

Mensch und Computer im Einklang – Die Doherty-Schwelle

Die Produktivität steigt, wenn Computer und Nutzer in einem bestimmten Tempo (< 400 ms) miteinander interagieren, sodass keiner auf die Gegenseite warten muss.

Wichtige Erkenntnisse

- Gewährleisten Sie innerhalb von 400 ms eine Systemrückmeldung, um die Aufmerksamkeit der Anwender aufrechtzuerhalten und die Produktivität zu steigern.
 - Machen Sie sich die wahrgenommene Leistung zunutze, um die Reaktionszeiten zu verbessern und Wartezeiten scheinbar zu verkürzen.
 - Animationen bieten die Möglichkeit, Menschen während des Ladens oder eines im Hintergrund ablaufenden Verarbeitungsprozesses visuell zu beschäftigen.
 - Unabhängig von ihrer Genauigkeit tragen Fortschrittsbalken dazu bei, Wartezeiten erträglicher zu gestalten.
 - Eine gezielte Verlangsamung eines Prozesses kann dessen wahrgenommenen Wert erhöhen und Vertrauen vermitteln, selbst wenn der eigentliche Vorgang viel weniger Zeit in Anspruch nimmt.
-

Überblick

Die Performance gehört zu den entscheidenden Merkmalen einer guten User Experience. Wenn der Benutzer beim Versuch, eine Aufgabe zu erledigen, mit langsamen Verarbeitungsgeschwindigkeiten, fehlendem Feedback oder zu langen Ladezeiten zu kämpfen hat, kann das schnell zu Frustration führen und sich nachhaltig negativ auswirken. Häufig wird Geschwindigkeit eher als technisch wünschenswerte Zielsetzung verkannt, obwohl sie eigentlich als wesentliches und für positive Benutzererfahrungen entscheidendes Designmerkmal betrachtet werden sollte. Unabhängig davon, ob es sich um die anfängliche Ladezeit des Produkts oder der Dienstleistung handelt oder um die Reaktionsgeschwindigkeit auf Eingaben und bei der Ausgabe von Rückmeldungen oder

um die Ladegeschwindigkeit von Folgeseiten – mit der Reaktionsgeschwindigkeit eines Systems steht und fällt die gesamte User Experience.

Es gibt mehrere Faktoren, die sich auf die Leistung von Websites und Apps auswirken können, der wichtigste ist jedoch das *Page Weight*, also die Gesamtgröße einer Seite mit allen zugehörigen Dateien, Skripten und Medien. Leider hat diese Kennzahl durchschnittlich betrachtet im Laufe der Jahre exponentiell zugenommen. Laut *HTTP Archive* betrug die durchschnittliche Größe einer Desktop-Seite im Jahr 2023 schon über 2 MB (2286 KB), während mobile Webseiten mit knapp über 2 MB (2007 KB) nicht wesentlich kleiner waren. Das ist ein gewaltiger Zuwachs gegenüber den durchschnittlichen Seitengrößen in den Jahren 2010 bis 2011: 634 KB auf dem Desktop und 260 KB auf Mobilgeräten (Abbildung 10-1).

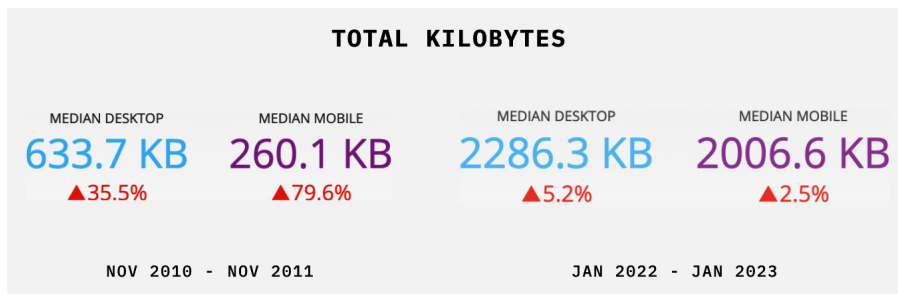


Abbildung 10-1: Die durchschnittliche Datenmenge pro Seite nimmt jedes Jahr zu (Quelle: HTTP Archive, 2023).

Dieser Trend führt zu längeren Wartezeiten, und niemand wartet gerne, wenn es eine Aufgabe zu erfüllen gilt. Unzählige Studien bestätigen die Tatsache, dass Anwender mit steigender Wartezeit eher frustriert werden und das verfolgte Ziel vielleicht sogar ganz abschreiben.

Außerdem verringern träge Reaktionszeiten des Systems die Produktivität der Schnittstellenanwender. Während eine Reaktionszeit von 100 ms als unmittelbar empfunden wird, ist eine Verzögerung zwischen 100 und 300 ms für das menschliche Auge bereits erkennbar, und die Anwender bekommen das Gefühl, weniger Kontrolle über das Geschehen zu haben. Sobald die Verzögerung über 1000 ms (1 Sekunde) hinausgeht, fangen die Nutzer an, über andere Dinge nachzudenken; ihre Aufmerksamkeit schweift ab, und zur Erfüllung der Aufgabe wichtige Informationen gehen allmählich verloren, was unweigerlich zu Leistungseinbußen führt. Die zur Fortsetzung der Aufgabe erforderliche kognitive Belastung nimmt zu und die gesamte User Experience leidet.

Ursprung

In der Anfangszeit des Desktopcomputings galten 2 Sekunden bei der Ausführung einer Aufgabe noch als akzeptabler Schwellenwert für die Reaktionszeit eines Computers. Der Grund für diesen weithin akzeptierten Standard war, dass er dem Benutzer Zeit ließ, über seine nächste Aufgabe nachzudenken. Im Jahr 1982 stellte dann die von den IBM-Mitarbeitern Walter J. Doherty und Ahrvind J. Thadani im *IBM Systems Journal* veröffentlichte Studie »The Economic Value of Rapid Response Time« diesen früheren Standard durch die Feststellung infrage, dass »die Produktivität bei einer Verringerung der Reaktionszeit überproportional ansteigt«, solange diese über der kritischen Schwelle von 400 ms liegt.⁴³ Die Studienautoren erklärten: »Wenn ein Computer und seine Benutzer in einem Tempo miteinander interagieren, das sicherstellt, dass keiner auf den anderen warten muss, schnellte die Produktivität in die Höhe, die Kosten der am Computer geleisteten Arbeit sinken, die Mitarbeiter empfinden ihre Arbeit als befriedigender und deren Qualität verbessert sich tendenziell.« Damit setzten sie einen neuen Standard, der als *Doherty Threshold* bekannt wurde und der auf Dohertys Beobachtung eines unverhältnismäßig großen Einflusses der Reaktionsgeschwindigkeit von Computern auf die Produktivität basierte.

Psychologisches Konzept

Flow

Flow ist ein geistiger Zustand, in dem eine Person bei einer Tätigkeit voll und ganz in das Gefühl der energiegeladenen Konzentration, der vollen Hingabe und der Freude an dieser Aktivität eintaucht. Der Begriff wurde 1970 von dem Psychologen Mihály Csíkszentmihályi geprägt und ist in vielen Bereichen weit verbreitet (besonders in der Beschäftigungstherapie), obwohl das Konzept angeblich schon seit Tausenden von Jahren unter anderen Namen existiert.⁴⁴

Der Flow-Zustand tritt bei einem Gleichgewicht zwischen der Schwierigkeit einer Aufgabe und dem Fähigkeitsniveau bei der jeweiligen Aufgabe ein. Er zeichnet sich durch intensive und fokussierte Konzentration auf die Gegenwart aus, verbunden mit einem Gefühl der totalen Kontrolle. Eine zu schwierige Aufgabe kann etwa zu erhöhter Frustration führen, während eine zu leichte Aufgabe zu Langeweile führen kann. Für das richtige Gleichgewicht muss die Herausforderung an die Fähigkeiten des Benutzers oder der Benutzerin angepasst werden. Wenn sich eine Person in einem Flow-Zustand befindet, geht sie völlig in ihren Aufgaben auf, ist frei von innerer Kritik und bis zu fünfmal produktiver.⁴⁵

43 Die Originalstudie ist mittlerweile verloren gegangen, aber Sie finden eine Kopie in Jim Elliotts Mainframe-Blog: <https://oreil.ly/8Gpua>.

44 Mihály Csíkszentmihályi, »Flow: The Psychology of Optimal Experience«, New York: Harper & Row, 1990.

45 Joshua Gold und Joseph Ciorciari, »A Review on the Role of the Neuroscience of Flow States in the Modern World«, *Behavioral Sciences* 10, Nr. 9 (2020): 137, <https://doi.org/10.3390/bs10090137>.

Wir können das Eintreten des Flow-Zustands erleichtern, indem wir das nötige Feedback geben, damit der Benutzer weiß, was er getan und erreicht hat, die Effizienz und die Reaktionsfähigkeit des Systems optimieren, unnötige Reibungsverluste vermeiden, Inhalte und Funktionen zum Ausprobieren bereitstellen und damit die Benutzeroberfläche interessanter gestalten.

Beispiele

In manchen Fällen dauert die Verarbeitung länger als die durch den Doherty Threshold vorgeschriebenen > 400 ms, ohne dass man viel dagegen unternehmen könnte. Das heißt aber nicht, dass wir den Benutzern nicht rechtzeitig eine Rückmeldung geben können, während im Hintergrund die erforderliche Verarbeitung stattfindet. Mit dieser Technik lässt sich der Eindruck erwecken, dass eine Website oder eine Anwendung schneller läuft, als es tatsächlich der Fall ist.

Ein gängiges Beispiel, das von Plattformen wie Instagram verwendet wird (Abbildung 10-2), ist die Darstellung eines Bildschirm-Skeletts beim Laden von Inhalten. Durch diese Technik scheint die Website schneller zu laden, indem in den Bereichen, in denen schließlich der Inhalt erscheinen wird, unmittelbar Platzhalterblöcke angezeigt werden. Die Blöcke werden nach und nach durch echte Texte und Bilder ersetzt, sobald diese geladen wurden. Dadurch wird der Warteeindruck verringert, was die Wahrnehmung von Geschwindigkeit und Ansprechverhalten erhöht, auch wenn der Inhalt langsam geladen wird. Darüber hinaus verhindern Bildschirm-Skelette das irritierende und störende Umherspringen von Inhalten beim Laden von angrenzendem Material, weil so für jedes Element im Voraus der entsprechende Platz reserviert wird.

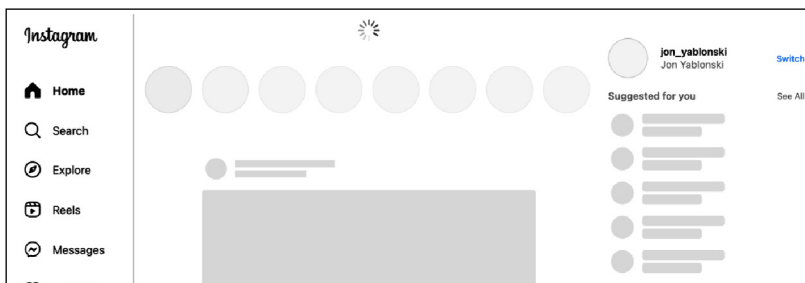


Abbildung 10-2: Das Bildschirmskelett von Instagram erweckt den Anschein, dass die Website schneller geladen wird (Quelle: Instagram, 2023).

Eine andere Möglichkeit zur Optimierung von Ladezeiten ist als »Blur-up«-Technik bekannt. Dieser Ansatz konzentriert sich speziell auf Bilder, die oft auch die Hauptursache für übermäßig lange Ladezeiten sowohl im Web als auch bei nativen Anwendungen sind. Dabei wird zunächst eine sehr kleine Bildversion geladen und auf den Bereich hochskaliert, in den schließlich das größere Bild geladen wird. Ein Gauß'scher Weichzeichner kaschiert die offensichtliche Pixelbildung und das Bildrauschen durch die Vergrößerung des niedrig aufgelösten Bilds (Abbildung 10-3). Sobald die größere Bildversion im Hintergrund nachgeladen wurde, wird sie hinter der niedrig aufgelösten Version platziert und durch Ausblenden des oberen Bilds sichtbar gemacht. Diese Technik gewährleistet nicht nur schnellere Ladezeiten, sondern reserviert vorab auch die komplette Bildschirmfläche für Bilder, was ein Springen der Seiten nach dem vollständigen Laden der hochauflösenden Bildversion verhindert.

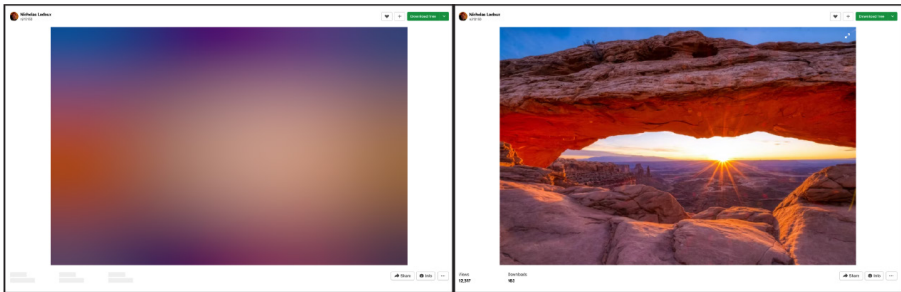


Abbildung 10-3: Unsplash verwendet die »Blur-up«-Technik, um ein schnelleres Laden der Seite zu ermöglichen (Quelle: Unsplash, 2019).

Animationen sind eine weitere Möglichkeit, Benutzer visuell einzubinden, während im Hintergrund ein Lade- oder Verarbeitungsprozess stattfindet. Ein gängiges Beispiel sind sogenannte Fortschrittsbalken oder prozentuale Fortschrittsanzeigen. Untersuchungen haben gezeigt, dass der bloße Anblick eines Fortschrittsbalkens die Wartezeiten unabhängig von seiner Genauigkeit erträglicher wirken lassen kann.⁴⁶ Dieses einfache UI-Element ist aus mehreren Gründen so wirksam:

- Es gibt den Anwendern die Gewissheit, dass ihre Aktion verarbeitet wird.
- Es bietet visuelle Reize während des Wartens.
- Es verringert den Eindruck des Wartens, indem es den Fokus vom eigentlichen Wartevorgang hin zur Animation des Fortschrittsbalkens verschiebt.

⁴⁶ Brad A. Myers, »The Importance of Percent-Done Progress Indicators for Computer-Human Interfaces«, in *CHI, 85: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, New York: Association for Computing Machinery, 1985, 11–17.

Zwar können wir die erforderliche Verarbeitung und die entsprechenden Wartezeiten nicht immer umgehen, aber wir können die Wartebereitschaft der Benutzer durch visuelles Feedback steigern.

Ein Beispiel für eine Animation zur Verringerung der mit Wartezeiten einhergehenden Ungewissheit und Frustration findet sich in Googles E-Mail-Client Gmail (Abbildung 10-4). Der Ladebildschirm verwendet eine animierte Version des Logos zusammen mit einem einfachen Fortschrittsbalken. Diese einfache, aber unverkennbare Animation lässt die Wartezeit insgesamt kürzer erscheinen und verbessert zugleich die allgemeine User Experience, indem sie den Nutzern signalisiert, dass die Anwendung geladen wird.



Abbildung 10-4: Gmail verwendet eine einfache, aber markante Animation, um die gefühlte Wartezeit zu verkürzen (Quelle: Gmail, 2023).

Die allgemein anerkannte Grenze, um die Aufmerksamkeit des Benutzers auf der anstehenden Aufgabe zu halten, beträgt zehn Sekunden – sobald diese Zeitspanne überschritten wird, möchte der Anwender während der Wartezeit andere Dinge erledigen.⁴⁷ Wenn die Wartezeiten das Maximum von zehn Sekunden übersteigen müssen, sind Fortschrittsbalken weiterhin hilfreich, sie sollten allerdings auch um eine Schätzung der verbleibenden Restzeit und um eine Beschreibung der gerade ausgeführten Aktion ergänzt werden. Diese Zusatzinformationen geben den Anwendern eine bessere Vorstellung davon, wie lange sie bis zum Abschluss der Aktion warten müssen, und ermöglichen ihnen, sich zwischenzeitlich anderen Dingen zu widmen. Nehmen Sie zum Beispiel den Installationsbildschirm von Apple (Abbildung 10-5), der während einer Aktualisierung angezeigt wird.

Eine weitere raffinierte Technik zur Verbesserung der wahrgenommenen Leistung stellt das sogenannte *optimistische UI* dar. Es liefert bereits optimistische

⁴⁷ Robert B. Miller, »Response Time in Man-Computer Conversational Transactions«, in *Proceedings of the December 9-11, 1968, Fall Joint Computer Conference, Part I*, Band. 33, New York: Association for Computing Machinery, 1968, 267–277.

Rückmeldungen, dass eine Aktion erfolgreich war, während diese noch bearbeitet wird, statt erst nach Abschluss der Aktion ein Feedback zu geben. So zeigt beispielsweise Instagram Kommentare zu Fotos schon vor ihrer tatsächlichen Veröffentlichung (Abbildung 10-6). Dadurch wirkt die Reaktionszeit der App kürzer, als sie es tatsächlich ist: Sie vermittelt sofort ein visuelles Feedback, das davon ausgeht, dass der Kommentar erfolgreich veröffentlicht wird, und zeigt erst nachträglich einen Fehler an, falls die Aktion doch schiefgeht. Die erforderliche Verarbeitung findet im Hintergrund weiterhin statt, aber das Leistungsempfinden der Benutzer der App wird verbessert.

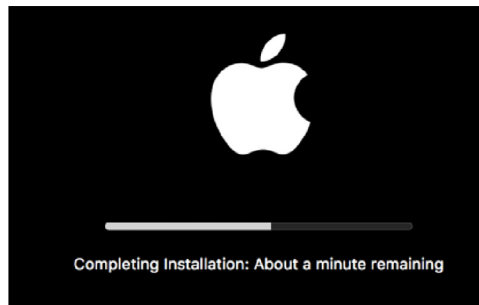


Abbildung 10-5: Während der Aktualisierungen gibt Apple zusammen mit einem Fortschrittsbalken die geschätzte Zeit bis zur Fertigstellung an (Quelle: Apple macOS).

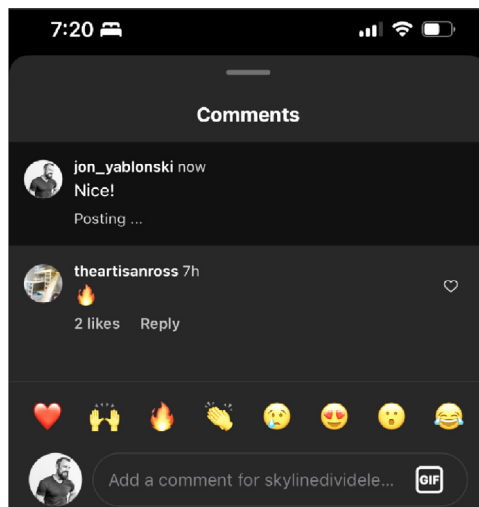


Abbildung 10-6: Instagram zeigt Fotokommentare in optimistischer Weise bereits vor der eigentlichen Veröffentlichung an, um die empfundene Leistung zu verbessern (Quelle: Instagram, 2023).

Schlüsselüberlegung

Reibung

Die meisten Probleme im Zusammenhang mit Reaktionszeiten resultieren daraus, dass diese zu lang sind. Es mag nicht besonders einleuchtend wirken, aber bedenken Sie, dass Reaktionszeiten eventuell auch *zu kurz* sein können. Wenn das System schneller reagiert, als vom Benutzer erwartet, können einige Probleme auftreten. Zunächst einmal kann eine etwas zu schnell erfolgte Änderung völlig übersehen werden – das gilt insbesondere dann, wenn die Änderung nicht als Ergebnis einer Benutzeraktion, sondern automatisch ausgeführt wurde. Ein weiteres Problem, das ein zu rasantes Ansprechverhalten nach sich ziehen kann, besteht darin, dass es für den Benutzer mitunter schwierig wird, das Geschehene nachzuvollziehen, da das Tempo der Veränderung nicht genügend Zeit für die mentale Verarbeitung lässt. Schließlich kann eine zu kurze Reaktionszeit Misstrauen erwecken, wenn sie nicht mit den Erwartungen des Benutzers an die ausgeführte Aufgabe übereinstimmt. Eine absichtlich herbeigeführte Verzögerung kann den wahrgenommenen Wert eines Prozesses erhöhen und mehr Vertrauen vermitteln, selbst wenn der eigentliche Vorgang viel weniger Zeit in Anspruch nimmt.⁴⁸ Reibung kann sorgfältig eingesetzt werden, wenn eine überlegte Reaktion verlangt wird. Zum Beispiel kann eine Bestätigungsanzeige genug Reibung erzeugen, um den aufwendigen und evaluativen mentalen Prozess des »langsamen Denkens« zu aktivieren (siehe Kapitel 7) und so die Wahrscheinlichkeit von Fehlern oder Irrtümern zu verringern. Ein weiteres Beispiel ist Googles Datenschutzprüfung (Abbildung 10-7), die Ihr Konto auf potenzielle Sicherheitslücken prüft. Google nutzt die Gelegenheit, um seine Nutzer darüber aufzuklären, was gescannt wird, und verlängert den Vorgang zusätzlich, um mehr Vertrauen in die Gründlichkeit des Scans zu schaffen.

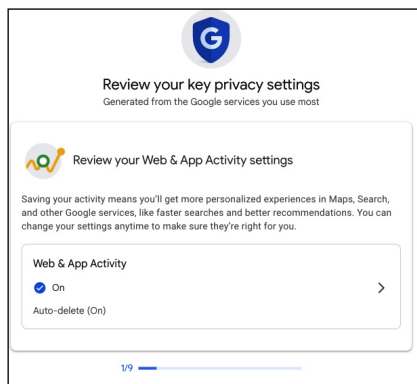


Abbildung 10-7: Die Datenschutzprüfung von Google scannt Ihr Konto auf potenzielle Sicherheitslücken. Dabei wird die Zeit, die der Prozess eigentlich benötigt, verlängert und die Gelegenheit zur Aufklärung genutzt (Quelle: Google, 2023).

48 Mark Wilson, »The UX Secret That Will Ruin Apps for You«, Fast Company, 6. Juli 2016, <https://www.fastcompany.com/3061519/the-ux-secret-that-will-ruin-apps-for-you>.

Fazit

Die Performance ist nicht nur eine technische Überlegung für unsere Kollegen in der Programmierabteilung – sie ist ein wesentliches Gestaltungsmerkmal. Es liegt in unserer Verantwortung als Designer, dass die Nutzer unserer Produkte und Dienstleistungen ihre Aufgaben so schnell und effizient wie möglich erledigen können. In diesem Zusammenhang ist es wichtig, das richtige Feedback zu geben, die gefühlte Performance ausnutzen und Fortschrittsbalken einzusetzen, um die gefühlte Wartedauer insgesamt zu verringern.

Inhaltsverzeichnis

	Einleitung	7
1	Der Mensch ist ein Gewohnheitstier – Jakobs Gesetz	19
2	Minimaler Aufwand für maximalen Erfolg – Fitts’ Gesetz	31
3	Kleine Einheiten, großer Erfolg – Millers Gesetz	43
4	Weniger Auswahl, schnellere Reaktion – Hicks Gesetz	53
5	Vorausschauend planen – Postels Gesetz	65
6	Höhepunkte schaffen – Die Höchststand-Ende-Regel	77
7	Das Gesetz der Schönheit – Der Aesthetic-Usability-Effekt	89
8	Aus der Masse herausstechen – Der Restorff-Effekt	101
9	Ein bisschen Komplexität muss sein – Teslers Gesetz	113
10	Mensch und Computer im Einklang – Die Doherty-Schwelle	123
11	Psychologische Prinzipien in der Gestaltung anwenden	133
12	Macht bedeutet Verantwortung	143
	Index	158