

Performance – (k)ein Thema für Usability Professionals?

Dr. Herbert A. Meyer
artop - Institut an der Humboldt-
Universität zu Berlin
Christburger Str. 4
10405 Berlin
meyer@artop.de
www.artop.de

Petra Vogt
Gomez Deutschland AG
Flughafenstr. 52a
22335 Hamburg
petra.vogt@gomez.com
www.gomez.com

Marco Glier
Cuxhavener Straße 579
21149 Hamburg
glier@gmx.net

Abstract

Das durch das Zusammenwirken von Hard- und Software bedingte zeitliche Verhalten eines Computersystems (Performance) kann Effektivität, Effizienz und Zufriedenstellung beim Umgang mit interaktiven Systemen beeinflussen. Es wird gezeigt, wie Performance als »Usability Requirement« in die software-technische und software-ergonomische Bewertung und Gestaltung von

Softwareprodukten einfließen kann. Zur Begründung der These, Performance sei eine Grundlage jeder Usability, wird ein kurzer Einblick in den Stand der Forschung gegeben. Dabei werden Befunde und Mythen zur Auswirkung von Performance auf Benutzerverhalten sowie Empfehlungen zur Gestaltung von Systemantwortzeiten dargestellt.

Keywords

Performance, Systemantwortzeit, Usability Requirement

1.0 Einleitung

Für die Benutzer zeigt sich die Performance eines interaktiven Systems durch das Auftreten von Verzögerungen zwischen Eingabe und Ausgabe (Antwortzeit, in der Benutzersicht Ladezeit genannt) und die Zeit, die zur Darstellung benötigt wird (Ausgabezeit, Bildschirmaufbau). Sie ist abhängig von einem komplexen Zusammenspiel mehrerer Hardware-Komponenten mit der Software sowie bei Netzwerkanwendungen auch noch der Netzlaufzeit.

Für die Gestaltung und Bewertung interaktiver Systeme ist die Berücksichtigung der Performance unabdingbar, da das zeitliche Verhalten eines Computersystems darauf Einfluss nimmt, inwiefern Benutzer ihre Ziele effektiv, effizient und zufriedenstellend erreichen. Der vorliegende Beitrag geht zunächst darauf ein, wie Performance traditionell aus software-technischer Sicht behandelt wird und wie sie beim Usability Engineering unter software-ergonomis-

chen Gesichtspunkten berücksichtigt werden kann. Danach geben wir einige Beispiele für die Auswirkung der Performance, sowie grundlegende Hinweise zur Gestaltung der Systemantwortzeit.

2.0 Performance als »Usability Requirement«

2.1 Performance aus software-technischer Sicht

In der Softwaretechnik wird Performance als Merkmal der Softwarequalität behandelt. Das Qualitätsmodell ISO 9126 beispielsweise bestimmt die »Effizienz« eines Softwareprodukts u.a. über das Teilmerkmal »Zeitverhalten«. Mit Zeitverhalten ist die Antwort- und Verarbeitungszeit sowie der Durchsatz bei der Funktionsausführung – und damit Performance – gemeint.

Performance gehört aus software-technischer Sicht zu den nicht-funktionalen Anforderungen, die den Rahmen festlegen, in dem die gewünschte Verarbeitungsfunktionalität umgesetzt werden muss. Es sollte also in einer Anforderungsanalyse spezifiziert werden, in welchem Bereich sich die Performance unter festgelegten Bedingungen bewegen sollte. Wenn die Randbedingungen (z.B. Nutzeraufgaben, Hardware, Kosten) bekannt sind, kann Performance mittels spezieller Verfahren (z.B. Last-, Zeit- oder Stresstests) gegen die Anforderungsspezifikation getestet werden. Eine typische Spezifikation wäre beispielsweise: »Ein Benutzer soll bei mindestens 95% aller Transaktionen nicht mehr als eine Sekunde auf die Verarbeitung der Eingabe und Darstellung der Ausgabe warten.«

Es stellt sich die Frage, auf welcher Grundlage die quantitativen Angaben festgelegt werden. Gibt es empirische Befunde, die kontext-unabhängig Gültigkeit beanspruchen, können die für

die Anforderungsspezifikation benötigten exakten Angaben aus ihnen abgeleitet werden. Dies ist beispielsweise bei der Festlegung des Antwortzeitverhaltens bei der Direktmanipulation der Fall. Hier ist hinreichend empirisch gesichert, dass schnelle Antwortzeiten zwischen 20 und 100 Millisekunden elementare Voraussetzung für das Gelingen einer dynamischen Interaktion sind (vgl. Shneiderman & Plaisant, 2004). Doch wie sollte vorgegangen werden, wenn keine abgesicherten Werte zur Verfügung stehen? Möglicherweise können Guidelines und Empfehlungen zu Rate gezogen. Allerdings enthalten sie selten exakte Angaben zur Lösung des Antwortzeitproblems (siehe Abschnitt 3.3).

Insgesamt betrachtet ist nicht auszuschließen, dass in software-technischer Hinsicht bei der Anforderungsspezifikation ein Vorgehen angewendet wird, das Teal und Rudnicky (1992) »guesstimate« nennen. Zukünftige Benutzer müssen sich dann beim Umgang mit den Softwaresystemen unter Umständen an ein relativ willkürlich festgelegtes Zeitverhalten gewöhnen. Auffällig ist, dass diese Gewöhnung in aller Regel gelingt. Wie sich im Laufe der noch jungen Geschichte der Mensch-Computer Interaktion gezeigt hat, sind Benutzer, wenn sie keine andere Wahl haben, in der Lage, ihr Handeln flexibel auf unterschiedliche Antwortgeschwindigkeiten interaktiver Systeme einzustellen. Die Anpassung des Menschen an die Technik ist unter software-ergonomischen Gesichtspunkten allerdings nicht vertretbar.

2.2 Performance in software-ergonomischer Sicht

In der Software-Ergonomie hat sich bekanntlich der englische Begriff Usability als Leitkonzept für Softwarequalität durchgesetzt. Usability wird zwar auch in der ISO 9126 definiert, maßgeblich ist heute allerdings das in der ISO 9241-Serie dokumentierte Konzept. Herausragendes Merkmal des Konzepts ist der kontextuelle Ansatz, der beim Gestalten und Bewerten von Softwareprodukten immer den jeweiligen Nutzungskontext in den Vordergrund stellt.

Performance wird in der Normenserie nicht ausdrücklich berücksichtigt. Zur Sprache kommt sie allerdings bei den Empfehlungen zum Gestaltungsgrundsatz »Erwartungskonformität« und als Systemvoraussetzung für die direkte Manipulation von Fenstern. Wie kann nun eine in software-ergonomischer Hinsicht akzeptable Festlegung der Performance erfolgen, ohne auf das genannte »guesstimate«-Vorgehen zurückzugreifen? Ein aktueller Forschungsbericht von Geis, Dzida und Redtenbacher (2004) mit Vorschlägen zur Revision der ISO 9241-Serie gibt dazu Auskunft: Zunächst muss grundsätzlich darauf geachtet werden, dass das Antwortzeitverhalten eines Computersystems den Erwartungen der Benutzer entspricht. Ist eine Erwartungskonformität gegeben, empfinden Benutzer die Verzögerungen zwischen Eingabe und Ausgabe nicht als Wartezeit. Gibt es bei bestimmten dynamischen Interaktionsformen empirische Evidenz dafür, dass fast alle Benutzer in fast allen Situationen ein bestimmtes Antwortzeitverhalten benötigen, kann darauf verzichtet werden, den Nutzungskontext zu berücksichtigen. Eine bestimmte Performance kann in diesem Fall direkt als Produktmerkmal festgeschrieben werden (Beispiel

Direktmanipulation). Gibt es hingegen keine allgemeingültige und quantitativ exakte Grundlage, und dieses ist beim aktuellen Stand der Forschung häufig zu erwarten (siehe Abschnitt 3), muss die Performance auf den konkreten Nutzungskontext eingestellt werden. Eine nicht-funktionale Anforderungsspezifikation lautet dann beispielsweise wie folgt: »Ein Benutzer soll nicht auf die Verarbeitung der Eingabe und Darstellung der Ausgabe warten.«

In software-technischer Sicht ist diese Formulierung ungewöhnlich unscharf und auf den ersten Blick nicht zu akzeptieren. Auf den zweiten Blick wird eine gemeinsame Sicht – und ebenfalls ein gemeinsames Vorgehen – von Software-Technik und Software-Ergonomie möglich. Denn auch in software-ergonomischer Sicht muss die allgemeine Formulierung operationalisiert und damit quantifiziert werden. Geis et al. (2004) schlagen dazu eine Vorgehensweise vor, die sie als »Usability Requirement Engineering« bezeichnen. Verfahren der Wahl, um ein Usability Requirement festzustellen, ist dabei u.a. die teilnehmende Beobachtung. Die im Qualitätsmanagement vorgesehenen Verfahren zur Überprüfung der Performance (Lasttests, Zeittests, Stresstests u.a.) werden bei einem solchen Vorgehen erst dann durchgeführt, wenn die Anforderungsspezifikation nutzungsangemessen und quantifizierbar formuliert werden konnte.

3.0 Befunde, Mythen und Empfehlungen

3.1 Befundlage zur Antwortzeitproblematik

Aus den Forschungsbemühungen im Bereich der Mensch-Computer-Interaktion und insbesondere der Software-Ergonomie ist bekannt, dass Benutzer relativ sensibel auf das Zeitverhalten von Computersystemen reagieren und ihr eigenes Verhalten darauf einstellen – sowohl kognitiv als auch emotional (Holling, 1989; Meyer & Hildebrandt, 2002; Shneiderman & Plaisant, 2004; Teal & Rudnick, 1992). Die menschliche Anpassungsfähigkeit ist in dieser Hinsicht sehr groß. Schnell bilden sich jedoch – willentlich oder unwillentlich – Erwartungshaltungen heraus, die die Wahrnehmung der Anwendung und den Umgang mit ihr deutlich beeinflussen.

Die Auswirkungen der technischen Performance auf das menschliche Verhalten sind bislang jedoch nur unzureichend erforscht. Es wurden Einzelbefunde erhoben, die nur selten aufeinander bezogen sind. Zudem ist die methodische Güte der Untersuchungen nicht als sehr hoch einzuschätzen, d.h. die erzielten Ergebnisse sind kaum generalisierbar (vgl. Holling, 1989). Es mangelt vor allem an generellen theoretischen Konzepten, die als roter Faden zum Verständnis der Antwortzeitproblematik beitragen könnten und Anhaltspunkte für systematische Untersuchungen liefern. Ben Shneiderman beklagte die Theoriearmut bereits 1984 in dem ersten Überblicksartikel zum Thema Systemantwortzeit (Shneiderman, 1984). Bezeichnenderweise wiederholt er die Klage in der aktuellen Neuauflage seines Standardwerks »Designing the user interface: Strategies for effective human-computer interaction« (Shneiderman & Plaisant, 2004). 20 Jahre sind vergangen,

ohne dass es einen Erkenntnisfortschritt gegeben hätte.

3.2 Mythen

3.2.1 »Die unendlich schnelle Maschine«

Software-Techniker neigen bei der Gestaltung von Software-Produkten dazu, das Qualitätsmerkmal Performance zu unterschätzen. Alan Dix zeigte bereits 1987, wie sich der Irrglaube, dass ein Computersystem von sich heraus immer schnell genug reagieren wird, stillschweigend in das Software Engineering eingeschlichen hat (Dix, 1987). Dieser Mythos existiert auch unter Netzbedingungen. So gibt es viele Anwendungen mit einer faszinierenden grafischen Web-Oberfläche, deren Benutzer aufgrund einer miserablen Performance bereits nach kurzer Zeit das Interesse verlieren. Der von Raskin (2000) beklagte Zustand, dass Anwendungen beim Starten viel zu lange brauchen und dadurch Benutzer in ihrer Interaktion schwerwiegend behindern, ist vermutlich auch ein Ausläufer des Glaubens an die unendlich schnelle Maschine.

3.2.2 »Je schneller, desto besser«

Die Maxime »Je schneller, desto besser« wird spätestens seit Einführung der Direktmanipulation regelmäßig geäußert (z. B. Smith, 1983). Sie geht unter anderem auf Untersuchungen zurück, die bis Mitte der 80er Jahre in Forschungszentren von IBM durchgeführt wurden. Es sollte nachgewiesen werden, dass schnelle Antwortzeiten der Schlüssel zu mehr Produktivität bei der Arbeit mit Timesharing-Systemen sind (Doherty & Thadhani, 1982). Der kommerzielle Erfolg der im Anschluss an diese Untersuchungen erfolgten Umstellung der Systemperformance spricht für die Plausibilität der Annahme. Andere Studien kommen allerdings zu

dem Ergebnis, dass es keinen linearen Zusammenhang zwischen Performance und Produktivität gibt (z. B. Barber & Lucas, 1983).

Die für die Direktmanipulation berechnete Forderung von sehr schnellen Antwortzeiten kann nicht verallgemeinert werden. Nach Hüttner, Wandke und Rätz (1995) ist es sogar der häufigste Fehler bei der Gestaltung der Antwortzeiten, sie in jedem Fall auf das technische Minimum zu reduzieren. »Der dadurch erzielte sehr schnelle Ablauf führt zwar meist momentan zu einer hohen Zufriedenheit mit dem System, jedoch haben Untersuchungen gezeigt, dass insbesondere bei sehr kurzen, konstanten Antwortzeiten die Beanspruchung des ungeübten Benutzers sich langfristig deutlich erhöht. Dies äußert sich in einer messbaren Zunahme der körperlichen Beschwerden und in einem Anstieg der Fehlerrate« (Hüttner et al., 1995). Gegen »Je schneller, desto besser« sprechen noch weitere Sachverhalte. Beispielsweise scheint es so zu sein, dass längere Antwortzeiten akzeptiert und sogar erwartet werden, wenn ihnen komplexe Eingaben vorausgehen.

3.2.3 »Die optimale Wartezeit«

Gibt es eine optimale Wartezeit? Eine Antwort auf diese regelmäßig wiederkehrende Frage steckt in den Gegenfragen: Für welche Tätigkeit? Unter welchen Umständen? Es gibt zurzeit kaum empirische Evidenz, die eine exakte Bestimmung einer optimalen Antwortzeit unabhängig vom Nutzungskontext möglich macht. Alle kursierenden Zahlen – 1, 2, 3, 4, 8, 10, 12 oder 15 Sekunden – gelten, wenn überhaupt, für spezielle Anwendungsfälle.

In diesem Zusammenhang sollte allerdings nicht übersehen werden, dass Benutzer ihre Erwartungen an das Zeitverhalten von einer Anwendung auf eine andere übertragen können. Aus der Praxis kann dabei vor allem für Web-Anwendungen empfohlen werden, einen relativen Wert im Wettbewerbsumfeld anzugeben statt eines absoluten Wertes. Ein Usability Requirement könnte zum Beispiel lauten: »Die Website soll im Mittel in ihrer Ladezeit die Top 3 des Wettbewerbsumfeldes erreichen.«

3.3 Empfehlungen

Die folgenden Empfehlungen, die sich im wesentlichen auf Hüttner et al. (1995) beziehen, bilden eine gute Grundlage für zukünftige Arbeiten in Theorie und Praxis. Die einzelnen Hinweise sind nicht unabhängig voneinander zu verstehen (zur Begründung und Vertiefung sei auf den online zur Verfügung stehenden Text von Hüttner et al. verwiesen).

- Systemantwortzeiten sind prinzipiell möglichst kurz zu gestalten, aber es ist auch darauf zu achten, dass sie nicht zu kurz sind.
- Systemantwortzeiten sind immer im engen Zusammenhang mit der Anwendung und der Dialogsituation zu gestalten.
- Systemantwortzeiten sollten nicht zu stark variieren. Die Streuung der Systemantwortzeiten ist kleiner als die Hälfte des Mittelwertes zu gestalten.
- Systemantwortzeiten im Sekundenbereich sind durch eine besondere Anzeige zu kennzeichnen. Bei längeren Pausen (spätestens ab 10 Sekunden) bietet sich eine Information über den Systemzustand an.

- Bei sehr langen Systemantwortzeiten (spätestens ab 30 Sekunden) sollte die verbleibende Wartezeit analog angezeigt werden. Möglich ist auch eine Ausgabe von Zwischenergebnissen als Vorabinformation bzw. das Ermöglichen der parallelen Bearbeitung von anderen Aufgaben.
- Systemantwortzeiten sollten für die zukünftige Benutzung veränderbar sein, so dass für ungeübte Anfänger und versierte Nutzer ein unterschiedliches Antwortzeitverhalten realisiert werden kann.

4.0 Fazit und Ausblick

»Performance ist eine Grundlage jeder Usability«. So prägnant lässt sich das Credo unseres Beitrages zusammenfassen. Dabei gibt es noch sehr viele Möglichkeiten, den Stand der Kunst in Theorie und Praxis zu verbessern. Ob das im Rahmen dieser Arbeit vorgestellte Usability Requirement Engineering produktiv eingesetzt werden kann, um die Performance-Qualität interaktiver Systeme zu verbessern, wird die Zukunft zeigen. Zunächst muss nachgewiesen werden, dass durch dieses Verfahren eine bessere interaktive Qualität erreicht wird als über das »guesstimate«-Vorgehen, welches immer in Gefahr steht, Mythen zu unterliegen. Auf jeden Fall ist es für Usability Professionals ein lohnenswertes Unterfangen, sich näher mit Performance zu befassen, da nahezu alle Anwendungen, die in verteilten Systemen ausgeführt werden, mit dem Antwortzeitproblem belastet sind.

5.0 Literatur

- Barber, R. E. & Lucas, H. C. (1983). System response time, operator productivity, and job satisfaction. *Communications of the ACM*, 26, 972-986.
- Dix, A. (1987). The myth of the infinitely fast machine. In D. Diaper & R. Winder (Eds.), *People and Computers III - Proceedings of HCI 1987* (pp. 215-228). Cambridge, NY: Cambridge University Press. Zugriff via <http://www.comp.lancs.ac.uk/computing/users/dixa/>
- Doherty, W. J. & Thadhani, A. J. (1982). The economic value of rapid response time (IBM Technical Report GE20-0752-0). Zugriff via <http://www.vm.ibm.com/devpages/jelliott/>
- Geis, T., Dzida, W. & Redtenbacher, W. (2004). Specifying usability requirements and test criteria for interactive systems: consequences for new releases of software-related standards within the ISO 9241 series (Publication series from the Federal Institute for Occupational Safety and Health, research report Fb 1010). Bremerhaven: Wirtschaftsverlag NW.
- Holling, H. (1989). Psychische Beanspruchung durch Wartezeiten in der Mensch-Computer Interaktion. Berlin: Springer.
- Hüttner, J., Wandke, H. & Rätz, A. (1995). Benutzerfreundliche Software. Psychologisches Wissen für die ergonomische Schnittstellengestaltung. Berlin: Paschke. Zugriff via <http://www.bmp.de/ISBN/3-929711-06-0/>
- Meyer, H. A. & Hildebrandt, M. (2002). Towards time design: Pacing of hypertext navigation by system response times. In L. Terveen & D. Wixon (Eds.), *CHI 2002 Conference on Human Factors in Computing Systems (Extended Abstracts)*, pp. 824-825). New York: ACM Press.
- Raskin, J. (2000). *The humane interface. New directions for designing interactive systems*. Reading, MA: Addison-Wesley.
- Shneiderman, B. (1984). Response time and display rate in human performance with computers. *ACM Computing Surveys*, 16, 265-285.
- Shneiderman, B. & Plaisant, C. (2004). *Designing the user interface: Strategies for effective human-computer interaction (4th ed.)*. Reading, Mass.: Addison-Wesley.
- Smith, D. (1983). Faster is better: A business case for subsecond response time. *Computerworld*, 18, 1-11.
- Teal, S.L. & Rudnick, A.I. (1992). A performance model of system delay and user strategy

5.0 Literatur

Barber, R. E. & Lucas, H. C. (1983). System response time, operator productivity, and job satisfaction. *Communications of the ACM*, 26, 972–986.

Dix, A. (1987). The myth of the infinitely fast machine. In D. Diaper & R. Winder (Eds.), *People and Computers III - Proceedings of HCI 1987* (pp. 215-228). Cambridge, NY: Cambridge University Press. Zugriff via <http://www.comp.lancs.ac.uk/computing/users/dixa/>

Doherty, W. J. & Thadhani, A. J. (1982). The economic value of rapid response time (IBM Technical Report GE20-0752-0). Zugriff via <http://www.vm.ibm.com/devpages/jelliott/>

Geis, T., Dzida, W. & Redtenbacher, W. (2004). Specifying usability requirements and test criteria for interactive systems: consequences for new releases of software-related standards within the ISO 9241 series (Publication series from the Federal Institute for Occupational Safety and Health, research report Fb 1010). Bremerhaven: Wirtschaftsverlag NW.

Holling, H. (1989). *Psychische Beanspruchung durch Wartezeiten in der Mensch-Computer Interaktion*. Berlin: Springer.

Hüttner, J., Wandke, H. & Rätz, A. (1995). *Benutzerfreundliche Software. Psychologisches Wissen für die ergonomische Schnittstellengestaltung*. Berlin: Paschke. Zugriff via <http://www.bmp.de/ISBN/3-929711-06-0/>

Meyer, H. A. & Hildebrandt, M. (2002). Towards time design: Pacing of hypertext navigation by system response times. In L. Terveen & D. Wixon (Eds.), *CHI 2002 Conference on Human Factors in Computing Systems (Extended Abstracts, pp. 824–825)*. New York: ACM Press.

Raskin, J. (2000). *The humane interface. New directions for designing interactive systems*. Reading, MA: Addison-Wesley.

Shneiderman, B. (1984). Response time and display rate in human performance with computers. *ACM Computing Surveys*, 16, 265–285.

Shneiderman, B. & Plaisant, C. (2004). *Designing the user interface: Strategies for effective human-computer interaction* (4th ed.). Reading, Mass.: Addison-Wesley.

Smith, D. (1983). Faster is better: A business case for subsecond response time. *Computerworld*, 18, 1–11.

Smith, D. (1983). Faster is better: A business case for subsecond response time. *Computerworld*, 18, 1–11.

Teal, S.L. & Rudnick, A.I. (1992). A performance model of system delay and user strategy selection. *Proceedings of ACM CHI 2002 Conference on Human Factors in Computing Systems* (p. 295–305). Monterey, CA: Addison-Wesley.

»Es ist erlaubt digitale und Kopien in Papierform des ganzen Papers oder Teilen davon für den persönlichen Gebrauch oder zur Verwendung in Lehrveranstaltungen zu erstellen. Der Verkauf oder gewerbliche Vertrieb ist untersagt. Rückfragen sind zu stellen an den Vorstand des GC UPA e.V. (Postfach 80 06 46, 70506 Stuttgart). *Proceedings of the 3rd annual GC UPA Track Linz*, September 2005 © 2005 German Chapter of the UPA e.V.«

