

Eine Taxonomie für Längsschnittstudien in der MCI

Jens Gerken & Harald Reiterer

Arbeitsgruppe Mensch-Computer Interaktion, Universität Konstanz

Zusammenfassung

Während klassische Experimente oder Usability Tests oft nur eine Momentaufnahme darstellen, ermöglichen es Längsschnittstudien auch Veränderungen zu beobachten und zu einem ökologisch validen Gesamtbild der Gebrauchstauglichkeit zu gelangen. Bislang werden derartige Verfahren aber noch sehr selten in der Mensch-Computer Interaktion eingesetzt. In diesem Beitrag wird auf Basis einer umfangreichen Literaturanalyse eine Taxonomie vorgestellt, die einen Überblick über die vielfältigen Möglichkeiten von Längsschnittstudien ermöglicht. Hierdurch kann sowohl der wissenschaftliche Diskurs verbessert werden, um die bekannten Ansätze weiter zu verbessern, als auch dem praktischen Anwender Hilfestellung in Planung und Durchführung einer Längsschnittstudie gegeben werden.

1 Einleitung

Der Mensch ist tagtäglich mit Veränderungen konfrontiert – sowohl er selbst als auch die Umwelt sind in stetem Wandel. Viele empirische Untersuchungsmethoden in der Mensch-Computer Interaktion basieren jedoch auf der Annahme, durch eine einmalige Datenerhebung das gesamte Spektrum der Nutzung abbilden zu können – sei es um die Gebrauchstauglichkeit eines interaktiven Produktes zu bewerten, Anforderungen für die Gestaltung eines solchen zu erheben oder um mehr über die Beziehung zwischen Mensch und Computer zu erfahren. Letztlich ist aber davon auszugehen, dass die gewonnenen Erkenntnisse aus solchen Querschnittstudien dem Lichtkegel einer Taschenlampe gleichen – mit etwas Pech ist der erhellte Bereich in keinster Weise repräsentativ. Längsschnittstudien versuchen dieser Problematik entgegenzutreten, indem sie zu mehreren Zeitpunkten Daten erheben und diese in Beziehung setzen und somit den Wandel oder die Dynamik, wie Wilamowitz-Moellendorff et al. es nennen (Wilamowitz-Moellendorff et al. 2008) in Handlungen oder Einstellungen sichtbar machen. In den Sozialwissenschaften und der Psychologie ist diese methodische Herangehensweise bereits weiter verbreitet (Taris 2000). Formal definieren lassen sich derartig erhobene Daten wie folgt durch Kontrastierung mit Querschnittsdaten: „longitudinal data present information about what happened to a set of research units during

a series of time points. In contrast, cross-sectional data refer to the situation at one particular point in time” (Taris 2000). Nach Courage et al. (2009) ist eine Studie dann als Längsschnitt zu bezeichnen „if at least one dimension – e.g. users, design, research questions – is compared over a time period“. Hieraus geht bereits hervor, dass es sich bei Längsschnittstudien nicht um eine spezifische Methode handelt, sondern vielmehr um ein Untersuchungsparadigma. Im englischsprachigen Raum wird daher auch zumeist der umfassendere Begriff „longitudinal research“ verwendet (z.B. Menard 2007). Auch in der Mensch-Computer Interaktion sind Längsschnittstudien in dem Sinne keine neue Entdeckung. Bereits 1978 führten Card et al. (1978) ein Experiment im Längsschnitt durch, um das Lernverhalten bei Nutzung eines neuen Eingabegerätes zu untersuchen. Ebenso sind viele Feldstudien, beispielsweise am Arbeitsplatz oder zur Anforderungsanalyse prinzipiell im Längsschnitt angelegt. Eine systematische Auseinandersetzung mit der Thematik im Allgemeinen scheint verstärkt aber erst in den letzten Jahren stattzufinden (z.B. Veranstaltungen von Courage et al. (2009) auf Fachtagungen wie der CHI oder UPA Conference). Auch andere Forscher betonen immer wieder die Notwendigkeit von Längsschnittstudien, beispielsweise um besser untersuchen zu können, wie Informationsvisualisierung in die Arbeitspraxis integriert wird (González & Kobsa 2003) oder um zu untersuchen, welche Art von Usability Problemen auch über die Zeit hinweg bestehen bleiben (Kjeldskov et al. 2006). Der erhöhte zeitliche Aufwand der Durchführung und die gesteigerte Komplexität bezüglich Planung und Analyse schränken den breiten Einsatz dieser Forschungsmethodik jedoch weiterhin ein. Menard (2007) führt zudem aus, dass im Prinzip sämtliche Fallstricke von Querschnittstudien auch auf Längsschnittstudien übertragbar sind und dazu noch einige neue hinzukommen. In den existierenden Studien in der MCI werden darüberhinaus nur gelegentlich methodische Aspekte diskutiert (z.B. Wilamowitz-Moellendorff 2007, Shneiderman & Plaisant 2006), ebenso fehlt bislang eine umfassende Diskussion in Textbüchern wie „Interaction Design“ (Sharp et al. 2007). Dies erschwert nicht nur die Anwendung von Längsschnittstudien im konkreten Fall mangels Hilfestellung sondern auch den wissenschaftlichen Diskurs, um die Vorgehensweise methodisch voran zu bringen. In diesem Beitrag möchten wir daher, aufbauend auf einer eingehenden Analyse existierender Längsschnittstudien eine Taxonomie für eben diese in der MCI vorstellen. Diese soll sowohl einen Überblick über den Stand der Anwendung bieten als auch die wesentlichen Aspekte, die für Planung und Durchführung relevant sind, definieren und in Zusammenhang stellen. Ziel ist es, für die weitere Forschung in diesem Gebiet eine solide Diskussionsgrundlage zu schaffen, die es zudem Praktikern ermöglicht, besser abschätzen zu können, ob eine Längsschnittstudie eine lohnende Alternative darstellen könnte.

2 Eine Taxonomie für Längsschnittstudien

Die grundlegende Basis für die Taxonomie bietet eine umfangreiche Analyse der wissenschaftlichen Fachliteratur in der Mensch-Computer Interaktion. Hierzu wurden etwa 80 Beiträge zum Thema Längsschnittstudien, welche auf einschlägigen Fachtagungen (z.B. CHI, Interact, Mensch & Computer) und in Zeitschriften (z.B. Int. J. of Human-Computer Studies) veröffentlicht wurden gesichtet. Dabei wurde offensichtlich, dass viele Beiträge nur wenige Informationen hinsichtlich der Methodik liefern sondern lediglich anmerken, dass eine

Längsschnittstudie eingesetzt wurde. Für eine eingehende Analyse wurden daher 32 Beiträge ausgewählt, welche aus methodischer Sicht den höchsten Informationsgehalt boten (Eine Übersicht findet sich hier: http://hci.uni-konstanz.de/downloads/longitudinal_studies.pdf). Ziel war es daraufhin, die methodischen Gemeinsamkeiten und Unterschiede herauszuarbeiten und somit ein umfassendes Bild für Längsschnittstudien in der Mensch-Computer Interaktion zu erhalten. Weiterhin wurde die einschlägige Literatur in den Sozialwissenschaften ebenfalls in diese Analyse mit einbezogen, insbesondere das Ende 2007 erschiene Handbook of Longitudinal Research (Menard 2007). Im Folgenden soll die Taxonomie zunächst grob umrissen werden woraufhin im Anschluss ausgewählte Aspekte im Detail diskutiert werden. Zur besseren Übersicht zeigt Abbildung 1 zudem die Taxonomie grafisch auf.

Als zentrales Unterscheidungskriterium der diversen analysierten Längsschnittstudien lassen sich zu aller erst die Unterschiede hinsichtlich der **Forschungs- bzw. Untersuchungsfragen** heranziehen. Während die Sozialwissenschaften hier explizit den Wandel und Veränderungen (z.B. unterschiedliches Wahlverhalten über die Jahre) untersuchen möchten, so werden in der Mensch-Computer Interaktion Längsschnittstudien auch eingesetzt, um ein ganzheitlicheres Bild einer Produktnutzung zu erhalten oder um die Veränderung als konfundierenden Faktor ausschließen zu können. In Kapitel 2.1 wird detailliert anhand von Beispielen erläutert, welche konkreten Fragestellungen dies sein können.

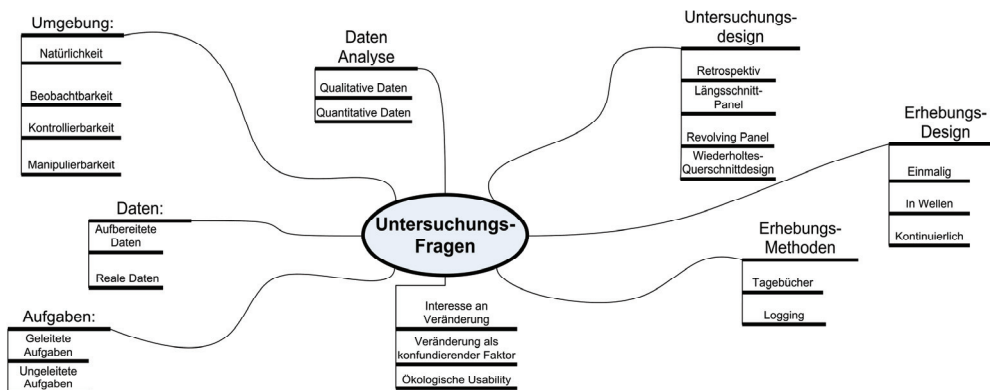


Abbildung 1: Grafische Darstellung der Taxonomie

Die weiteren Aspekte der Taxonomie bauen auf diesem zentralen Element auf. Zunächst seien hier die verschiedenen möglichen **Untersuchungsdesigns** genannt. Die Mensch-Computer Interaktion zeigt sich hier bislang wenig kreativ und setzt zumeist auf Panel-Designs und retrospektive Analysen. In Kapitel 2.2 werden die Vor- und Nachteile dieser Designs diskutiert und zudem noch weitere Design-Möglichkeiten aus den Sozialwissenschaften aufgezeigt, die einige der bekannten Probleme beheben können. In direktem Zusammenhang zu den Untersuchungsdesigns stehen die **Erhebungsdesigns**. Darunter versteht man den Ablauf der Datenerhebung, beispielsweise ob diese einmalig, kontinuierlich oder in Wellen erfolgt. Dieser Punkt wird in Kapitel 2.3 eingehend erörtert. Bezüglich der **Erhebungsmethoden** werden zum einen prinzipiell bekannte Methoden aus Querschnittstudien

eingesetzt, beispielsweise Interviews oder kontrollierte Beobachtungen im Experiment, zum anderen aber verstärkt Methoden die implizit für längere Untersuchungen ausgelegt sind, beispielsweise Tagebücher oder Logging-Verfahren. Diese werden in Kapitel 2.4 eingehender diskutiert.

Darüberhinaus lassen sich Längsschnittstudien hinsichtlich der verwendeten **Aufgaben**, der hierfür zugrunde liegenden **Daten** und der **Umgebung**, in welcher diese durchgeführt werden, unterscheiden. Im Rahmen dieses Beitrags können diese drei Aspekte jeweils nur knapp angerissen werden (siehe Kapitel 2.5) – eine eingehende Diskussion wird in Folgearbeiten weitergeführt werden. Abschließend muss noch der Aspekt der **Datenanalyse** benannt werden. Dieser Aspekt ist insbesondere relevant, wenn der Wandel in der Benutzung untersucht werden soll, da ansonsten klassische Verfahren der Datenanalyse herangezogen werden können, seien es z.B. grounding theory Ansätze für qualitative Daten oder Signifikanztests im Falle von quantitativen Daten. Wenn aber der Wandel selbst untersucht werden soll offenbaren diese Verfahren methodische Schwächen. In Kapitel 2.6 wird somit abschließend diskutiert, welche Herausforderungen es hier zu meistern gilt.

2.1 Untersuchungsfragen

Jede empirische Untersuchung sollte von ihren Forschungsfragen getrieben werden. Daher sind es auch diese, die letztlich die Durchführung einer Längsschnittstudie nahe legen müssen. Wie bereits aufgeführt lassen sich in den existierenden Studien insgesamt drei zentrale Untersuchungsfragen identifizieren – wobei gewisse Überschneidungen nicht ausbleiben.

Interesse an Veränderung: Auch in der MCI ist die Veränderung an sich natürlich eine Forschungsfrage, welche von Interesse ist. Beispielsweise ist es für den Softwarehersteller, der jede neue Version einem abschließenden Usability Test unterzieht durchaus von Interesse, ob die neue Version weniger Probleme bereitet als die vorige und ob die ergriffenen Maßnahmen diese Veränderung in der Produktnutzung hervorgerufen haben. Weiterhin gibt es Untersuchungen, die sich damit beschäftigen, wie das Erlernen einer Software oder eines Eingabegerätes genau abläuft, welche Strategiewechsel im Laufe der Zeit erfolgen und wie dies unterstützt werden kann (Mendoza & Novick 2005). Das Ziel hierbei ist nicht zwangsläufig die konkrete Verbesserung eines Produktes sondern die Steigerung des allgemeinen Verständnisses der Mensch-Computer Interaktion (z.B. van der Kleij et al. 2005, Bergman et al. 2008). Dieses kann dann wiederum für zukünftige Produktentwicklungen von Nutzen sein.

Veränderung als ein konfundierender Faktor: Auf der anderen Seite kann die Veränderung auch einen Störfaktor darstellen, der kontrolliert werden soll. Nicht die Veränderung selbst ist von Interesse, sondern was an ihrem Ende steht. Beispielsweise beim Vergleich von neuen mit bereits gelernten Eingabegeräten (z.B. MacKenzie & Zhang 1999, Bieg 2008) oder bei der Frage, welche Usability Probleme auch nach einiger Nutzungszeit existieren und somit als besonders kritisch zu betrachten sind (z.B. Kjeldskov et al. 2005).

Ökologische Usability: Mittels einer Längsschnittstudie kann eine umfassende und reale Benutzungssituation simuliert ebenso wie eine reale Nutzungssituation beobachtet werden. Daraus ergibt sich die Möglichkeit, eine validere Bewertung des Produkts vorzunehmen. Studien mit dem Fokus auf diesem Aspekt klammern die Veränderung oftmals weitgehend

aus – teilweise sind auch die Zeiträume zu kurz, um überhaupt systematische Veränderungen feststellen zu können (z.B. Gerken et al. 2008, McLachlan et al. 2008, Jain 2008).

2.2 Untersuchungsdesign

Das Untersuchungsdesign gibt vor, in welchem Rahmen die Datenerhebung stattfinden kann und welche Methoden einsetzbar sind. Zudem lassen sich mit unterschiedlichen Untersuchungsdesigns auch unterschiedliche Fragestellungen abdecken.

Retrospektives Design: Beispielsweise durch Befragung von Probanden, welche erlebte Ereignisse im Nachhinein rekonstruieren und wiedergeben (z.B. Corpus Interview Technik, Wilamowitz-Moellendorff et al. 2007) ermöglicht ein retrospektives Design den Blick in die Vergangenheit. Die erhobenen Daten sind dabei sehr stark subjektiv beeinflusst und für den Forscher nur schwer überprüfbar. Der Vorteil liegt in der Möglichkeit, Phänomene zu untersuchen, welche bei ihrem Auftreten übersehen wurden oder nicht untersuchbar waren. Zudem sind sie sehr Ressourcen schonend.

Panel Design: Dies ist das klassische, prospektive Längsschnittdesign (z.B. Kjeldskov et al. 2005, Bieg 2008). Hierbei werden ein oder mehrere Untersuchungsobjekte (also zumeist Probanden) mehrmals oder durchgehend über einen längeren Zeitraum beobachtet. Dies ermöglicht die Untersuchung von Veränderungen und deren Ursachen, insbesondere wenn der Forscher aktiv Variablen manipuliert, also beispielsweise ein neues Feature in ein Produkt integriert und die Reaktion darauf beobachtet. Allerdings ist ein Panel-Design auch nicht frei von Nachteilen. Zum einen muss grundsätzlich damit gerechnet werden, dass im Laufe der Studie Probanden nicht mehr zur Verfügung stehen – also die Studie abbrechen. Zum anderen muss beachtet werden, dass die Studie selbst Einfluss auf sich nimmt – Probanden ändern eventuell ihre Meinung oder ihre Art der Benutzung, weil sie durch die Studie dazu angeregt werden, diese zu hinterfragen. Wenn der Forscher aktiv Variablen manipuliert, ist dies insbesondere der Fall. Abhilfe kann hierbei das folgende Design schaffen.

Revolving-Panel Design: In dieser Variante eines Längsschnitt-Panel Designs werden zu festgelegten Zeitpunkten neue Probanden systematisch integriert, wohingegen ältere ausscheiden. Hierdurch lässt sich beispielsweise zu Zeitpunkt X feststellen, ob ein beobachteter Effekt in der Längsschnittgruppe auch bei den neu hinzugekommenen Probanden beobachtbar ist und somit eventuell auf externe Effekte zurückzuführen ist. Weiterhin eignet sich dieses Design gut, um während der Studie neue Fragestellungen zu integrieren. Dies ist ein wichtiger Aspekt sämtlicher Längsschnittstudien – durch die Dauer können sich viele Aspekte, auch die Forschungsfrage ändern. Die Gefahr, die bislang gesammelten Daten zu verlieren, wenn diese Änderungen auch im Design berücksichtigt werden, ist jedoch sehr groß. Das Revolving-Panel Design erlaubt es, hier flexibler neue Fragestellungen parallel zu der Integration neuer Probanden einzuführen. Soweit es uns bekannt ist, wurde in der MCI jedoch bislang kein derartiges Design eingesetzt – ein Grund könnte in dem erhöhten Planungsaufwand liegen

Längsschnitt-Fall Studie: Ein Spezialfall eines Panel-Designs stellt die Längsschnitt-Fall-Studie oder Case Study dar. In dieser wird nur eine sehr geringe Anzahl an Versuchsperso-

nen über einen möglichst langen Zeitraum beobachtet. Ziel ist es hierbei, möglichst tief in Details vorzustoßen. Im Gegenzug kann es aber schwer sein, die gewonnenen Ergebnisse zu verallgemeinern. In der MCI wurde dieser Ansatz insbesondere von Shneiderman & Plaisant (2006) unter dem Namen MILC – Multi-dimensional in-depth Case Study eingeführt.

Wiederholtes Querschnitt-Design (auch Trend-Study): In diesem Fall wird eine Querschnittstudie mehrfach wiederholt, allerdings unter der Annahme, dass der geänderte Zeitpunkt einen Einfluss hat bzw. es wurden vom Forscher selbst Variablen manipuliert, um die Veränderung zu beobachten. Dabei wird jedoch nicht auf die gleiche Probandengruppe zurückgegriffen, sondern es wird lediglich versucht, eine gleichwertige Gruppe für die Studie heranzuziehen. Sinnvoll ist dieser Ansatz immer dann, wenn die Dauer zwischen den Testphasen relativ lange sind, beispielsweise beim Vergleich einer Software mit dessen Vorgängerversion. Dieser Ansatz steigert somit die Praktikabilität einer Längsschnittstudie, denn oftmals stehen in solchen Fällen Versuchspersonen nicht über mehrere Jahre verlässlich zur Verfügung. Allerdings kann auch nicht untersucht werden, inwieweit eventuell äußere Einflüsse, wie beispielsweise die Verbreitung eines bestimmten Interaktionskonzeptes in Konkurrenzprodukten, die Benutzung des eigenen Produkts verändert haben. Zudem ist es nicht möglich, intra-individuelle Veränderungen zu analysieren, beispielsweise das Erlernen einer neuen Visualisierung.

2.3 Erhebungsdesign

Die Art der Datenerhebung ist ein entscheidender Aspekt, welcher Längsschnittstudien von Querschnittstudien unterscheidet. Die Ausnahme bilden retrospektive Designs, die ebenso nur auf eine einmalige Datenerhebung angewiesen sind. Üblicherweise findet die Datenerhebung jedoch *in Wellen* statt (z.B. MacKenzie & Zhang 1999, Kjeldskov et al. 2005), also in zumeist regelmäßigen Abständen (beispielsweise einmal in der Woche, einmal pro Monat, etc.), um somit klare Aussagen über den Einfluss des Zeit-Faktors treffen zu können. Unregelmäßige Abstände sind dann sinnvoll, wenn die Erhebung ereignisbasiert erfolgt, beispielsweise immer wenn eine neue Version einer Software veröffentlicht wird oder der Proband ein neues Mobiltelefon kauft. Ebenso gibt es Datenerhebungsmethoden (siehe nächstes Kapitel), welche inhärent eine ereignisbasierte Datenerhebung ermöglichen (z.B. Tagebuch-techniken). Zunehmend häufiger werden zudem kontinuierliche Erhebungs-Designs, welche entsprechend der Bezeichnung ohne Unterbrechung Daten erheben (z.B. Shneiderman & Plaisant 2006) und somit ein ganzheitlicheres Bild der Benutzung ermöglichen. Möglich wird dies mit technischen Logging Verfahren (z.B. Interaktion, Bild, Ton) und auch entsprechenden Analysetechniken, ohne welche die hierbei anfallenden enormen Datenmengen nicht beherrschbar wären. Die Dauer einer Längsschnittstudie ist letztlich von der Untersuchungsfrage abhängig, wird aber natürlich auch durch das gewählte Erhebungs-Design beeinflusst. Eine kontinuierliche Datenerhebung über mehrere Jahre hinweg scheint zumindest zum jetzigen Zeitpunkt noch impraktikabel zu sein. Prinzipiell kann eine Längsschnittstudie aber von wenigen Stunden, über Tage bis hin zu Monaten oder gar Jahren dauern.

2.4 Erhebungsmethoden

Im Prinzip werden bei Längsschnittstudien oftmals dieselben Methoden zur Datenerhebung eingesetzt wie bei traditionellen Querschnittstudien, also beispielsweise Interviews, unterschiedlich standardisierte Formen der Beobachtung oder Fragebögen (für einen Überblick siehe Sharp et al 2007). Dies ist insbesondere der Fall, wenn die Studie im Labor durchgeführt wird. Im Feld sind jedoch insbesondere Logging Verfahren und ereignisbasierte Verfahren wie Tagebuch-Techniken hervorzuheben. Logging Verfahren werden eingesetzt, um im Hintergrund über einen längeren Zeitraum Daten zu sammeln (z.B. Alexander & Cockburn 2009). Der Vorteil liegt darin, große Datenmengen ohne viel Aufwand für den Forscher zu erlangen. Die Schwierigkeit wiederum besteht darin, aus diesen Daten dann entsprechende Informationen zu extrahieren. Fehlen bei quantitativen Interaktionsdaten oftmals die Kontextinformationen so macht bei Video oder Audiodaten schlicht der Umfang der Daten Probleme. Insbesondere wenn Veränderungen untersucht werden sollen muss das Logging zudem bereits im Vorfeld sehr genau auf die entsprechende Fragestellung ausgerichtet sein und überlegt werden, anhand von welchen Ereignissen Veränderungen überhaupt festgestellt werden können. Ein anderer Ansatz stellen Tagebuchtechniken dar (Rieman 1993), bei welchen dem Proband die Rolle der Datenerhebung zukommt und dieser, je nach Art der Durchführung, selbst entscheiden kann, wann er Informationen in das ausgehändigte Tagebuch niederschreibt. Auch hier wird der Forscher also bei der Erhebung entlastet, jedoch steigt auch die Unsicherheit, da Probanden ihre Einträge auch manipulieren können. Ebenso wie bei Logging-Verfahren müssen zudem Aspekte integriert werden, die eine Beobachtung von Veränderung möglich machen. Ein einfaches Beispiel hierfür sind Rating-Skalen für die Bewertung von Produktaspekten.

2.5 Aufgaben, Daten und die Umgebung

Auf diese Aspekte kann im Rahmen dieses Aufsatzes nur verkürzt eingegangen werden. Prinzipiell bieten Längsschnittstudien bei Aufgaben, Daten und Umgebung eine vielfältige Auswahl. Bezüglich der Aufgaben können beispielsweise **geleitete Aufgaben**, bei denen der Forscher die Tätigkeit direkt vorgibt (z.B. Fitts' Tapping Test in Bieg 2008) ebenso eingesetzt werden wie **ungeleitete Aufgaben**, bei denen der Proband seiner gewohnten Tätigkeit nachgeht und der Forscher lediglich als Beobachter auftritt (Saraiya et al. 2006). Erstere ermöglichen das leichtere Aufdecken von Veränderungen, da die Aufgaben speziell hierfür gestaltet werden können. Letztere bieten ein ökologisch valideres Bild der Benutzung, sind also für derartige Forschungsfragen vorzuziehen. In direktem Zusammenhang hierzu stehen auch die verwendeten Daten – für geleitete Aufgaben werden zumeist **aufbereitete Daten** vom Forscher vorgegeben, welche wiederum die Analyse erleichtern. Ungleitete Aufgaben müssen schon vom Prinzip her auf **reale Daten** der Probanden zurückgreifen, da diese sonst nicht ihrer gewohnten Tätigkeit nachgehen können. Bezüglich der Umgebung halten wir die einfache Unterscheidung zwischen Feld und Labor für nicht mehr zeitgemäß. Moderne Techniken der Datenerhebung ermöglichen zum Beispiel auch im Feld ein viel höheres Maß der Kontrolle (z.B. Logging mit Bilddaten). Hierdurch ist der einfache Rückschluss: Feld = qualitativ & unkontrolliert und Labor = quantitativ & kontrolliert nicht mehr möglich. Wir

schlagen daher eine Unterscheidung nach dem Grad der Veränderung der Umgebung durch den Forscher vor. Hierbei lassen sich natürliche Umgebungen (kein Eingriff), beobachtbare Umgebungen (passive Teilnahme), kontrollierbare Umgebungen (der Forscher kann Variablen konstant halten) und manipulierbare Umgebungen (der Forscher kann Variablen aktiv verändern) unterscheiden. Gerade Längsschnittstudien ermöglichen es durch die Integration der Zeit-Variable eine viel natürlichere Umgebung zu schaffen und trotzdem, beispielsweise durch unaufdringliche Logging-Techniken ein gewisses Maß an Kontrolle beizubehalten.

2.6 Datenanalyse

Die Analyse von Daten in Längsschnittstudien ist insbesondere dann eine Herausforderung, wenn Veränderungen analysiert werden sollen. Im Falle von quantitativen Daten, beispielsweise bei der Analyse von Lerneffekten bei Eingabegeräten wird auf klassische Verfahren wie Repeated-Measures ANOVAs oder Helmert-Kontraste vertraut. Komplexere Verfahren, welche insbesondere die bei Längsschnittstudien verletzte Unabhängigkeit der Daten berücksichtigen, werden bislang in der MCI kaum eingesetzt. Zu nennen sei hier beispielsweise das von Hilbe & Hardin vorgestellte Verfahren, das „generalized estimation equation“ (GEE, in Menard 2007), welches versucht, genau dies zu berücksichtigen. Eine weitere Herausforderung stellt das Erkennen von zeitlichen Mustern dar, also beispielsweise immer wiederkehrende Interaktionsabläufe. Die von Noldus vertriebene Software „Theme“ bietet hier Möglichkeiten zur Analyse. Im Falle von qualitativen Daten lässt sich ebenfalls noch keine überzeugende Methodik erkennen, welche es ermöglicht, systematisch in diesen Veränderungen festzustellen. Dementsprechend wird die Thematik der Datenanalyse in den analysierten Beiträgen auch nur sehr oberflächlich betrachtet und wenig diskutiert. Dieser Stand zeigte sich auch bei der Diskussion innerhalb des Workshops „Best Practices in Longitudinal Research“ (Courage et al. 2009) auf der CHI Konferenz 2009. Letztlich muss festgehalten werden, dass in diesem Bereich noch deutlich mehr Forschung notwendig ist, um die für Längsschnittstudien in der MCI relevanten Methoden und Techniken der Datenanalyse, sowohl für quantitative als auch qualitative Daten, zu identifizieren und entsprechend anzupassen.

3 Implikationen für die MCI & Ausblick

Die Analyse der existierenden Studien zum Thema Längsschnittstudien hat gezeigt, dass bei aller Heterogenität dennoch eine gemeinsame Basis zu finden ist, wofür mit dieser Taxonomie der erste Grundstein gelegt wurde. Damit liefert sie mehrere Implikationen für die MCI auf zwei unterschiedlichen Ebenen: Der Ebene des methodisch-wissenschaftlichen Diskurses sowie der Ebene der praktischen Anwendung von Längsschnittstudien in Wissenschaft und Industrie.

Methodisch-wissenschaftlicher Diskurs: Eine einheitliche Sprachregelung ist für jeden Diskurs wesentlich. Diese ist bislang für Längsschnittstudien nicht gegeben. Beispielsweise existiert keine systematische Unterscheidung zwischen Untersuchungsdesign, Erhebungsdesign und Erhebungsmethode, was aber gerade bei Längsschnittstudien essentiell ist. Die

Taxonomie erlaubt es, existierende Studien auf ihr gewähltes Vorgehen hin zu analysieren und diese somit auch kritisch zu hinterfragen und alternative Möglichkeiten zu erkennen. Ebenfalls erleichtert sie den Vergleich verschiedener Ansätze bei ähnlichen Forschungsfragen. Letztlich liefert sie zudem eine Basis für die methodische Weiterentwicklung von Längsschnittstudien in der MCI, indem die Taxonomie selbst immer wieder hinterfragt und erweitert werden kann. Lässt sich eine Forschungsfrage beispielsweise bislang nicht klassifizieren, so regt dies auch an zu überprüfen, inwieweit hierfür vielleicht auch neue Untersuchungsdesigns oder Erhebungsmethoden notwendig sind. Andererseits kann die Entwicklung einer neuen Erhebungsmethode ebenso direkten Einfluss auf die Gestaltung und Durchführung einer Studie haben.

Praktische Anwendung in Industrie und Forschung: Bislang werden Längsschnittstudien in Textbüchern nur unzureichend bis gar nicht behandelt – oftmals werden auch lediglich Methoden wie Logging oder Tagebuchstudien aufgeführt (z.B. Interaction Design). Die Taxonomie erlaubt es, das weite Gebiet der Längsschnittstudien greifbarer zu machen. Hierbei dient die Taxonomie auch als Ideenpool bei der Planung einer Studie, indem die komplette Bandbreite der Möglichkeiten aufgezeigt wird. Zudem kann direkt kritisch hinterfragt werden, ob die konkrete Forschungsfrage überhaupt in das Raster von Längsschnittstudien passt oder ob nicht vielleicht doch eine klassischere Herangehensweise zu bevorzugen ist. Weiterhin kann sie aber auch sehr konkret dazu benutzt werden, sämtliche relevanten Aspekte einer derartigen Studie bei der Planung zu diskutieren und zu berücksichtigen. Hierzu gehören gerade auch Fragen wie „wo sollte die Studie stattfinden?“ und „welches Erhebungsdesign nehmen wir?“.

Für die weitere Forschung im Bereich Längsschnittstudien halten wir insbesondere die Aspekte Untersuchungsdesign und Erhebungsdesign für entscheidend. Je nach Fragestellung macht es zum Beispiel durchaus einen Unterschied, ob die Erhebungswellen im täglichen oder wöchentlichen Rhythmus stattfanden oder vielleicht innerhalb der Studie gar variabel gehalten wurden (z.B. Bieg 2008 vs. Castellucci & MacKenzie 2008). Diese Zusammenführung von konkreten Forschungsfragen und Untersuchungsdesign würde zudem auf lange Sicht zu einer Sammlung von bekannten Mustern führen, welche gerade dem Praktiker die Anwendung nochmals erleichtern würde. Darüberhinaus sollte die Rolle der Umgebung weiter untersucht werden. In vielen Fällen wäre es wünschenswert, wenn diese innerhalb einer Studie auch variabel gestaltet werden könnte. Hierdurch könnten die Vorteile aus unterschiedlichen Umgebungen kombiniert werden, also beispielsweise die Erhebung von standardisierten Leistungsmaßen im Labor und der ökologisch valideren Beobachtung in natürlicher Umgebung. Wie bereits aufgezeigt bedarf zudem das Thema der Datenanalyse der gesteigerten Aufmerksamkeit, wobei aus unserer Sicht insbesondere die Frage, wie in qualitativen Daten systematisch Veränderungen festgestellt werden können, hervorzuheben ist.

4 Literaturverzeichnis

Alexander, J., Cockburn, A. (2008). An Empirical Characterisation of Electronic Document Navigation. *Proceedings of the 2008 Conference on Graphics Interface 2008* p.123-130.

- Bergman, O., Beyth-Marom, R., Nachmias, R., Gradovitch, N., Whittaker, S. (2008). Improved search engines and navigation preference in personal information management. *ACM Transactions on Information Systems*, vol. 26, n. 4, p. 20.
- Bieg, H.-J. (2008). *Laserpointer and eye gaze interaction - design and evaluation*. Master-thesis, University of Konstanz, 2008. <http://hci.uni-konstanz.de/intehrdis/Bieg2008.pdf>
- Card, S. K., English, W. K., & Burr, B. J. (1978). Evaluation of Mouse, Rate-Controlled Isometric Joystick, Step Keys, and Text Keys for Text Selection on a CRT. *Ergonomics*, 21(8), 601-613.
- Castellucci, S. J., MacKenzie, I. S. (2008). Graffiti vs. Unistrokes: An Empirical Comparison. *In Proceedings of CHI 2008*. ACM.
- Courage, C., Jain, J., Rosenbaum, S. (2009). Best practices in longitudinal research. *In CHI 2009 extended abstracts*, ACM.
- Gerken, J., Demarmels, M., Dierdorf, S., Reiterer, H. (2008). Modellierungs- und Zoomtechniken für Punktdiagramme. *In Mensch & Computer 2008*, Oldenbourg Verlag.
- González, V. and Kobsa, A (2003). A workplace study of the adoption of information visualization systems. *In Proceedings of I-KNOW'03*, Graz, Austria, p92-102.
- Jain, J. (2008). Multimodal Capture of Consumer Intent in Retail. *In CHI 2008 extended abstracts*.
- Kjeldskov, J., Skov, M. B., Stage, J. (2005). Does time heal?: a longitudinal study of usability. *In Proceedings of the OZCHI*, ACM.
- Van der Kleij, R., Paashuis, R., Schraagen, J. M. (2005). On the passage of time: temporal differences in video-mediated and face-to-face interaction. *In Int. J. of Human-Computer Studies*, 62.
- MacKenzie, I. S. Zhang, S. X. (1999). The design and evaluation of a high-performance soft keyboard. *In CHI 1999*. ACM.
- McLachlan, P., Munzner, T., Koutsoufios, E., North, S. (2008). LiveRAC: Interactive Visual Exploration of System Management Time-Series Data. *In Proceedings of CHI 2008*. ACM.
- Menard, S. (2007). *Handbook of Longitudinal Research*, Academic Press.
- Mendoza, V.; Novick, D.G. (2005): *Usability over time*. Proceedings of SIGDOC2005, S.151-158.
- Rieman, J. (1993). The diary study: A workplace-oriented research tool to guide laboratory efforts. *INTERCHI 1993*, ACM.
- Saraiya, P., North, C., Lam, V., Duca, K. (2006). An Insight-Based Longitudinal Study of Visual Analytics. *IEEE Transactions on Visualization and Computer Graphics* 12, 6, p. 1511-1522.
- Sharp, H., Rogers, Y., Preece, J. (2007). *Interaction Design: Beyond Human-Computer Interaction*. Wiley and Sons.
- Shneiderman, B. and Plaisant, C. (2006). Strategies for evaluating information visualization tools: multi-dimensional in-depth long-term case studies. *AVI 2006 Workshop BELIV '06*. ACM.
- Taris, T. W. (2000). *A Primer in longitudinal data analysis*. London: SAGE Publications.
- Wilamowitz-Moellendorff, M. v., Hassenzahl, M., & Platz, A. (2007). Veränderung in der Wahrnehmung und Bewertung interaktiver Produkte. *Mensch & Computer 2007*. Oldenbourg Verlag.

Kontaktinformationen

Jens Gerken

Arbeitsgruppe Mensch-Computer Interaktion
Universität Konstanz
D-78457 Konstanz
Universitaetsstrasse 10, Box D-73
Tel.: +49 (0) 7531 88 3748
Fax: +49 (0) 7531 88 4772
E-Mail: jens.gerken@uni-konstanz.de

Harald Reiterer

Arbeitsgruppe Mensch-Computer Interaktion
Universität Konstanz
D-78457 Konstanz
Universitaetsstrasse 10, Box D-73
Tel.: +49 (0) 7531 88 3704
Fax: +49 (0) 7531 88 4772
E-Mail: harald.reiterer@uni-konstanz.de