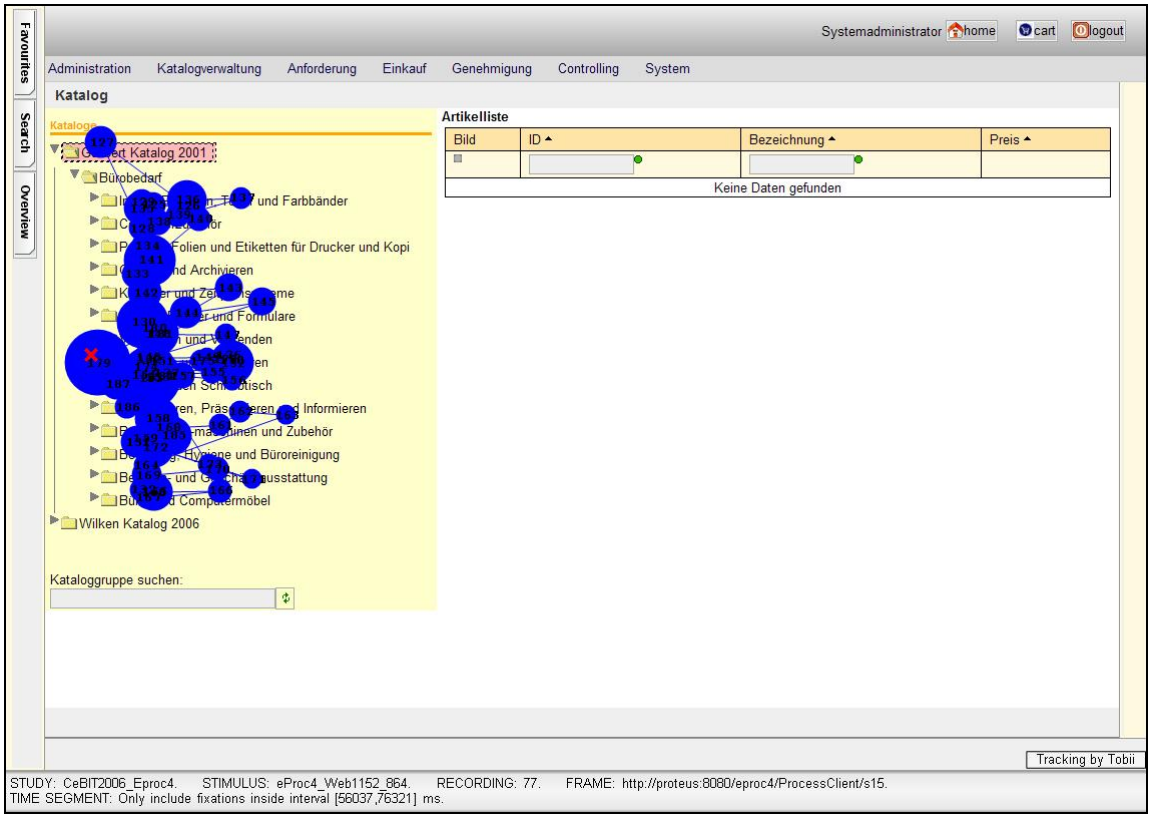

Abbildung 31: Seite Katalog, erste Baumebene expandiert

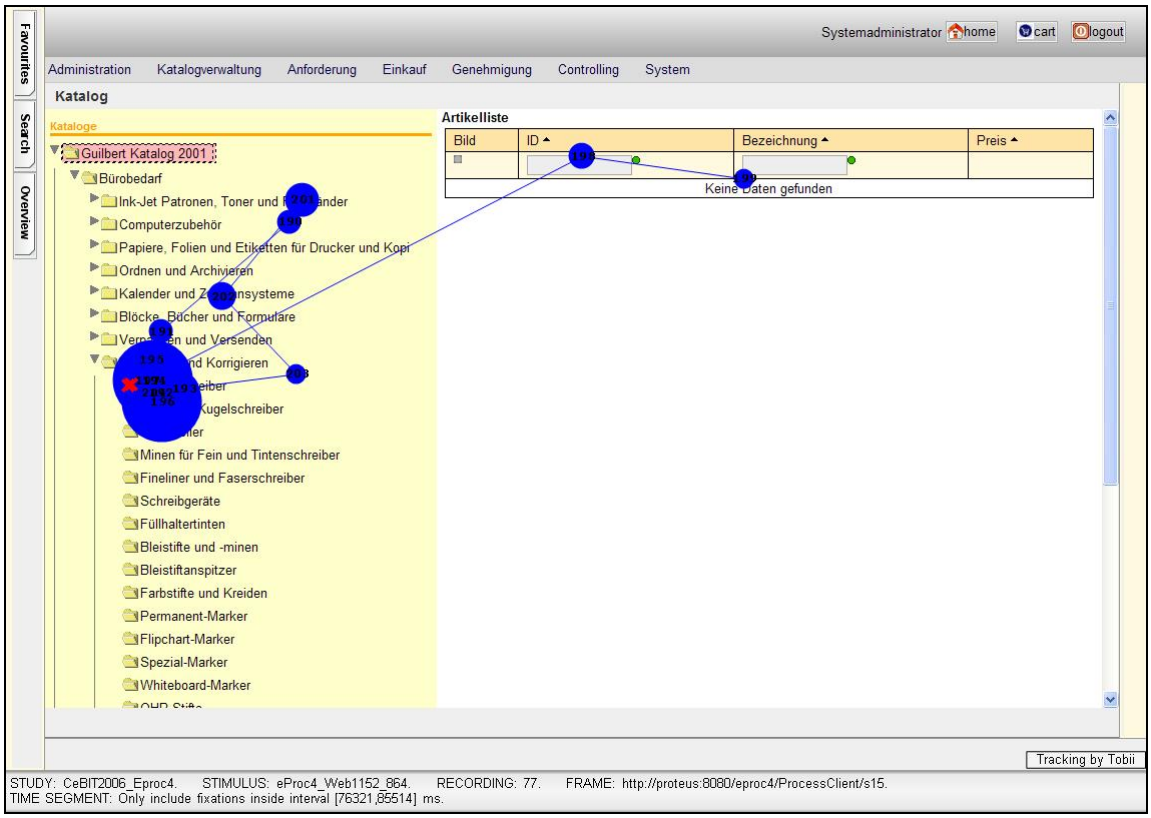

Abbildung 32: Seite Katalog, zweite Baumebene expandiert

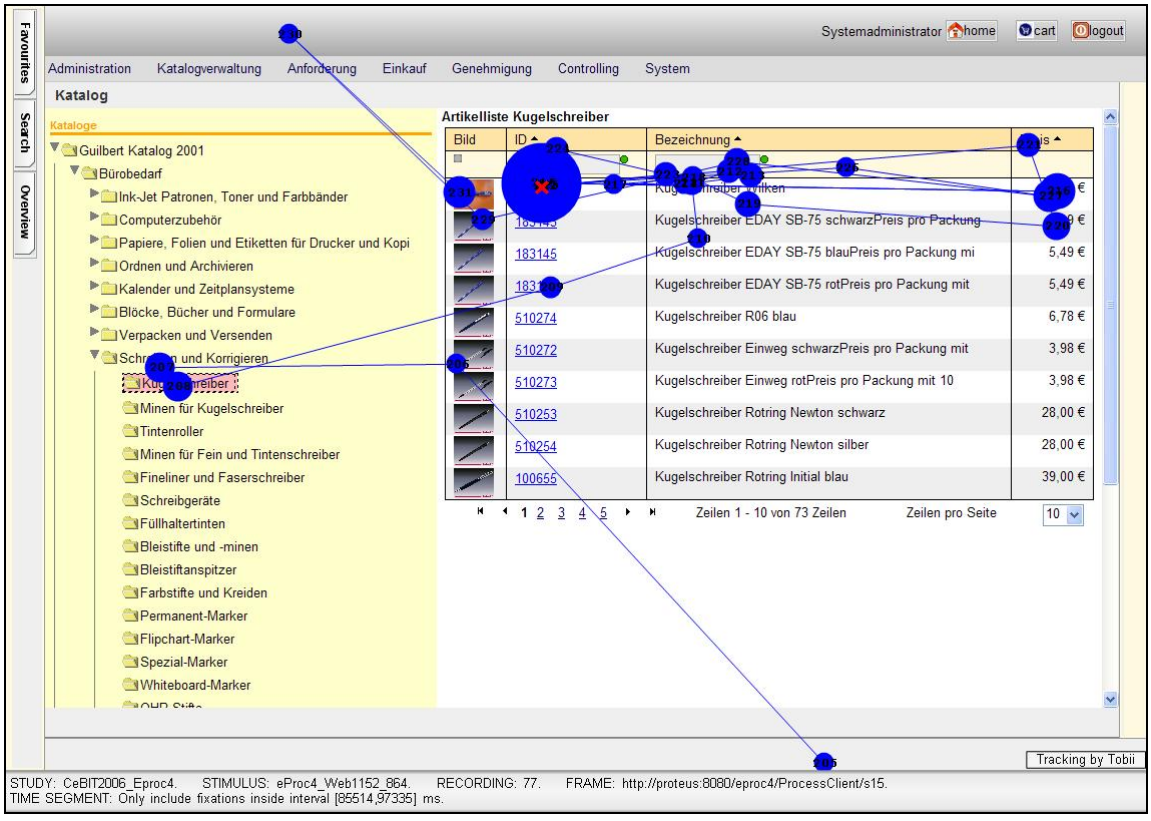

Abbildung 33: Seite Katalog mit Artikelliste

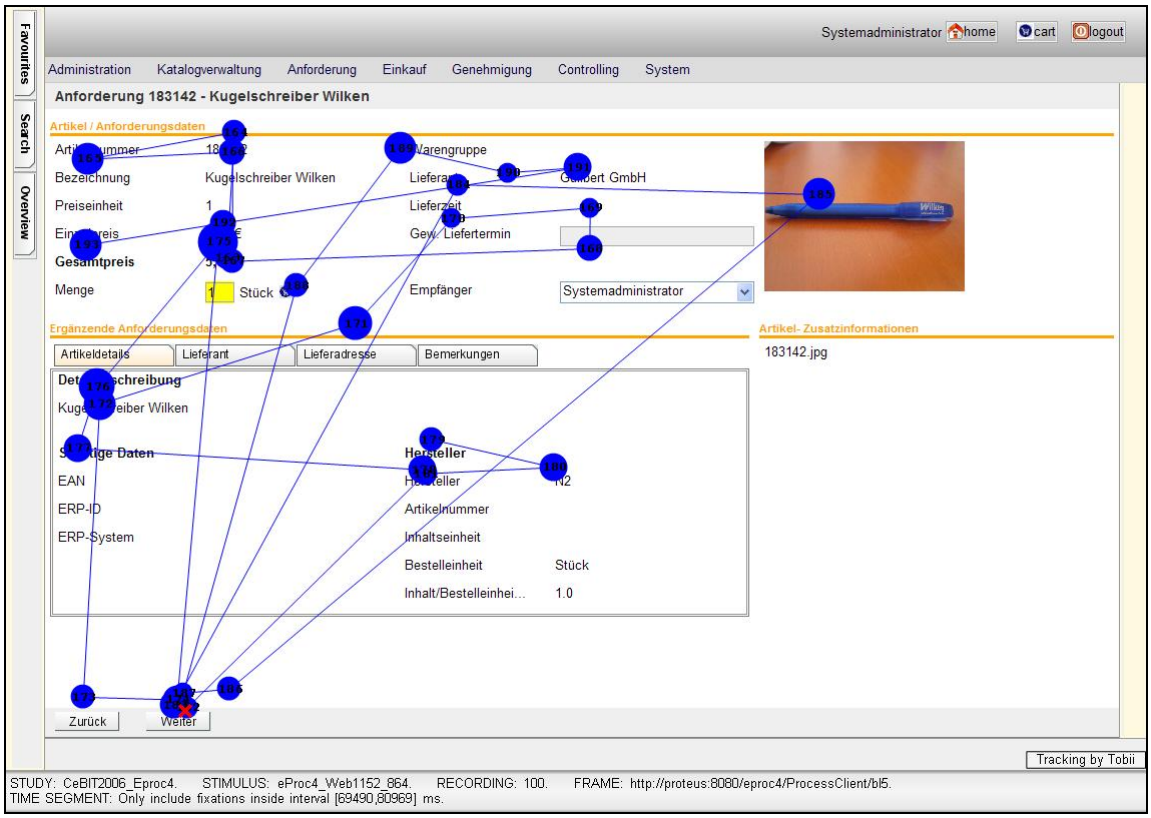

Abbildung 34: Seite Anforderung Artikel

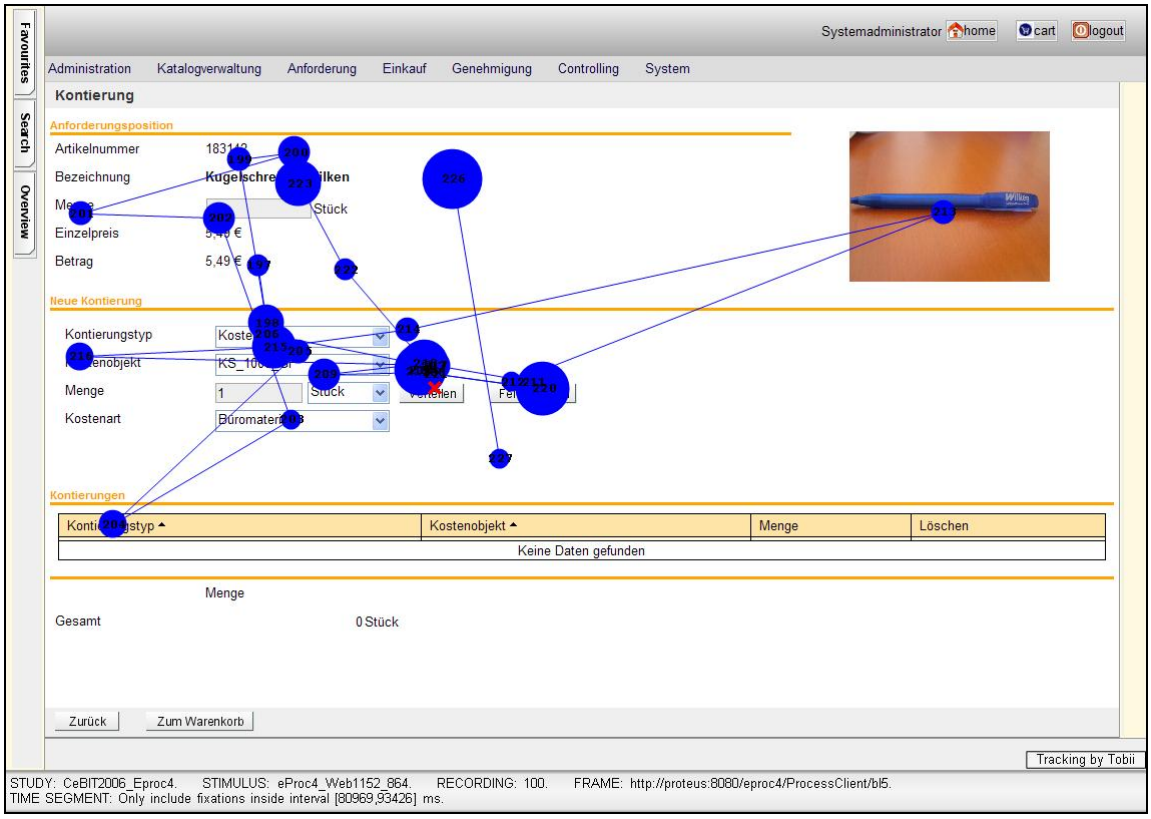

Abbildung 35: Seite Kontierung

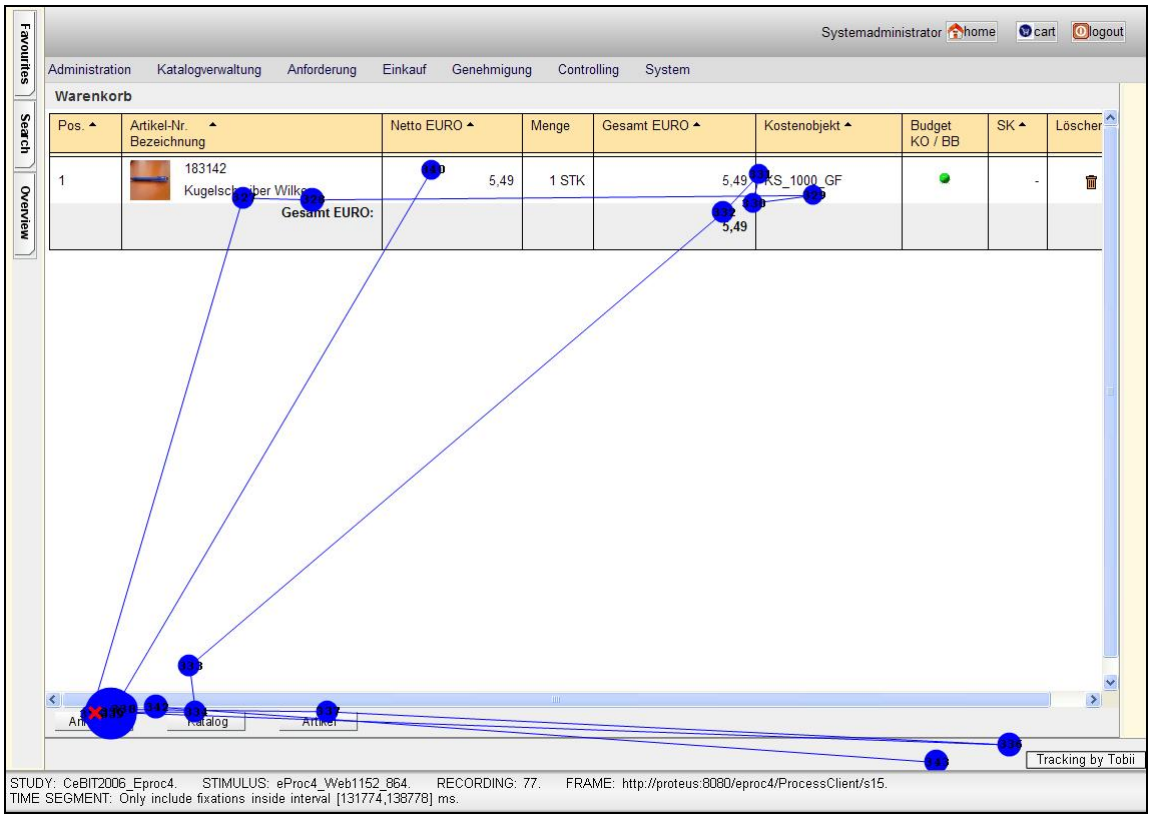

Abbildung 36: Seite Warenkorb

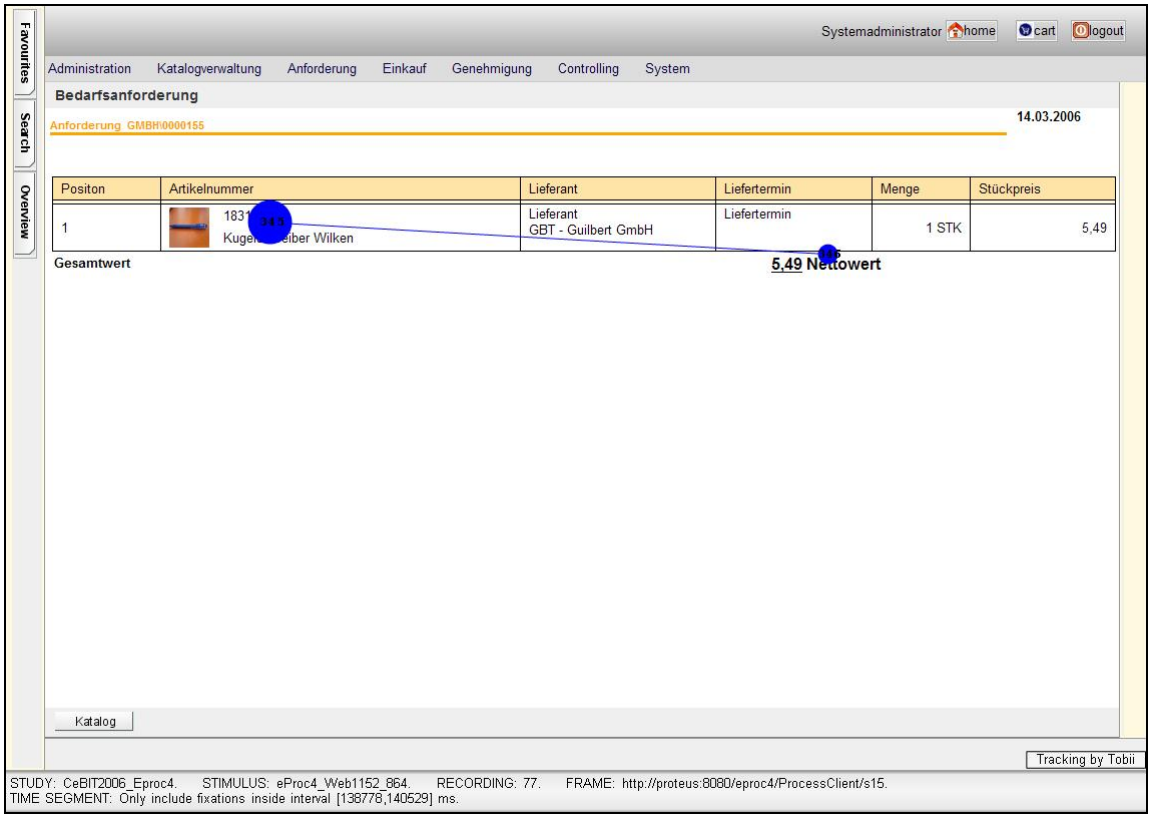

Abbildung 37: Seite Bedarfsanforderung