

# Flashlight: Online Erfassung von Mausbewegungen

Annika S. Boldt<sup>1</sup>, Michael Schulte-Mecklenbeck<sup>2</sup>, Ryan O. Murphy<sup>3</sup>

Humboldt-Universität zu Berlin<sup>1</sup>, University of Bergen<sup>2</sup>, ETH Zürich<sup>3</sup>

## **Zusammenfassung**

Für die Usability-Forschung sind oftmals nicht nur die Zeiten interessant, die für die Bewältigung von Aufgaben benötigt wurden, sondern auch die Wege, auf denen diese Lösungen zustande kamen. Da die Erfassung von Blickbewegungen jedoch aufwendig und teuer ist, wird oft stattdessen die Bewegung der Maus registriert. Hier wird mit der Covering-Methode Flashlight der Umgang mit einer Suchmaske einer Versuchspersonendatenbank untersucht, bei der der Grad der Komplexität experimentell variiert wurde. Flashlight eignet sich sowohl zur Testung als auch Generierung von Hypothesen.

## 1 Einleitung

Die Interaktion mit Software ist ein dynamischer Prozess, daher reicht es in der Usability-Forschung oftmals nicht aus, nur die zur Lösung von Aufgaben benötigte Zeit zu betrachten, sondern auch die Wege, auf denen diese zustande gekommen sind. Für die Erfassung solcher Prozesse eignet sich vor allem die Blickbewegungsmessung. Diese ist jedoch sehr teuer und zeitaufwendig, weshalb im hier vorgestellten Ansatz nur der Anspruch erhoben wird, außerhalb des Labors und online über das Internet Daten zu erfassen.

In den letzten Jahren wurde die Bewegung der Maus zu einer wichtigen abhängigen Variable. Problematisch ist jedoch, dass sich Blick- und Mausbewegungen nicht eins zu eins aufeinander abbilden lassen. Eine Lösung bieten die so genannten Covering-Methoden mit der Grundidee der ortsbasierten Aufmerksamkeit (Posner et al. 1980). Dort werden die Teile der Seite verdeckt, die sich aktuell nicht im Fokus der Blick- bzw. Mausposition befinden. Beispiele sind Mouselab (Willemsen & Johnson 2006), der Restricted Focus Viewer (Bednarik & Tukiainen 2005), die Lorem-Ipsum-Covering-Methode (Kain et al. 2006) und das 2007 entwickelte Flashlight.

## 2 Flashlight

Bei dieser Methode wird angenommen, dass Maus-Bewegungen üblicherweise mit den Blickbewegungen übereinstimmen und Orte von Interesse kennzeichnen. Diese werden zuvor als „Areas of Interest“ (AIO) definiert.



Abbildung 1: Beispiel für Flashlight: links der unverzerrte Stimulus und rechts wie ihn der Teilnehmer sieht

Wie aus Abbildung 1 hervorgeht, ist jeweils nur der Teil des Screenshots sichtbar, über dem sich die Maus befindet. Der restliche Teil wird graduell bis zur Unleserlichkeit verzerrt. Flashlight erhebt mit einer Auflösung von etwa 10 Hz die Anzahl und Dauer der Maus-Fixationen pro AOI. Außerdem können Übergangswahrscheinlichkeiten und Heatmaps<sup>1</sup> errechnet werden. Flashlight basiert auf HTML, Javascript, PHP und MySQL. Voraussetzungen sind daher grundlegende Programmierkenntnisse und Webspace. Die Versuchsperson benötigt einen Javascript-fähigen Browser. Das Open Source-Programm und MATLAB-Auswertungsscripts stehen unter <http://taschenlampe.sourceforge.net> kostenlos zur Verfügung.

Flashlight wurde bereits mit Blickbewegungsmessung verglichen und eine gute Übereinstimmung beider Methoden festgestellt, sofern die Stimuli maximal acht AOIs enthielten. Flashlight ist dabei in der Analyse von Entscheidungsprozessen eingesetzt worden, hier soll nun eine mögliche Eignung für die Usability-Domäne getestet werden.

## 3 Methoden

Untersuchungsobjekt der Evaluation ist die Online-Versuchspersonendatenbank PESA<sup>2</sup>, die seit 2007 an der Humboldt-Universität zu Berlin genutzt wird. Ein Workshop zu PESA im Frühjahr 2008 offenbarte zahlreiche Mängel, zum Teil im Usability-Bereich. Daher wurde die Suchmaske für Versuchsleiter als Evaluationsobjekt ausgewählt. Probleme hierbei sind

<sup>1</sup> Visualisierungen von häufig betrachteten Elementen auf dem Bildschirm

<sup>2</sup> Psychologischer Experimental-Server Adlershof: <http://web.psychologie.hu-berlin.de/pesa>

die Unübersichtlichkeit und die Verwendung von Mengenoperatoren (<, >) statt sprachlicher Beschreibungen. Es wurde als Gegenentwurf eine weniger komplexe Maske erstellt.

Die Screenshots wurden mit unterschiedlichen Daten erstellt, um Lerneffekte auszuschließen, und mit einem Gauss-Filter verzerrt. Das Online-Experiment bestand nun aus einem Fragebogen zu demografischen Daten und 36 Screenshots. Die Teilnehmer wurden dabei gebeten, Fragen zur Maske zu beantworten. Sie sollten dazu mit der Maus die entscheidende Information suchen und beantworten, ob diese dort enthalten war oder nicht.

Verglichen wurde der Umgang mit der alten und der neuen Maske für jede Versuchsperson. Die abhängigen Variablen waren Verweilzeiten, Fehlerwerte, Anzahl der Maus-Fixationen und Anteil korrekter Fixationen. Wir nahmen bei der optimierten Suchmaske eine bessere Performanz in Bezug auf alle Variablen an. Außerdem erwarteten wir, dass die mit PESA bereits erfahrenen Probanden die Aufgaben besser und schneller lösen würden.

## 4 Ergebnisse

Die 27 Teilnehmer waren im Mittel 26,6 Jahre alt. 17 waren weiblich und 13 hatten bereits mit der Maske gearbeitet. Die Auswertung der Daten geschah über oben genannte MATLAB-Skripts. Dazu wurden 5 AOIs definiert. Anzahl der Fixationen<sup>3</sup> pro AOI wurden berechnet.

Die Hypothesen wurden mit Varianzanalysen getestet. Die Wechselwirkungseffekte wurden dabei nicht signifikant ( $p = .379$  bis  $.930$ ). Wir berichten hier nur die Ergebnisse der post hoc durchgeführten t-Tests.

Die Analysen zeigten, dass wie erwartet die Probanden weniger Zeit benötigten, um die Aufgabe mit der optimierten Maske zu lösen (7,1 vs. 8,7 Sekunden). Dieser Unterschied war signifikant ( $p = .027$ ). Auch in den Fehlerraten lag dieser Effekt in der angenommenen Richtung vor (8,6 % bei der neuen Maske und 11,4 % bei der alten), allerdings wurde dieser Unterschied nicht signifikant ( $p = .204$ ).

Wir erwarteten zudem eine geringere Anzahl an Maus-Fixationen für die optimierte Maske, da diese weniger komplex war. Tatsächlich fixierten die Probanden diese im Mittel 7,9 mal und die alte Maske 9,2 mal. Dieser Unterschied erweist sich im t-Test bei zweiseitiger Testung jedoch nur als tendentiell signifikant ( $p = .081$ ).

Innerhalb dieser Maus-Fixationen wurde die Rate der Fixationen errechnet, die sich auf die AOIs beziehen, also die für die Aufgabenbewältigung relevanten Teile. Ein Mittelwertsunterschied zeigte sich ebenso: Während 46,5 % der Fixationen in der alten Maske auf die relevanten Bereiche gerichtet waren, waren es bei der optimierten Fassung 56,6 %. Dieser Unterschied ist statistisch signifikant ( $p < .001$ ).

---

<sup>3</sup> Als Fixation galt es, wenn sich die Maus in 300 ms nicht mehr als 20 Pixel bewegte.

Weitere Analysen zeigten, dass die erfahrenen PESA-User auf diesen Variablen ebenso bessere Ausprägungen besaßen. Diese Unterschiede wurden für alle Vergleiche außer den Fehlerwerten signifikant ( $p = .002$  bis  $.036$ ). Außerdem ermöglicht Flashlight neben der Hypothesentestung auch eine Analyse von Ausreißern. Mit Grafiken, die das genaue Fixationsmuster der Probanden abbilden, lassen sich diese gut analysieren.

## 5 Diskussion

Die Ergebnisse des Experiments zeigen, dass Flashlight sensitiv für den Unterschied zwischen den Masken ist. Ein Mehrwert im Vergleich zur bloßen Analyse von Fehlerzahlen besteht darin, dass gezeigt werden kann, wie sich der Anteil irrelevanter Maus-Fixationen zugunsten relevanter verschiebt. Anzumerken sei jedoch, dass die hier verwendeten Screenshots statisch sind. Eine mögliche Erweiterung für interaktive Seiten wäre wünschenswert. Auch sind Störvariablen bei Online-Testungen wesentlich schwerer zu kontrollieren, Hinweise hierzu gibt es bei Reips (2002).

Ein Vorteil von Flashlight ist jedoch, dass es günstig und flexibel zu handhaben ist. Auch kann eine Analyse von Ausreißern helfen, Hypothesen für die zukünftige Forschung und Anregungen für eine Verbesserung des Interfaces zu generieren. Abschließend ist also zu sagen, dass Flashlight eine gute und günstige Alternative zur konventionellen Blickbewegungsmessung darstellt. Für die Zukunft wäre ein direkter Vergleich mit konventioneller Blickbewegungsmessung angewandt in einer Usability-Testung wünschenswert.

### Literaturverzeichnis

- Bednarik, R. & Tukiainen, M. (2005). Effects of Display Blurring on the Behavior of Novices and Experts during Program Debugging. In Proceedings of CHI'05: *Extended abstracts of the ACM Conference on Human Factors in Computing Systems*. Portland: ACM Press. (S. 1204–1207).
- Kain, S., Polkehn, K., Thüning, M. & Wandke, H. (2006). Lorem-Ipsum Covering as a method to measure attention allocation on websites - An alternative to Site-Covering? Poster presented at the *General Online Research 2006 (G.O.R. 06)*, Bielefeld.
- Posner, M.I., Snyder, C.R.R. & Davidson, B.J. (1980). Attention and the detection of signals. *Journal of Experimental Psychology: General*, 109, 160-174.
- Reips, U. D. (2002). Internet-Based Psychological Experimenting: Five Dos and five Don'ts. *Social Science Computer Review*, 20(3), 241-249.
- Willemsen, M. C. & Johnson, E. J. (2006). *MouselabWEB: Monitoring information acquisition processes on the Web*. Retrieved March 23, 2009, from <http://www.mouselabweb.org/>

### Kontaktinformationen

Annika S. Boldt: [boldtann@hu-berlin.de](mailto:boldtann@hu-berlin.de)