

Wie schnell ist „schnell“ bei Business-Software?

Analyse zur Performance bei der Nutzung von Business-Software

Wolfgang Bonhag

DATEV eG
Fürther Straße 111
90329 Nürnberg
wolfgang.bonhag@datev.de

Doreen Feindt

GfK SE
Burchardstraße 19
20095 Hamburg
doreen.feindt@gfk.com

Dr. Siegfried Olschner

DATEV eG
Fürther Straße 111
90329 Nürnberg
siegfried.olschner@datev.de

Ulf Schubert

DATEV eG
Fürther Straße 111
90329 Nürnberg
ulf.schubert@datev.de

Abstract

Die Performance einer Software ist ein wichtiger Einflussfaktor für die User Experience. Dabei hängt die Zufriedenheit mit der Performance eines Systems nicht nur von der tatsächlichen Geschwindigkeit, sondern auch von der Erwartungshaltung der Anwender ab. Wir stellen in diesem Beitrag die Ergebnisse einer Studie vor, in der wir der Frage nachgegangen sind, welche Geschwindigkeit bei spezifischen Interaktionskategorien vom Anwender akzeptiert wird. Dabei wurden 733 Teilnehmern simulierte System-Bearbeitungszeiten innerhalb einer simulierten Business-Software vorgegeben. Es wurden insgesamt 7 Aufgaben/Benutzerinteraktionen abgebildet (aus drei unterschiedlichen Interaktionskategorien), für welche jeweils 4 unterschiedliche System-Bearbeitungszeiten simuliert wurden. Aufgrund der erhobenen Daten konnte für zwei der drei Interaktionskategorien eine klare Empfehlung für die Performance von Business-Software abgeleitet werden.

Keywords

Performance-Messung, Interaktionskategorien, Bearbeitungszeiten, Ladezeiten, System-Bearbeitungszeiten

Einleitung

Die Geschwindigkeit von Software (Performance) wird von ihren Anwendern in der Regel nicht objektiv, sondern subjektiv eingeschätzt. Die Einschätzung basiert dabei nicht allein auf der tatsächlichen Reaktionszeit, sondern auch auf den mit vergleichbaren Systemen gemachten Erfahrungen und den Erwartungen an die Performance.

Um die einzelnen Aspekte zu berücksichtigen, welche die subjektiv wahrgenommene Performance ausmachen, verwendet DATEV das nachfolgende Modell (Schubert 2013).

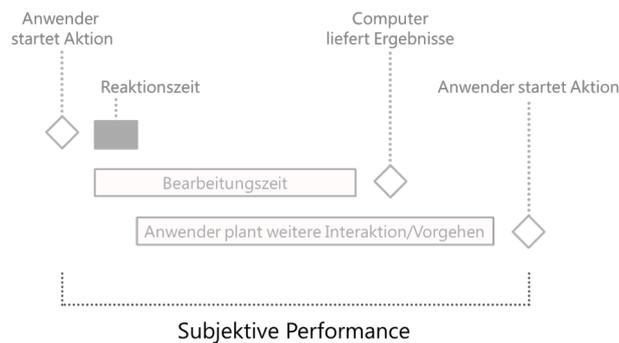


Abbildung 1: Elemente subjektiver Performance (Schubert 2013) – In diesem Artikel wird anstelle des Begriffs Bearbeitungszeit die Benennung System-Bearbeitungszeit verwendet.

Nach diesem Modell besteht die subjektiv wahrgenommene Performance aus der Reaktionszeit, der System-Bearbeitungszeit und der Zeit, die Anwender benötigen, um die nächsten Schritte zu planen. Die Reaktionszeit ist die Zeitdauer, die eine Anwendung benötigt, um auf die Eingabe des Anwenders zu reagieren. Die Reaktion kann entweder das direkte Ausführen der gewünschten Aktion oder ein Systemfeedback über die zu erwartende System-Bearbeitungszeit sein. Die Reaktionszeit ist etwas, worauf bei der Entwicklung häufig am meisten geachtet wird, denn diese kann man gut messen. Sie macht aber oft nur einen kleinen Teil der subjektiv erlebten Performance aus. Für das subjektive Erlebnis spielt die System-Bearbeitungszeit eine große Rolle. Diese beschreibt die Dauer, die Anwender und Anwendung benötigen, um eine Arbeitsaufgabe zu bearbeiten und abzuschließen. Betrachtet man die Performance umfassend, muss bei der Bewertung derselben auch die Zeit betrachtet werden, die ein Anwender zur Planung der weiteren Interaktionsschritte benötigt.

Für die Reaktionszeit ist es nicht notwendig, Zeitvorgaben zu ermitteln. Auf die Eingabe des Anwenders muss ein System generell innerhalb von wenigen Millisekunden reagieren - minimal mit einer Anzeige, dass eine Wartezeit erfolgt, maximal mit einer Meldung über die geschätzte Dauer der System-Bearbeitungszeit (Systemfeedback).

In der vorliegenden Studie haben wir uns auf die System-Bearbeitungszeit konzentriert. Die Hauptfragestellung war, wie lange die System-Bearbeitungszeit eines Systems bei

unterschiedlichen Interaktionen dauern darf, ohne die Zufriedenheit der Anwender mit dem System negativ zu beeinflussen. Dabei war es insbesondere interessant herauszufinden, welche Erwartungen Anwender an unterschiedliche Interaktionen haben. Weiterhin ging es uns um die Ableitung von Empfehlungen von System-Bearbeitungszeiten für bestimmte Interaktionsformen, die als Leitlinie für die Entwicklung bzw. als nicht-funktionale Anforderungen verwendet werden können.

Untersuchungsmethode

Allgemeines Vorgehen

Zur Vorgabe der System-Bearbeitungszeiten an eine große Teilnehmeranzahl und zur Abfrage einer Bewertung derselben wurde im Jahr 2012 von GfK und DATEV ein Online-Fragebogen konzipiert, der neben klassischen Frageformen auch einen Klickdummy enthielt. Dieser Klickdummy war in Form einer fiktiven Business-Software mit klassischem Bildschirmaufbau gestaltet und stellt nacheinander mehrere Screenshots mit einer simulierten System-Bearbeitungszeit dar.

Die dargestellten fiktiven Oberflächen enthielten vor allem Listen und Tabellen. Ein optischer Bezug zu einer bestimmten im Markt befindlichen Business-Software wurde vermieden.

Die Genauigkeit, in der die System-Bearbeitungszeiten simuliert werden konnten, wurde leicht durch den verwendeten Browser und die Internet-Verbindungsgeschwindigkeit beeinflusst. Um die Messfehler zu mitteln, war eine hohe Anzahl an Messungen notwendig.

Stichprobe

Insgesamt beteiligten sich 733 Personen an der Studie. Das Durchschnittsalter betrug 42,4 Jahre (zwischen 18 und 80). 45,6 % der Teilnehmer waren weiblich, 54,4 % der Teilnehmer waren männlich.

Alle Teilnehmer verwendeten bei ihrer beruflichen Arbeit Business-Software und nutzten diese mindestens „mehrmals pro Woche“. Weiterhin war die Nutzung einer DSL-Internetverbindung Voraussetzung, und es konnten nur Nutzer von Desktop-PCs oder Notebooks teilnehmen.

Die Rekrutierung der Teilnehmer erfolgte aus einem Online-Panel mit Mitgliedern aus unterschiedlichen Unternehmen.

Zur Überprüfung, ob bezüglich Performance eine einheitliche Erwartungshaltung besteht, wurde mit Hilfe einer allgemeinen Kontrollfrage nach der subjektiv erlebten Performance von Microsoft Word gefragt. 86 % der Teilnehmer bewerten die Performance von Word als "angemessen" (Ladezeit zwischen einem Klick und dem Erscheinen des Ergebnisses). Deswegen betrachten wir die Stichprobe als ausreichend homogen.

Maskierung des Untersuchungsziels

Um eine zu starke Fokussierung der Teilnehmer auf die Warte- bzw. System-Bearbeitungszeit zu vermeiden, wurde das Hauptziel der Studie nicht explizit genannt. Die Befragung wurde allgemein als „Untersuchung der Qualität einer Business-Software“ angekündigt.

Neben den Fragen zur Einschätzung der System-Bearbeitungszeit wurden zwei zusätzliche Fragen zur Gestaltung der Software und zur Auffindbarkeit von Bedienelementen gestellt. Diese wurden nicht ausgewertet.

Die Fragen lauteten:

Wie beurteilen Sie die Ladezeit zwischen Ihrem Klick und dem Erscheinen der Ergebnisse? (Antwortmöglichkeiten: „Zu schnell“, „Angemessen“ und „Zu langsam“).

Maskierungsfrage: Wie schwierig war es für Sie, das richtige Bedienelement zum Start der Aktion zu finden? (Antwortmöglichkeiten: „Überhaupt nicht schwierig“, „Etwas schwierig“ und „Sehr schwierig“).

Maskierungsfrage: Wie hat Ihnen die Gestaltung der Oberfläche gefallen? (Antwortmöglichkeiten: „Sehr gut“, „Gut“, „Teils-teils“, „Schlecht“ und „Sehr schlecht“).

Bei der Frage zur System-Bearbeitungszeit/Ladezeit wurde bewusst eine 3-stufige Skala gewählt („Zu schnell“, „Angemessen“ und „Zu langsam“). Eine stärkere Unterteilung wie z.B. in "etwas zu langsam" und "viel zu langsam" hätte für die spätere Festlegung eines Bewertungssystems keinen Mehrwert geliefert, da sowieso eine dichotome Kategorisierung angestrebt wurde (Ladezeit in Ordnung vs. nicht in Ordnung). Die Antwortkategorie "zu schnell" entspricht in diesem Zusammenhang eher einer Kontrolloption.

Versuchsbeschreibung

Interaktionskategorien

Im Rahmen dieser Grundlagen wurden unterschiedliche Interaktionsformen zu Interaktionskategorien gruppiert. Alle Interaktionsformen beziehen sich auf die Interaktion mittels Maus und Tastatur (Shneiderman 1984; Seow 2008). Eine Interaktionskategorie (IK) beschreibt eine Gruppe von Interaktionsformen, für die eine ähnliche Erwartungshaltung hinsichtlich Performance besteht. Das Konzept der Interaktionskategorien wurde von DATEV auf die eigenen Bedürfnisse angepasst. Insgesamt wurden fünf Interaktionskategorien definiert. Die System-Bearbeitungszeiten bei diesen fünf Interaktionskategorien reichen von wenigen Millisekunden in der Interaktionskategorie 1 (IK1) bis hin zu mehreren Minuten oder gar Stunden in der Interaktionskategorie 5 (IK5). Die konzipierten Interaktionskategorien werden wie folgt definiert:

IK1: Alle Interaktionen, die in Zusammenhang mit der Hand-Auge-Koordination stehen und bei denen der Anwender vom System ein direktes Feedback erwartet und benötigt. (z. B. Tastatureingabe, Anzeigen eines Mouseovers, Anzeigen eines Textcursors nach Klick in einem Eingabecontrol).

IK2: Alle Interaktionen, die mit der direkten Manipulation bzw. Änderung der Oberfläche zusammenhängen und bei denen der Anwender eine zeitnahe Umsetzung der angeforderten Aktion erwartet (z. B. Ein-/Ausblenden eines Bereichs der Bedienoberfläche, Sortieren einer Liste, Wechsel einer Registerkarte).

IK3: Alle Interaktionen, die in Arbeitsprozesse integriert sind. "Integriert" bedeutet, dass vor und nach dieser Interaktion weitere Interaktionen vom Anwender zu tätigen sind (z. B. Öffnen eines Dialogfensters mit umfangreichen Daten, Start einer Unteranwendung aus einer bereits geöffneten Anwendung heraus, Refresh der gesamten Oberfläche).

IK4: Alle Interaktionen zur Vorbereitung oder am Ende eines Arbeitsprozesses oder Tätigkeit, wie z. B. Neustart einer Anwendung, umfangreiche Berechnungen oder Erstellen von Reports. Beim Beenden eines Arbeitsprozesses wird davon ausgegangen, dass der Anwender danach seine Interaktion mit dem System in einem anderen Arbeitsprozess fortsetzt.

IK5: Alle Interaktionen zur Vorbereitung oder am Ende eines Arbeitsprozesses, bei denen der Anwender von vornherein eine längere System-Bearbeitungszeit erwartet, wie z. B. Installation, Systemwartung, Update. Üblicherweise dauern diese Interaktionen länger.

- Die vorliegende Studie konzentriert sich auf die Interaktionskategorien 2 bis 4, da für die Interaktionskategorien 1 und 5 Zeitangaben entweder offensichtlich oder nicht sinnvoll sind.

System-Bearbeitungszeit und Systemfeedback

Für die verwendeten drei Interaktionskategorien wurde jeweils ein eigenes Set von vier System-Bearbeitungszeiten vorgegeben, z. B. 0, 1, 2 oder 3 Sekunden bei IK2 [Tabelle 1], wobei 0 Sekunden einen idealen Wert darstellen. Ebenso wurde für jede Interaktionskategorie ein übliches spezifisches Systemfeedback angezeigt, z. B. eine Sanduhr bei IK2 und ein Fortschrittsbalken bei IK4.

Aufgabenstellungen

Insgesamt wurden jedem Teilnehmer vier Aufgaben vorgelegt. Pro Interaktionskategorie gab es zwei bis drei verschiedene Interaktionen, die mit jeweils vier verschiedenen System-Bearbeitungszeiten getestet wurden, woraus sich 28 Aufgaben-Bearbeitungszeit-Kombinationen ergeben. Die Auswahl der vier Aufgaben pro Teilnehmer aus dem Set von 28 möglichen Kombinationen erfolgte zufällig. Allerdings wurde ausgeschlossen, dass ein Teilnehmer die gleiche Interaktion zweimal beurteilt, nur jeweils mit unterschiedlichen Zeiten. Jede der 28 Kombinationen wurde von mindestens 100 Teilnehmern bearbeitet. Die

Ergebnisdaten der Kombinationen zeichnen sich somit durch eine hohe Teilnehmeranzahl aus.

Nr.	IK	Beispiel für eine Aufgabenstellung	System-Bearbeitungszeiten in Sek.	Feedback während der System-Bearbeitungszeit
1	2	Wechsel einer Liste von einer Listenansicht in eine Detailansicht	0, 1, 2, 3	Sanduhr (ab 1 sec System-Bearbeitungszeit)
2	2	Sortieren einer Liste nach einer anderen Spalte	0, 1, 2, 3	
3	3	Öffnen eines Arbeitsblattes aus der Übersicht links	3, 5, 7, 9	Animierte Grafik (drehende Uhr mit Erläuterungstext)
4	3	Wechsel des angezeigten Dokuments über einen Navigator	3, 5, 7, 9	
5	3	Datenbearbeitung aus der laufenden Anwendung	3, 5, 7, 9	
6	4	Start einer Anwendung vom Desktop	9, 13, 17, 21	Fortschrittsbalken mit festem Erläuterungstext
7	4	Erstellen eines umfangreichen Reports, incl. Grafiken	9, 13, 17, 21	

Tabelle 1: Sieben Aufgaben in jeweils vier System-Bearbeitungszeiten ergeben 28 Aufgaben -Bearbeitungszeit-Kombinationen.

Ziel einer jeden Aufgabe ist es, eine einzige Frage zu beantworten: War die gerade erlebte System-Bearbeitungszeit bezogen auf die Interaktion „zu schnell“, „angemessen“ oder „zu langsam“? Um die Erwartungshaltung zu setzen, enthielt jede Aufgabenbeschreibung eine Information über die zu erwartende Reaktion des Systems.

- Der Ablauf bei jeder der 28 Einzelaufgaben war stets gleich gestaltet. Die unterschiedlichen Aufgaben und System-Bearbeitungszeiten wurden den Teilnehmern zufällig rotiert vorgegeben:

Der Anwender erhält eine aufgabenspezifische Anleitung:

Beispiel IK3: Bitte öffnen Sie den Kontenplan, indem Sie auf den entsprechenden Eintrag im Dropdown-Menü klicken. Sie öffnen damit ein neues Fenster, in dem die gewählte Anwendung gestartet wird.

Nach dem Klick auf das angesprochene Bedienelement startet die System-Bearbeitungszeit, und entsprechend der Interaktionskategorie wird das Systemfeedback eingeblendet.

Nach Ablauf der System-Bearbeitungszeit wird das Ergebnis der Interaktion angezeigt, z. B. öffnet sich der Kontenplan (Abbildung 2).

Danach wurden die drei Bewertungsfragen vorgelegt.

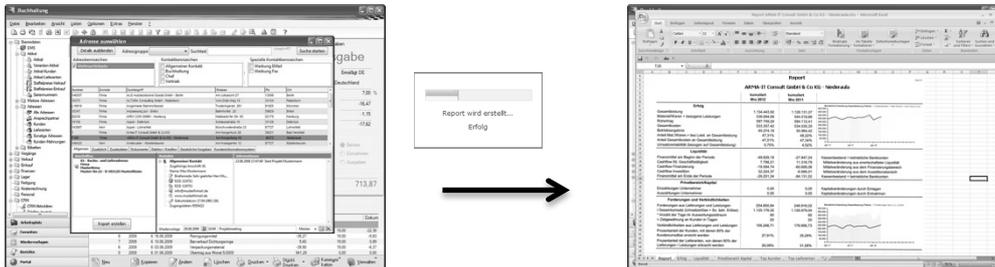


Abbildung 2: Beispiel für die Simulation des Öffnens eines Reports.

Ergebnisse

Interaktionskategorie 2

Die Ergebnisse der Aufgaben aus IK2 sind nicht aussagekräftig. Bei den Bewertungen ist kein Trend erkennbar; es werden sehr lange System-Bearbeitungszeiten toleriert. Vermutlich ist die Fragebogensituation nicht mit einer „echten“ Arbeitssituation vergleichbar. Abbildung 3 zeigt kaum Veränderungen bei den Prozentverteilungen der Antwortkategorien.

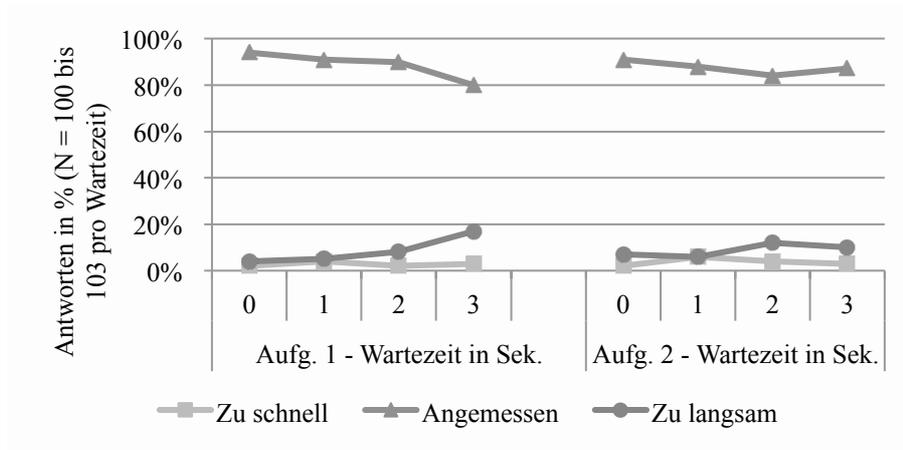


Abbildung 3: Bewertung der System-Bearbeitungszeit bei Wechsel von einer Listenansicht in eine Detailansicht (Aufgabe 1) und Sortieren einer Liste nach einer anderen Spalte (Aufgabe 2).

Interaktionskategorie 3

Bei IK3 ist bei zwei Aufgaben ein klarer Trend erkennbar. Alle drei Aufgaben zeigen eine interpretierbare Verlaufslinie für die Bewertung „zu langsam“ (Abbildung 4).

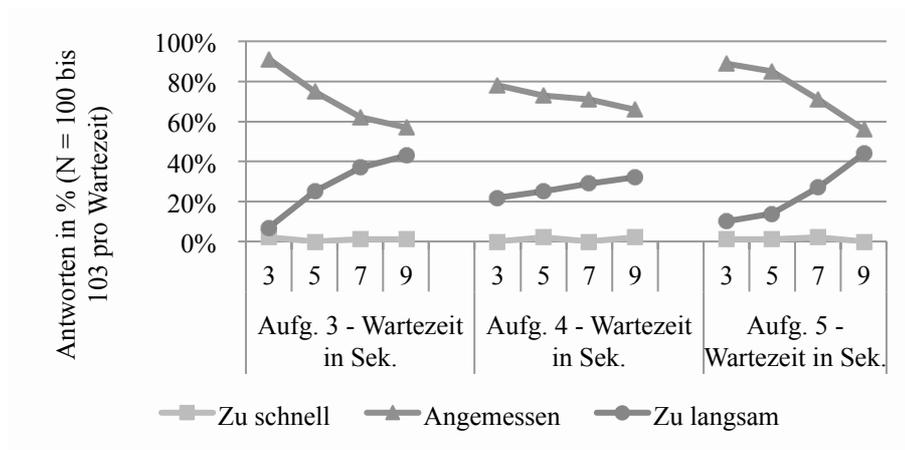


Abbildung 4: Öffnen eines Arbeitsblattes aus der Übersicht links (Aufgabe 3) und Wechsel des angezeigten Dokuments über einen Navigator (Aufgabe 4) und Datenbearbeitung aus der laufenden Anwendung (Aufgabe 5).

Interaktionskategorie 4

IK4 zeigt ebenfalls interpretierbare Messergebnisse. In einer Aufgabe kommt es zu einer Überschneidung der Linien „Angemessen“ und „Zu langsam“ [Abbildung 5].

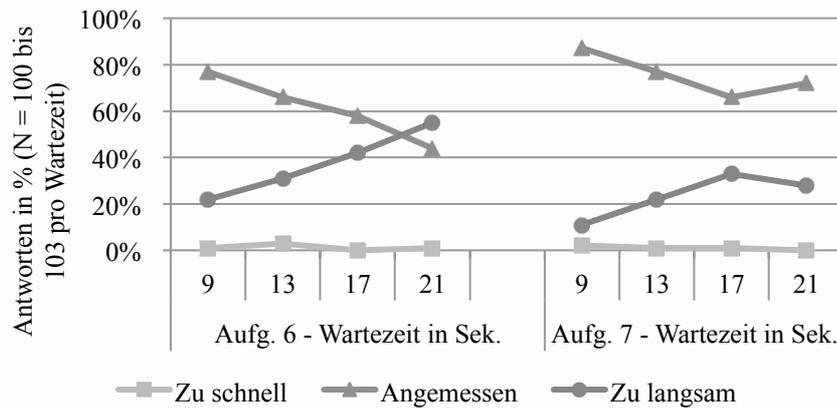


Abbildung 5: Start einer Anwendung vom Desktop (Aufgabe 6) und Erstellen eines umfangreichen Reports, incl. Grafiken (Aufgabe 7).

Interpretation der Bewertungen

Die in der Studie erhobenen Bewertungen sind nur aussagefähig in Verbindung mit einem Norm-System. Dieses wurde von uns individuell festgelegt und könnte je nach Anwendungskontext auch anders justiert werden.

Bei der Interpretation der Daten wird der Anteil der Studien-Teilnehmer, die mit einer erlebten System-Bearbeitungszeit unzufrieden sind, herangezogen. Diese bestimmen durch ihre Anzahl die Qualität der gezeigten Performance. Je höher der Anteil der unzufriedenen Teilnehmer bei einer System-Bearbeitungszeit ist, desto schlechter ist die Bewertung der System-Bearbeitungszeit. Tabelle 2 zeigt unseren Vorschlag für eine strenge Auslegung der Bewertungen (Norm-System).

„Schnell genug“	„Sollte schneller sein“	„Zu langsam“	„Viel zu langsam“
Bearbeitungszeiten, mit denen weniger als 10 % der Teilnehmer der Studie unzufrieden sind	Bearbeitungszeiten, mit denen 10 bis 20 % der Teilnehmer der Studie unzufrieden sind	Bearbeitungszeiten, mit denen 20 bis 30 % der Teilnehmer der Studie unzufrieden sind	Bearbeitungszeiten, mit denen mehr als 30 % der Teilnehmer der Studie unzufrieden sind

Tabelle 2: Beispiel für eine strenge Auslegung der Bewertung (Norm-System).

Übertragung der Werte in Anforderungstabellen

Anhand der Verteilungen der Bewertungen können auf Basis der in Tabelle 2 festgelegten Prozentwerte Anforderungen an die System-Bearbeitungszeiten in IK3 und IK4 formuliert werden. Tabelle 3 zeigt die Umsetzung in Sekundenwerte.

System-Bearbeitungszeiten, die im Bereich "sind schnell genug" liegen, bedürfen keiner Optimierung. Liegen die System-Bearbeitungszeiten allerdings in den Bereichen "sollten schneller sein", "sind zu langsam" oder "sind viel zu langsam", besteht Handlungsbedarf.

Die Bearbeitungszeiten ...	Interaktionskategorie 3	Interaktionskategorie 4
... sind schnell genug	Bis 3,5 Sekunden	Bis 7 Sekunden
... sollten schneller sein	3,5 bis 5,5 Sekunden	7 bis 10,5 Sekunden
... sind zu langsam	5,5 bis 7 Sekunden	10,5 bis 14,5 Sekunden
... sind viel zu langsam	Ab 7 Sekunden	Ab 14,5 Sekunden

Tabelle 3: Empfehlungen für System-Bearbeitungszeiten der Interaktionskategorie 3 und 4

Diskussion

Die große Anzahl der Teilnehmer ermöglicht es, mit dieser Studie valide Messwerte zu präsentieren. Die Übertragbarkeit der Ergebnisse auf reale Business-Software muss jedoch mit Vorsicht erfolgen, da die Messungen nicht in einer realen Arbeitssituation erfolgten. Bei der Übertragung der Ergebnisse ist es notwendig, die Erwartungshaltungen der Anwender an die Software zu berücksichtigen. Es ist davon auszugehen, dass sich die Erwartungshaltungen zwischen unterschiedlichen Anwendergruppen und Nutzungskontexten unterscheiden. Außerdem verändert die Erfahrung der Anwender im Umgang mit digitalen Produkten und Dienstleistungen die Erwartungshaltung im Zeitverlauf. Es ist davon

auszugehen, dass die Erwartungen hinsichtlich der Dauer von System-Bearbeitungszeiten steigen.

Die Empfehlungen für die System-Bearbeitungszeiten in IK3 und IK4 können aktuell als Orientierung für die Einschätzung der System-Bearbeitungszeiten in Business-Software herangezogen werden. Weiterhin können sie als nicht-funktionale Anforderungen in die Entwicklung von Business-Software übernommen werden.

Literatur

- Butler, T. W. (1983). Computer Response Time and Use Performance. In *Proceedings of ACM SIGCHI '83 Conference on Human Factors in Computer Systems (1983)*, 58-62. New York: ACM.
- Miller, R. B. (1968). Response time in man-computer conversational transactions. In *Proc. AFIPS Fall Joint Computer Conference Vol. 33*, 267-277. Montvale, NJ: AFIPS Press.
- Seow, S. C. (2008). *Designing and Engineering Time: The Psychology of Time Perception in Software*. Boston MA: Addison Wesley Professional
- Schubert, U. (2013) *Performance ist nicht objektiv messbar*. http://www.user-experience-blog.de/2013/08/performance_ist_nicht_objektiv. Letzter Aufruf: Juni 2015.
- Shneiderman, B. (1984). Response Time and Display Rate in Human Performance with Computers. In *ACM Computing Surveys*, Vol. 16, No. 3. New York: ACM