

RWTH Aachen University
Media Computing Group
Prof. Dr. Jan Borchers
Mensch-Maschine-Interaktion (HCI)
SS 2007

Evaluationstechniken

Ivo Jacobs
Sharon Bockholdt

03. Mai 2007

Inhaltsverzeichnis

1. Einleitung	1
2. Grundlagen der Evaluation	1
3. Evaluation durch Experten	2
3.1 Kognitiver Durchgang.....	2
3.2 Heuristische Evaluation	3
3.3 Modellbasierte Evaluation	4
3.4 Evaluation durch Literaturrecherche.....	4
4. Evaluation durch Benutzer.....	5
4.1 Grundsätzliches zu experimentellen Designs	5
4.2 Beobachtungstechniken	6
4.3 Befragungstechniken	7
4.4 Beobachten physiologischer Reaktionen.....	7
6. Zusammenfassung & Fazit	9
7. Literaturangaben.....	10

1. Einleitung

Im Folgenden möchten wir uns im Rahmen des Proseminars „Mensch-Maschine-Interaktion (HCI)“ näher mit Evaluationstechniken im Designprozess interaktiver Systeme beschäftigen.

Dabei beziehen wir uns im Wesentlichen auf Kapitel 9 „Evaluation techniques“ (S. 319-364) aus dem Buch „Human Computer Interaction“ (3. Auflage, Prentice Hall, 2004.) von Alan Dix, Janet Finlay, Gregory Abowd und Russell Beale.¹ Zusätzlich haben wir weitere Quellen recherchiert, die ebenfalls im Literaturverzeichnis angeführt sind.

Einleitend beginnen wir mit einer kurzen Begriffsklärung: Was ist Evaluation?

„Evaluation [...v...], die; -, -en <lat.> (Bewertung; Beurteilung); evaluieren“²

Ursprünglich wurde der Begriff im Zusammenhang mit der Bewertung der Arbeit von Bildungsinstitutionen verwendet, hat sich jedoch heute in vielen Bereichen als Synonym für Analyse- und Bewertungsmethoden etabliert.³

Wir werden uns mit der Evaluation bzw. Evaluationstechniken speziell im Designprozess interaktiver Systeme beschäftigen und zunächst einige Grundlagen der Evaluation klären.

2. Grundlagen der Evaluation

Beim Designprozess interaktiver Systeme ist eine begleitende Evaluation, also Bewertung durch verschiedene Evaluationstechniken, während und bereits vor der Entwicklung enorm wichtig. Die Evaluation kann bereits frühzeitig eine Rückmeldung hinsichtlich der Verbesserung und Optimierung des Systems aufzeigen und somit kostenintensive Änderungen zu einem späteren Entwicklungszeitpunkt ersparen.

Es gibt verschiedene Möglichkeiten Evaluation zu kategorisieren. Eine ist die Einteilung in summative und formative Evaluation.

Unter der summativen Evaluation versteht man die Überprüfung eines Systems auf bestimmte Kriterien hin, die durch quantitative Daten (z.B. durchschnittliche Zeit für die Bearbeitung einer Aufgabe bei der Beobachtung von Benutzern) eine Entscheidungsgrundlage für die Auswahl zwischen zwei oder mehr Versionen zur Verfügung stellt.

Die formative Evaluation hingegen bewertet einen Prototypen eines Systems durch qualitative Aussagen (z.B. Cognitive Walkthrough durch einen Experten) um auf Grundlage dessen Verbesserungsvorschläge herauszuarbeiten, die das Gestaltungskonzept an die Benutzeranforderungen anpasst.

Kurz gesagt:

*"When the cook tastes the soup, that's formative;
when the guests taste the soup, that's summative." - Robert Stakes⁴*

Im Sinne dieses Zitates unterteilen wir den Hauptteil dieser Arbeit in **Evaluation durch Experten** und **Evaluation durch den Benutzer** (ähnlich wie in „Human Computer Interaction“). Diese zwei Arten von Evaluationstechniken werden in den nächsten zwei Kapiteln näher behandelt.

3. Evaluation durch Experten

3.1 Kognitiver Durchgang

Der kognitive Durchgang, standardmäßig Cognitive Walkthrough genannt, ist eine analytische Evaluationstechnik, die von einem Experten, möglichst zu einem frühen Zeitpunkt des Entwicklungsprozesses, durchgeführt wird. Dabei „schlüpft“ der Experte in die Rolle des Benutzers und analysiert konkrete Handlungsabläufe, die sich bei der Benutzung der Schnittstelle (Interface) zwischen Mensch und Maschine ergeben. Jeder einzelne Schritt wird dabei genau geprüft und dokumentiert, um mögliche Fehlerquellen und Probleme bei der Benutzung herauszufiltern. Es werden keine empirischen Daten erhoben, sondern Urteile eines oder mehrerer Experten eingeholt. Dabei liegt das Hauptaugenmerk vor allem auf der Erlernbarkeit der Bedienung des Systems durch Ausprobieren. Es gilt herauszufinden, ob das System die Erwartungen eines potentiellen Benutzers erfüllt.

Bevor man einen kognitiven Durchgang starten kann, werden vier Dinge benötigt:

1. *Beschreibung der Spezifikation oder des Prototyps des Systems.*
2. *Beschreibung der Aufgaben: Welche Aufgabe(n) soll(en) analysiert werden?*
3. *Liste der Handlungen: Welche einzelnen Schritte sollen bei der Durchführung einer Aufgabe getätigt werden?*
4. *Klärung der Zielpersonen: Wer sind die Nutzer? Wer sind die Benutzergruppen? Welches Vorwissen und welche Erfahrungen bringen sie mit?*

(vgl. „Human Computer Interaction“, S. 321)

Um möglichst nah an potentielle Fehlerquellen heranzukommen, stellt sich der Experte bei jedem Schritt, den er bei der Aufgabe tätigt, folgende vier Leitfragen:

1. *Handelt der Benutzer so, dass er das von ihm erwartete Ergebnis erzielt?*
2. *Sieht der Benutzer, dass die Aktion verfügbar ist und er sie ausführen kann?*
3. *Verbindet der Benutzer die korrekte Aktion mit dem von ihm gewünschten Effekt?*
4. *Erhält der Benutzer nach korrekt getätigter Aktion ausreichend Rückmeldung (Feedback) vom System und versteht auch, was er mit der Aktion erreicht hat?*

(vgl. *Human Computer Interaction*, S. 321/322)

Das Ziel eines kognitiven Durchgangs ist es also, die unterschiedlichen Erwartungen an das Produkt von Entwickler und Benutzer aufzudecken. Vor allem soll dem Entwickler vermittelt werden, wo und warum es zu Problemen bei der Interaktion zwischen Benutzer und Produkt kommen kann.

3.2 Heuristische Evaluation

Das Wort Heuristik entstammt dem Griechischen und beschreibt „die Lehre von den Methoden zur Auffindung neuer wissenschaftlicher Erkenntnisse“.⁵ Die heuristische Evaluation ist beim Designprozess ein Verfahren, bei dem Benutzeroberflächen anhand von festgelegten Gestaltungsprinzipien, den Heuristiken, untersucht werden.

Die Idee hinter dieser Technik ist, dass eine bestimmte Anzahl von Evaluatoren, also Experten, unabhängig voneinander das System testen, um dabei auf mögliche benutzerunfreundliche Probleme zu stoßen. Diese Heuristiken geben den Evaluatoren vor, auf was sie bei der Auswertung achten sollen. Dazu haben Jakob Nielsen und Rolf Molich 1990 zehn Heuristiken entwickelt, die Nielsen 1994 noch einmal verfeinert hat.

1. ***Sichtbarkeit des Systemstatus:*** Konkret bedeutet dies, dass der Benutzer immer wissen soll, wo er sich gerade befindet und wo er von diesem Punkt aus hingelangen kann.
2. ***Das System muss sich an der Realität orientieren und sich dementsprechend anpassen:*** Dies bedeutet vor allem, dass die Schnittstelle in der Sprache des Benutzers sein muss.
3. ***Benutzerfreiheit und Kontrolle:*** Der Benutzer soll das System kontrollieren und dort frei agieren können. Wichtig hierbei ist, dass der Benutzer einen oder mehrere Schritte problemlos vor- oder zurückgehen kann.
4. ***Konsistenz und Standards:*** Das bedeutet in Klartext, dass bekannte Dinge nicht in einem anderen Kontext verwendet werden, um den Benutzer nicht zu irritieren. Es soll sich an Konventionen und Standards gehalten werden.
5. ***Fehler-Vermeidung:*** Heißt, es soll besonders erschwert werden, einen Fehler zu machen. Besser als gute Fehlermeldungen ist ein System, in dem es erst gar nicht zu solchen Fehlern kommt.

6. **Wiedererkennen statt wiederaufrufen:** Dabei geht es im Prinzip um die Eindeutigkeit von Elementen. Der Benutzer sollte erkennen können, wofür die Funktion gedacht ist und sich nicht erst erinnern müssen.
7. **Flexibilität und Effizienz:** Erfahrenen Benutzern sollte die Möglichkeit gegeben werden, ihre Handlungen innerhalb des Systems zu verkürzen. Auf der anderen Seite, sollten diese Möglichkeiten den unerfahrenen Benutzer nicht irritieren, damit beide Benutzer gut damit zurechtkommen, auf ihre Weise.
8. **Ästhetisches und minimalistisches Design:** Genauer gesagt geht es hier um die Optik des Interfaces, in welcher Art und welchem Umfang Informationen dargestellt werden. Zu viele unnötige Informationen sollten beispielsweise vermieden werden.
9. **Hilfestellung bei der Fehlerbehandlung:** Fehlermeldungen sollten verständlich sein, das Problem genau erläutern und dem Benutzer eine konstruktive Problembeseitigung bieten.
10. **Hilfe und Dokumentation:** Die meisten Systeme sollten und müssen eine Anleitung und Unterstützung für ihre Benutzung bereitstellen.

(vgl. „Human Computer Interaction“, S. 325/326)

Am Ende werden anhand der Beobachtungen der Evaluatoren, alle auftauchenden Probleme gesammelt und herausgearbeitet, gegen welche Heuristiken verstoßen wurde.

3.3 Modellbasierte Evaluation

Bei einer dritten expertenbasierten Evaluationstechnik, der Modellbasierten Evaluation, wird auf existierende Modelle, wie z.B. „GOMS“ zurückgegriffen. Die Abkürzung „GOMS“ steht für Goals, Operators, Methods and Selection rules. Diese beschreiben Vorgänge, die ein Benutzer durchführen muss, um ein System zu bedienen. Beim sog. „Keystroke-level model“ wird beispielsweise der Zeitaufwand für eine Aufgabe mathematisch berechnet. Ganz im Sinne der Frage „Wie viele Tasten, muss der Benutzer drücken und wie viel Zeit braucht er dabei, um die Aufgabe zu lösen?“. Diese Techniken können in einem frühen Stadium der Entwicklung wertvolle Informationen liefern, haben jedoch den Nachteil, dass sie Vorgänge vereinfachen, um sie berechenbar zu machen und daher nicht sehr praxisnah sind.

3.4 Evaluation durch Literaturrecherche

Um Zeit- und Kostenaufwand der Evaluation zu verringern, gibt es die Möglichkeit durch Literaturrecherche auf Studien und Forschungsergebnisse zurückzugreifen, die bestimmte Problemstellungen bereits untersucht haben. Man sollte jedoch abwägen, ob dies die Durchführung einer erneuten Evaluation ersetzen kann oder doch eine erneute Evaluation notwendig ist.

4. Evaluation durch Benutzer

4.1 Grundsätzliches zu experimentellen Designs

Experimentelle Designs unterscheiden sich grundsätzlich durch zwei Arten von Experimenten: Laborexperimente und Feldexperimente.

Laborexperimente finden in diesem Zusammenhang in einem so genannten „Usability Labor“ statt (siehe Abb. 1).



Abb. 1 Usability Labor

Feldexperimente werden hingegen in der gewohnten Arbeitsumgebung des Benutzers durchgeführt und können etwa durch den Einsatz eines mobilen Usability Labors (beinhaltet Videokamera, Notebook mit Analysetools, etc.) unterstützt werden (siehe Abb. 2).



Abb. 2 Mobiles Usability Labor

Durch Laborbedingungen erlangt man zwar eine hohe interne Validität (Gültigkeit der Ergebnisse), möglicher Nachteil ist jedoch die ungewohnte Situation für den Benutzer, der sich möglicherweise nicht so verhält, wie in seiner gewohnten Umgebung. Das wiederum ist der Vorteil des Feldexperiments, bei dem der Benutzer sich in seiner gewohnten Umgebung befindet: es bietet eine hohe externe Validität.

Grundsätzlich ist bei experimentellen Designs auf eine gezielte Auswahl der Versuchsgruppe(n) zu achten.

Hat man eine entsprechende Versuchsgruppe, formuliert man eine detaillierte Beschreibung der Aufgabe für die Testpersonen bzw. die Benutzer.

Es folgt die Bestimmung einer Hypothese, die durch das Experiment überprüft werden soll (Bsp: „Die Druckfunktion ist schwer zugänglich.“). Durch beeinflussen der unabhängigen Variablen (die Größe, welche variiert und verändert werden kann) wird geprüft, ob die Hypothese sich bestätigen lässt.

4.2 Beobachtungstechniken

Beobachtungstechniken werden eingesetzt, um Benutzer im konkreten Umgang mit einem System (oder einem Prototyp) zu beobachten. Der Benutzer bekommt vom Evaluator mehrere Aufgaben, die er mit Hilfe des Systems lösen soll.

Durch Video-/Audioaufnahme, Interaktionsaufzeichnung (zeigt die Eingaben des Benutzers am Bildschirm) und den Einsatz von speziellen Analysetools ist es möglich eine ausführliche Analyse durchzuführen um herauszufinden, wo das System noch Schwachstellen im Hinblick auf die Gebrauchstauglichkeit⁶ hat. Um die Gedankengänge des Benutzers besser nachzuvollziehen, wird dieser in der Regel gebeten bei der Bearbeitung der Aufgaben laut zu denken („Think-Aloud-Methode“).

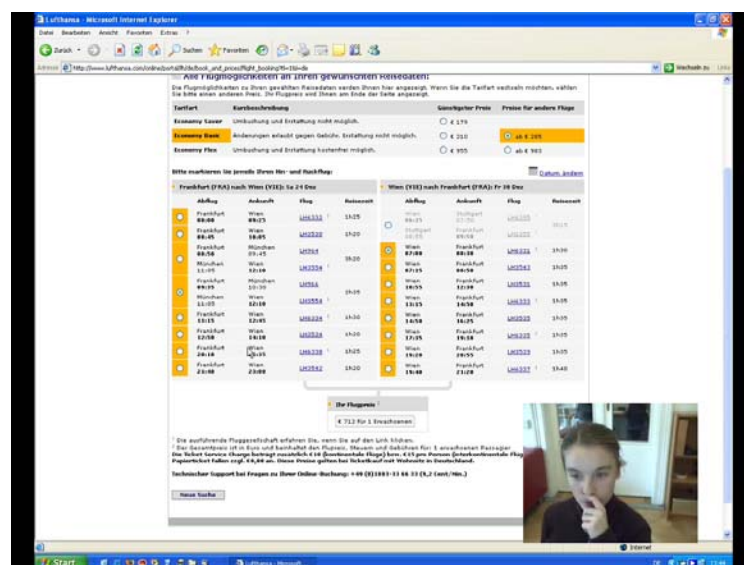


Abb. 3 Interaktions- und Videoaufzeichnung

Die Analyse und Auswertung einer Beobachtung kann durch die Aufzeichnungstechniken auch im Nachgang geschehen („post-task walkthrough“). Dadurch ergeben sich oft bemerkenswerte Zusammenhänge, die während des Experiments vom Evaluator nicht beachtet werden (besonders wenn man mehrere Evaluationstechniken in Verbindung miteinander anwendet).

4.3 Befragungstechniken

Im Allgemeinen dienen Befragungstechniken dazu, einen gewissen Teil an Problemen aufzudecken, die anhand von Beobachtungstechniken oder Techniken der Evaluation durch Experten nicht abgedeckt werden können.

Beim Interview erhält man gezielt Informationen über die Bedürfnisse, Eindrücke und Verhaltensweisen des Benutzers im Umgang mit dem System. Um Interviews mit unterschiedlichen Benutzern vergleichen zu können, sollten einige grundsätzliche Fragen formuliert werden, auf die das Interview aufbaut. Der Evaluator entscheidet dann, inwieweit er bei bestimmten Fragen nachhakt, um zusätzliche Informationen aus dem Interview herauszuziehen. Interviews eignen sich aus ökonomischer Sicht ebenfalls dazu, in Gruppen durchgeführt zu werden, um die Kosten zur Evaluation niedrig zu halten. Es gilt jedoch zu beachten, ob dies auch inhaltlich sinnvoll ist. (Wie setzt sich die Gruppe zusammen? Ist sie vergleichbar mit der Zielgruppe?)

Durch den Einsatz von Fragebögen lassen sich ebenso wertvolle Informationen sammeln. Das Fragebogendesign sollte jedoch gut durchdacht und die verschiedenen Fragemöglichkeiten (Fragen zur Person, Offene Fragen, Auswahlfragen, etc.) bewusst eingesetzt werden. Zudem ergibt sich der Vorteil, dass Fragebögen online beantwortet und automatisiert ausgewertet können.

4.4 Beobachten physiologischer Reaktionen

Eine weitere Technik der Evaluation durch den Benutzer ist das Beobachten physiologischer Reaktionen. Damit ist natürlich nicht gemeint, dass der Benutzer die physiologischen Reaktionen beobachtet, sondern dass die des Benutzers beobachtet werden.

Mit Hilfe von Methoden wie Eye-Tracking (zum Verfolgen der Augenbewegungen auf dem Bildschirm) (siehe Abb. 4) lassen sich beispielsweise exakte Daten über den Blickverlauf (siehe Abb. 5) eines Benutzers gewinnen. Anhand dieser Daten lassen sich auch so genannte „Hotspots“⁷ visualisieren, wodurch zentrale Bereiche der Aufmerksamkeit des Benutzers verdeutlicht werden.



Abb. 4 Augenbewegungen werden durch Sensoren erfasst



Abb. 5 Blickverlauf auf einer Webseite

Außer den Augenbewegungen lassen sich natürlich noch weitere physiologische Reaktionen eines Benutzers messen. An dieser Stelle nur ein paar Möglichkeiten:

- EEG – Elektroenzephalographie: mißt die elektrische Gehirnaktivität
- EMG – Elektromyographie: mißt elektrische Muskelaktivität

Durch die Kombination dieser Erkenntnisse mit anderen Techniken, ergeben sich komplexe Zusammenhänge, die Schwachstellen vorheriger Evaluationen aufzeigen können. Besonders im Bereich der Emotionalität des Benutzers.

Ein Beispiel für ein Programm, das verschiedene Techniken zusammenführt ist „The Observer XT“⁸. In Abb. 6 ist zu sehen, wie Beobachtungsnotizen, Herzschlag und Videoaufzeichnung zeitlich synchronisiert dargestellt werden. So lässt sich beispielsweise überprüfen, bei welcher Handlung die Testperson einen besonders hohen Herzschlag hatte.

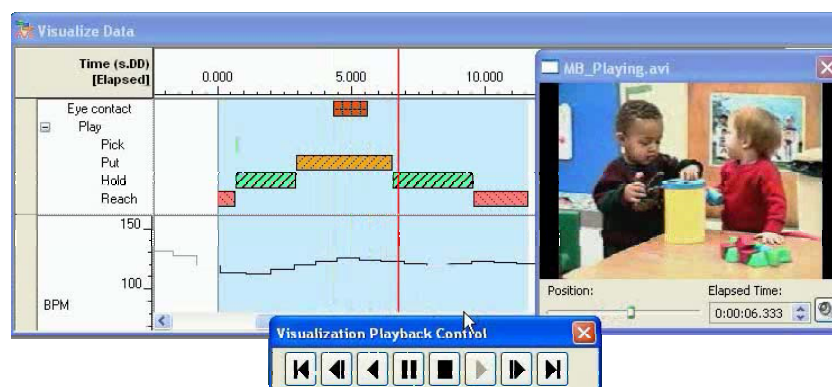


Abb. 6 Synchronisierung von Beobachtungsnotizen und Herzschlag in Verbindung mit Videobeobachtung

6. Zusammenfassung & Fazit

Während der Ausarbeitung ist deutlich geworden, dass Evaluation eigentlich zu jeder Entwicklungsphase eines Designprozesses stattfinden kann (auch davor). Im Rahmen eines solchen Designprozesses muss man sich im Klaren darüber sein, welche Form der Evaluation zu einem bestimmten Zeitpunkt am sinnvollsten ist. Zur Wahl stehen Techniken der Evaluation durch Experten oder den Benutzer. Sie alle haben ihre Vor- und Nachteile, tragen jedoch in jedem Falle zu einer Verbesserung des Systems bei. Durch die Kombination verschiedener Techniken lassen sich spezielle Probleme herausarbeiten, was aber zu hohen Kosten führen kann (besonders bei Evaluationstechniken durch Benutzer). Um zu einer geeigneten Wahl einer Evaluationstechnik zu gelangen muss man also besonders folgende Umstände beachten: Entwicklungszeitpunkt, Kostenaufwand, Evaluation durch Experten oder Benutzer und, ob sich gegebenenfalls Evaluationstechniken verbinden lassen.

7. Literaturangaben

¹ Alan Dix, Janet Finlay, Gregory Abowd, Russell Beale: Human Computer Interaction, 3. Auflage, Prentice Hall, 2004.

² DUDEN, Die Rechtschreibung, Band 1, 21., neu bearbeitete und erweiterte Auflage, Mannheim, 1996, S. 266

³ <http://de.wikipedia.org/wiki/Evaluation>

⁴ <http://eduinc.org/about/eval.html>

⁵ DUDEN, Die deutsche Rechtschreibung, Band 1, 21., neu bearbeitete und erweiterte Auflage, Mannheim, 1996, S. 347.

⁶ <http://www.fit-fuer-usability.de/1x1/basics/begriffszoo.html>

⁷ <http://www.uxmatters.com/MT/archives/000040.php>

⁸ <http://www.noldus.com>

Weitere Quellen:

<http://www.hcibook.com/e3/resources>

<http://www.evaluationstechniken.de>

<http://www.uxmatters.com>

<http://www.gc-upa.de/>

<http://www.useit.com/>

Abbildungen:

Abb. 1: <http://www.noldus.com>

Abb. 2: <http://www.noldus.com>

Abb. 3: <http://www.usability.at>

Abb. 4: <http://www.uxmatters.com>

Abb. 5: <http://www.uxmatters.com>

Abb. 6: <http://www.noldus.com>