

Blickgesteuerte Interaktion bei der visuellen Analyse komplexer Szenen zur Entlastung des Arbeitsgedächtnisses

Tamara Granacher

Schlüsselwörter: visuelle Klassifikation, Blicksteuerung, Szenenanalyse, Arbeitsgedächtnis, Blickmessung

Zusammenfassung

Bei der bildbasierten Szenenanalyse am Rechnerarbeitsplatz wird durch Verwendung von Blickmessung gegenüber herkömmlichen Interaktionstechniken eine Leistungssteigerung erwartet. Es wurden die zwei blickgesteuerten Interaktionstechniken *reine Fixation (RFix)* und *Fixation mit Bestätigung (FixMB)* und die konventionelle Interaktionstechnik mittels *Maus* mit der Human Processor Modelling Language (HPML) modelliert. Anschließend wurde anhand dieser Modelle der Indikator für die Belastung berechnet und mit der Leistung in Beziehung gesetzt. Hieraus wurde die Leistungsrangfolge *RFix* vor *FixMB* vor *Maus* vorhergesagt, umgekehrt zur Belastungsrangfolge der Interaktionstechniken. Die theoretische Vorhersage wurde durch eine anschließende, empirische Validierung bestätigt. Mit den belastungsärmeren Interaktionstechniken mittels Blickbewegungsmessung lässt sich bei der Szenenanalyse eine bessere Leistung erzielen als mit der Maus, obwohl die Versuchspersonen den Umgang mit dem Blickmessgerät nicht gewohnt und deshalb nicht genauso gut trainiert waren wie im Umgang mit der Maus. Die Gedächtnisfehleranzahl während einer Klassifikationsaufgabe war bei Benutzung der Maus merklich höher als bei der Benutzung einer blickgesteuerten Interaktionstechnik.

The performance of image-based scene analysis at computer workstations is expected to be improved by usage of eyetracking instead of conventional pointing techniques. Two interaction techniques controlled by eye movements: *pure fixation (RFix)* and *fixation with confirmation (FixMB)* and the conventional technique using a *computer mouse (Maus)* were modelled by the Human Processor Modelling Language (HPML). Subsequently, based on these models an indicator of workload was calculated and related to performance. The predicted ranking relating to performance was *RFix* before *FixMB* before *Maus*, which is reverse to the ranking relating to the workload of the interaction techniques. The theoretical prediction was confirmed by a subsequent empirical validation. With the usage of the less straining interaction technique via eye tracking, it is possible to achieve a higher performance in scene analysis than by using the computer mouse, although the probands were not accustomed to the eye-tracker and therefore were less well trained in usage of the eye-tracker than in the usage of the mouse. The number of memory errors during a classification task was much higher using the mouse than using interaction techniques controlled by eye movements.

Einleitung

Bei der bildbasierten Szenenanalyse, z.B. in der Erdbeobachtung mittels Satelliten, wird aufgrund des Vorhandenseins von bestimmten Objekten und deren örtlicher Relation zueinander die Szene klassifiziert. Aufgrund der hohen Komplexität der zur Verfügung stehenden Abbildungen der Szene ist es meist nicht möglich, sie auf einen Blick zu erfassen. Daher müssen mit gezielten Einzelbeobachtungen Detailausschnitte nacheinander analysiert und dabei evtl. einer Bildverbesserung unterzogen werden.

Der Szenenanalytiker verbessert in jedem Beobachtungsschritt seine Hypothese über die Szenenklasse, die er als mentales Modell in seinem Gedächtnis verwaltet, wobei das Arbeits-

gedächtnis die gerade bewussten Anteile speichert. Die aktuelle Klassenhypothese beeinflusst wiederum die Auswahl der als nächstes zu betrachtenden Bildregion (Region of Interest: ROI). Für diese Auswahl ist eine Interaktion mit dem Computer nötig. Die Interaktion belastet das Arbeitsgedächtnis zusätzlich zur mentalen Aufgabe der Szenenanalyse und kann, wenn die Interaktionstechnik ungeschickt gewählt ist, zu einer nennenswerten Leistungseinschränkung bei der Szenenanalyse führen. Je größer die Komplexität der Interaktion ist, desto größer wird die Wahrscheinlichkeit dafür, dass für die Szenenanalyse notwendige Merkeinheiten (Chunks) aus dem Arbeitsgedächtnis verdrängt, also vergessen werden. Dieser Umstand schlägt sich negativ in der Leistung nieder, da die vergessenen Chunks entweder neu geladen werden müssen, was Zeit kostet, oder aufgrund des Vergessens Fehler gemacht werden.

Die natürliche, visuelle Erkundung einer Abbildung geschieht mittels Blickbewegung. Deshalb liegt es nahe, den Blick anstelle des herkömmlichen Zeigens mit der Maus zur Unterstützung der Interaktion mit dem Bilddarstellungsprogramm heranzuziehen. Um zu untersuchen, ob dadurch eine Leistungssteigerung in der Szenenanalyse erreichbar ist, wurden zwei blickgesteuerte Interaktionstechniken für die ROI-Auswahl auf ihre Belastung des Arbeitsgedächtnisses hin der Interaktion mit der Computermaus gegenübergestellt:

Bei Interaktionstechnik *Reine Fixation* (kurz: *RFix*) werden die Bildausschnitte mittels genügend langer Fixation angefordert. Durch diesen Schwellenwert wird sichergestellt, dass keine sofortige, evtl. ungewollte Auslösung einer Aktion mit dem Blick erfolgt. Bei der zweiten Interaktionstechnik, hier *Fixation mit Bestätigung* (kurz: *FixMB*) genannt, muss die Auswahl eines Ausschnittes nach einer Vorselektion durch den Blick mit einem Tastendruck bestätigt werden. Dadurch ist ein Umherschweifen mit dem Blick möglich, ohne direkt eine Reaktion auszulösen. Dass ein Ausschnitt ausgewählt ist, wird mit einer Markierung desselben dargestellt.

Modellierung und theoretische Leistungsvorhersage

Ausgehend vom 3-Prozessor-3-Speicher-Modell nach Card et al. (1983) können mit der Human Processor Modelling Language HPML nach Geisler (2006) Taktfolgen der drei Prozessoren veranschaulicht und so eine quantitative Vorhersage über die Verdrängung von Chunks aus dem Arbeitsgedächtnis erreicht werden. Abb.1 zeigt die HPML-Modelle zu den drei untersuchten Interaktionstechniken.

Im Modell *RFix* entspricht Zyklus 1 ohne die Knoten B und C dem natürlichen, visuellen Klassifikationszyklus, der auftritt, wenn eine Szene ohne Interaktion klassifiziert werden kann. Zyklus 2 überprüft, ob sich der Blick an der richtigen Position befindet. Zyklus 3 dient zur indirekten Bestätigung für den Benutzer, dass seine Eingabe verstanden wurde, also die ROI verbessert angezeigt wird.

Bei *FixMB* ist zusätzlich der Tastendruck für die Bestätigung modelliert. Im Modell existieren zusätzlich zum Klassifikationszyklus 1 noch Zyklus 2, der prüft, ob der Blick auf der gewünschten Position ruht und Zyklus 3, der prüft, ob die ROI bereits ausgewählt wurde und somit für eine Bestätigung bereit ist. Die Markierung einer ROI dient dem Benutzer als direkte Bestätigung dafür, dass seine Eingabe angenommen wurde. Deshalb ist ein Zyklus analog zu Zyklus 3 des Modells *RFix*, der eine indirekte Bestätigung modelliert, nicht mehr nötig.

Beim Modell *Maus* führt der zusätzliche Regelkreis (2) zwischen der Maushand und dem Auge zur Steuerung der Zeigerposition zu einer deutlichen Vergrößerung des Modells. Zyklus 1 entspricht dem Klassifikationszyklus.

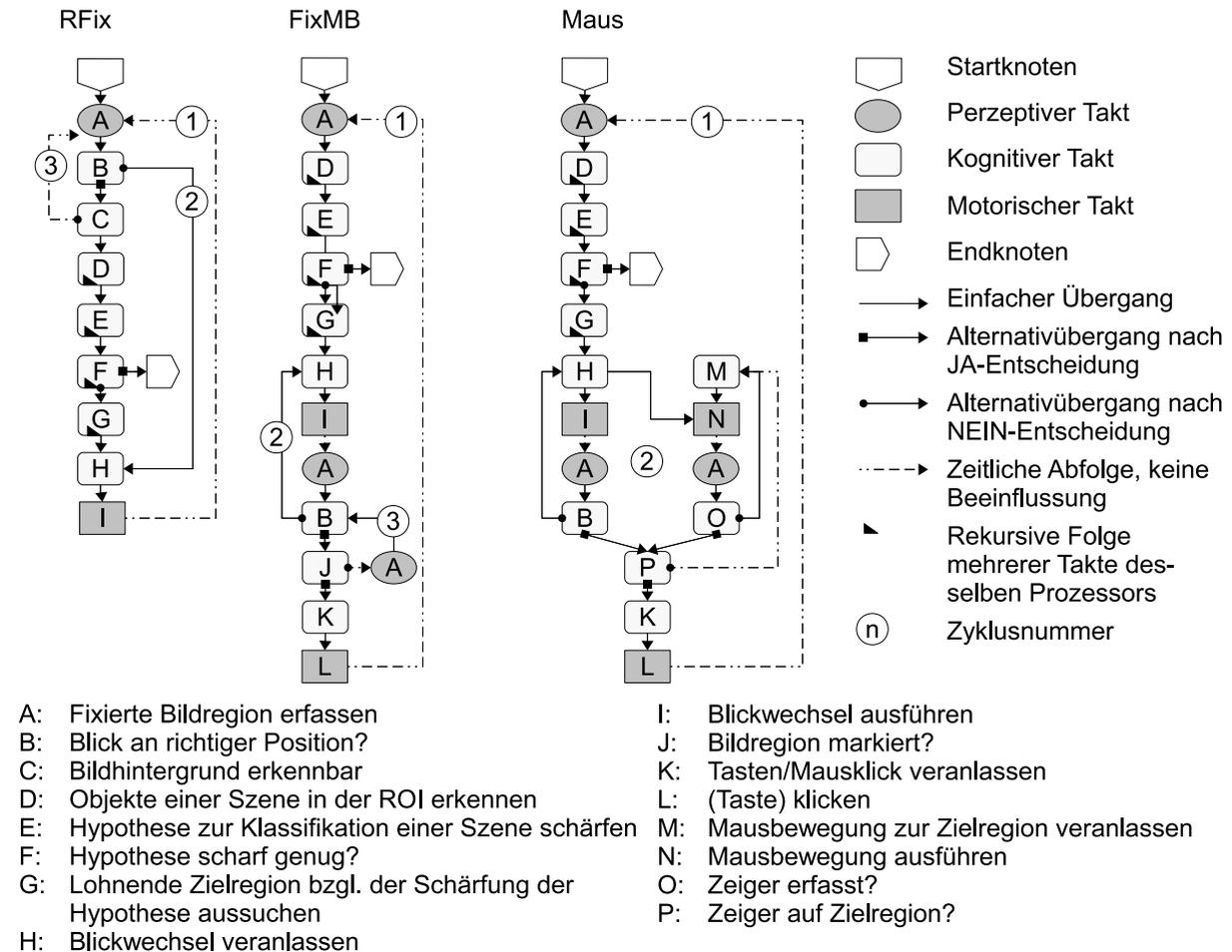


Abb.6: HPML-Modelle der drei untersuchten Interaktionstechniken

Um die unterschiedlichen Interaktionstechniken anhand ihrer HPML-Modelle zu vergleichen, wird als indirektes Belastungsmaß der von Geisler (2006) vorgestellte, indizierte Belastungsindikator B^* verwendet:

$$B^*(\Theta) = z + \sum_{i=1}^z S(i)$$

Dabei sei Θ die Interaktionstechnik, i der Zähler über die Prozessorknoten im Modell, z die Anzahl der Prozessorknoten und $S(i)$ die Anzahl der Schleifen (auch Rekursionen), in die der i -te Prozessorknoten eingebunden ist. Grundgewicht z stellt sicher, dass jeder Prozessorknoten mindestens einmal berücksichtigt wird. Einen Algorithmus zur Berechnung des Belastungsindikators B^* anhand des zugehörigen HPML-Modells stellen Geisler & Scheben (2007) vor.

Da die Leistung von Belastung und Beanspruchung abhängt, ist der Zusammenhang zwischen B^* und der Leistung nicht explizit bestimmbar. Deshalb erfolgt die theoretische Vorhersage der Leistungsunterschiede bei den verschiedenen Interaktionstechniken ordinal.

Je komplexer der Graph für eine Interaktionstechnik ist, desto mehr Zeit muss für die Interaktion aufgewendet werden und desto eher werden für die Szenenanalyse wichtige Chunks von den für die Interaktion notwendigen verdrängt. Beides wirkt sich negativ auf die Leistung aus. Die Rangfolge der Interaktionstechniken bei der Belastung entspricht somit der umgekehrten Rangfolge bei der Leistung. Da beide Funktionen monotone Funktionen sind, ist sichergestellt, dass eine Rangfolge bestimmbar ist. Die Leistung hängt mit der Beanspruchung auch von den individuellen Fähigkeiten des einzelnen Menschen ab. Die dadurch hervorgerufene Streuung

der Leistung in den zur Validierung durchgeführten Versuchen wird in Kauf genommen, da ein ausführliches Training zur Angleichung der Fähigkeiten vor den Versuchen nicht möglich war.

Der Belastungsindikator bei *RFix* hat mit $B^* = 29$ den niedrigsten Wert. Das erfüllt die Erwartungen, da diese Interaktionstechnik in Anlehnung an den visuellen Zyklus ohne zusätzlich benötigte Interaktion entworfen wurde. Es folgt das Modell *FixMB* mit einem Belastungsindikator von 36 und das *Maus*-Modell mit einem Wert von $B^* = 66$. Es ergibt sich also in Bezug auf die Leistung die Rangfolge *RFix* vor *FixMB* vor *Maus*.

Um die theoretisch vorhergesagte Rangfolge empirisch zu validieren, war ein allgemeines Beispiel nötig, das für alle Versuchsteilnehmer verständlich war und das wie in der Realität eine Grobstruktur aufweist, anhand der man auf charakteristische Teilobjekte schließen kann. Als Versuchsbeispiel wurden Fotos von einem verschiedenen eingedeckten Tisch verwendet.



Abb.7: links: Auswahl der vorgelegten Bilder, rechts: aufgedeckte ROI

In Abb.2 sind 4 der 15 Gedecke exemplarisch abgebildet, die aufgrund des Vorhandenseins von Gedeckteilen und deren örtlicher Relation zueinander nach dem bevorstehenden Mahl klassifiziert werden sollten. Dies ist möglich, weil die Anordnung der Komponenten des Tischgedecks fest geregelt ist. Beispielsweise liegt das Messer bei einem Hauptgang stets rechts neben dem Teller, die Gabel stets links neben dem Teller.

Bei dieser relativ einfachen Szene wäre eine Klassifikation auf einen Blick möglich. Bei komplexen Szenen muss die Szene mit Einzelbeobachtungen erkundet werden. Um dieses Verhalten zu erzwingen, wird das Bild durch eine graue Fläche verdeckt. Es darf jeweils nur das Gebiet (ROI) aufgedeckt werden, das gerade fokussiert wird (siehe Abb.2, rechts), so dass die Chunks zu den bereits gesehenen Objekten im Gedächtnis gespeichert werden müssen.

Empirische Validierung

Die Versuche wurden mit 20 Personen durchgeführt. Die vier weiblichen und 16 männlichen Versuchspersonen hatten ein Alter von 18 bis 45 Jahren. Alle Personen waren geübt im Umgang mit Computern und nahmen freiwillig an den Versuchen teil. Aus technischen Gründen konnten nur Personen ohne Brille an den Versuchen teilnehmen. Jeder Proband nahm jeweils an einer Versuchsreihe (6 Einzelversuche) mit einer blickgesteuerten Interaktionstechnik (*RFix* bzw. *FixMB*) und an einer Versuchsreihe mit der Interaktionstechnik *Maus* teil. Um eventuell auftretende Lerneffekte gleichmäßig auf die Interaktionstechniken zu verteilen, absolvierte eine Hälfte zuerst die Versuchsreihe *Maus*, die andere die Reihen *RFix* bzw. *FixMB*. Jede Versuchsperson hatte pro Einzelversuch 20 Sekunden Zeit, eine beliebige Anzahl ROIs der Reihe nach aufzudecken und ihre Klassenhypothese zu schärfen. War die Versuchsperson sich anschließend noch nicht sicher, hatte sie noch beliebig viel Zeit nachzudenken, bevor sie ihre Antwort gab. Die Zeitmessung wurde von den Probanden durch Tastendruck vor dem Öffnen der ersten ROI gestartet und vor Bekanntgabe der Antwort gestoppt. Bei jedem Einzelversuch wurden die benötigte Zeit für die Klassifikation und jede aufgedeckte ROI dokumentiert. Mindestens einen Tag vor dem Versuchstermin wurde ein Begrüßungsbogen verteilt, auf dem die Aufgabe, der Ablauf der Versuche und Grundsätzliches zur Blickmessung erklärt wurden. Außerdem enthielt der Bogen alle zu erkennenden Bilder mit der Aufforderung, diese auswendig zu lernen. Für eine detaillierte Aufstellung des Versuchsaufbaus siehe Granacher (2009).

Während der Versuchsbearbeitung können drei Arten von Gedächtnisfehlern auftreten:

Bei Fehlerart A) wird vergessen, nach dem Objekt der Szene in einer ROI überhaupt zu suchen. Da die Szenen vor dem Versuchsbeginn auswendig gelernt wurden und somit ins Langzeitgedächtnis übergegangen sein sollten, ist dieser Fehler jedoch nicht dem Arbeitsgedächtnis zuzuschreiben. Weil die getroffene Klassifikation durch Zufall trotzdem richtig sein kann, fällt erst bei einer nachträglichen Analyse der während des Versuchs betrachteten ROI auf, dass ein oder mehrere Objekte, die für eine korrekt abgeleitete Klassifikation notwendig sind, nicht überprüft wurden.

Bei Fehlerart B) muss ein Objekt der Szene in einer ROI nach einiger Zeit wiederholt angesehen werden, weil vergessen wurde, was dort schon gesehen wurde oder dass das Objekt vorher bereits angesehen wurde. Dieser Fehler ist dem Arbeitsgedächtnis zuzuordnen, weil die Objekte, die schon erkannt wurden, als Chunks im Arbeitsgedächtnis vorliegen. Werden diese Chunks durch andere Chunks verdrängt, so müssen sie durch wiederholtes Nachsehen wieder ins Arbeitsgedächtnis geladen werden. Der Fehler ist erst bei einer nachträglichen Analyse der gewonnenen Versuchsaufzeichnungen erkennbar, denn die Klassifikation ist aufgrund der Vergewisserung meist korrekt.

Bei Fehlerart C) wird vergessen, dass ein Objekt der Szene in einer ROI gesehen wurde. Im Gegensatz zu Fehlerart B) wird hier aber nicht bemerkt, dass etwas vergessen wurde, und zur Vergewisserung nicht noch einmal nachgesehen. Der Chunk zum vergessenen Objekt der Szene wird durch andere Chunks verdrängt. Deshalb wird der Fehler dem Arbeitsgedächtnis zugeschrieben. Er äußert sich darin, dass aus den gewonnenen Versuchsaufzeichnungen zwar hervorgeht, dass die Versuchsperson ein Objekt gesehen hat, am Ende jedoch eine falsche Klassifikation als Lösung nennt, die das Gesehene nicht berücksichtigt.

Abb.3 zeigt in Abhängigkeit der unterschiedlichen Interaktionstechniken die unterschiedlichen Arten und Häufigkeiten von Fehlern, die im Mittel pro Versuchsperson bei einer Versuchsreihe vorkamen. Bei Versuchsreihe *RFix* wurde im Mittel nur ein Fehler gemacht, während die Anzahl der Fehler bei *FixMB* im Mittel bei 1,7 Fehlern lag. Für die Interaktionstechnik *Maus* stieg die Anzahl der Fehler auf 4,2.

Die gemittelte Leistung der Versuchspersonen wurde nach richtig klassifizierten Szenen pro Versuchsreihe pro Zeit berechnet. Der Median der Leistung¹ bei der Versuchsreihe *RFix* hat mit 12,6 den besten Wert, gefolgt vom Median der Versuchsreihe *FixMB* mit einem Wert von 8,9. Versuchsreihe *Maus* folgt mit einem Median von 7,9. Die theoretische Vorhersage der Leistungsrankfolge entspricht somit der Rangfolge aus den Versuchen.

Die Beanspruchung ist aufgrund der sehr komplexen Beschreibung persönlicher Fähigkeiten und mentaler Zustände sehr schwer zu messen. Um eine qualitative Einschätzung der Beanspruchung zu ermöglichen, wurden die Versuchspersonen zu ihrem subjektiven Eindruck befragt. Insgesamt wurde an den blickgesteuerten Interaktionstechniken bemängelt, dass zu wenig Übung und Gewohnheit vorhanden war, vor allem weil sich erst im Laufe der Zeit ein Gefühl für den richtigen Umgang mit dem Blickmessgerät entwickelt. Die Interaktionstechnik *Maus* wurde oft als gewohnter und vertraut bezeichnet, andererseits gab es auch Stimmen, denen die zusätzliche Interaktion mittels Hand anstrengend und ablenkend erschien. Die Interaktionstechnik *RFix* beschrieben viele Teilnehmer als intuitiv, bequem und angenehm. Die Möglichkeit, bei der blickbasierten Interaktion mit dem Blick im Gegensatz zur Maus zu springen und die natürliche Herangehensweise wurden als leicht und unkompliziert empfunden.

¹Um besser lesbare Werte zu erhalten, ist die Leistung jeweils mit 100 multipliziert worden.

Ergebnis

Die Gedächtnisfehlerhäufigkeit stieg bei zunehmender Belastung. Die Rangfolge der Leistung mit den drei Interaktionstechniken entspricht der theoretischen Vorhersage. Mit Blickbewegung als Interaktionstechnik war eine höhere Leistung als mit der Maus möglich, obwohl es sich noch um eine ungewohnte Technik handelt (siehe dazu Abb.3). Aufgrund der Versuche hat sich ebenfalls gezeigt, dass die Ablaufbeschreibungssprache HPML nicht nur, wie von Geisler (2006) beschrieben, zur Modellierung visueller Detektion, sondern auch der Klassifikation verwendet werden kann. Detailliertere Ergebnisse finden sich in Granacher (2009).

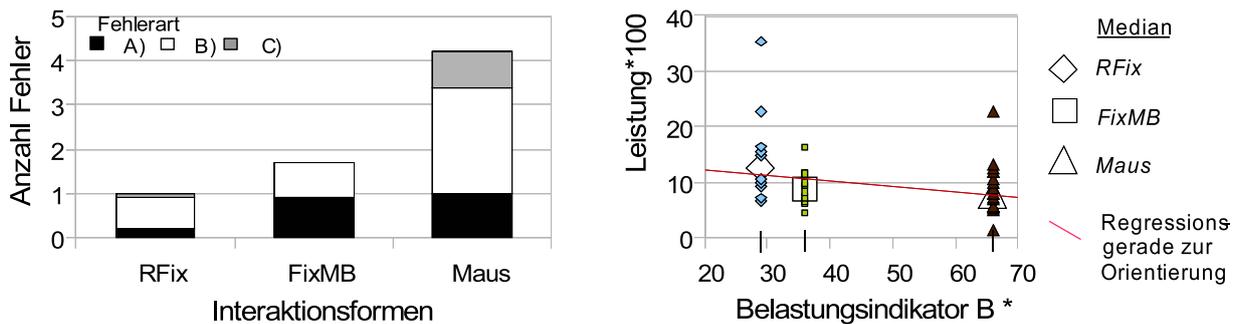


Abb.8: links: Mittlere Anzahl der Fehler in Abhängigkeit der unterschiedlichen Interaktionsformen pro Versuchsperson, rechts: experimentell ermittelte Leistungen der Versuchspersonen.

Danksagung

Die dargelegten Forschungsergebnisse entstanden im Rahmen meiner Diplomarbeit an der Fakultät für Informatik der Universität Karlsruhe (TH), die freundlicherweise vom Fraunhofer-Institut für Informations- und Datenverarbeitung IITB betreut wurde. Mein Dank gilt Herrn Prof. Dr. Syrbe und Herrn Dr. Geisler für die freundliche und engagierte Betreuung der Arbeit.

Literatur

- Card, S. K., Moran, T. P. & Newell, A. (1983). *The psychology of human-computer interaction*. Erlbaum
- Geisler, J. (2006). *Leistung des Menschen am Bildschirmarbeitsplatz. Das Kurzzeitgedächtnis als Schranke menschlicher Belastbarkeit in der Konkurrenz von Arbeitsaufgabe und Systembedienung*. Karlsruhe: Universitätsverlag Karlsruhe.
- Geisler, J., Scheben, C. (2007): Human processor modelling language (HPML): estimate working memory load through interaction. In: Min, Daihwan (Ed.): 10th IFAC/IFIP/IFORS/IEA Symposium on Analysis, Design, and Evaluation of Human-Machine Systems (IFAC-HMS 2007), Seoul, Korea 4.-6.9.2007.
- Granacher, T. R. (2009). Blickmessung als Interaktionstechnik für die Beobachtungssteuerung bei der visuellen Analyse komplexer Szenen im Hinblick auf die Belastung des menschlichen Arbeitsgedächtnisses. Diplomarbeit, Universität Karlsruhe, 2009 Verfügbar über: <http://publica.fraunhofer.de>