

Touching Problems – Formative Evaluationen eines touristischen Informationssystems für ein neuartiges Elektrofahrzeug

Benjamin STRENGE, Ludger SCHMIDT

*Fachgebiet Mensch-Maschine-Systemtechnik, Universität Kassel
Mönchebergstraße 7, D-34125 Kassel*

Kurzfassung: Das neuartige Elektrofahrzeug „E2V“ soll von älteren und in ihrer Mobilität eingeschränkten Menschen ohne Führerschein verwendet werden können. Über ein integriertes Informationssystem sollen Nutzer ohne Computerkenntnisse u. a. touristische Informationen abrufen und Fahrrouten planen können. Dieser Beitrag beschreibt die systematische Auswahl, Kombination und Verwendung geeigneter Methoden zur nutzerorientierten Evaluation dieses Informationssystems. So wurden verschiedene leichtgewichtige Evaluationsmethoden eingesetzt und Nutzertests in einem Fahrsimulator durchgeführt, um Probleme bzgl. der Benutzbarkeit zu identifizieren. Die Ergebnisse reflektieren insbesondere spezielle Bedürfnisse der Nutzergruppe älterer Personen ohne Computerefahrung.

Schlüsselwörter: Mensch-Computer-Interaktion, Usability Engineering, Methodenauswahl, Evaluation

1. Einleitung und Projektkontext

Der im Jahr 2013 zum UNESCO-Welterbe erklärte Kasseler Bergpark Wilhelmshöhe stellt mit einer Fläche von mehr als 2,4 km² und über 240 m Höhenunterschied zwischen seinen bekanntesten Sehenswürdigkeiten, dem Schloss Wilhelmshöhe und dem Oktogon mit Herkules-Statue, erhebliche Ansprüche an die körperliche Leistungsfähigkeit seiner Besucher. Das an der Universität Kassel entwickelte, einachsige Elektrofahrzeug „E2V“ soll zukünftig auch älteren und körperlich eingeschränkten Menschen, die nicht über einen Führerschein verfügen, die problemlose Erkundung derartiger nicht-urbaner Gebiete ermöglichen. Dabei sollen auch Nutzer ohne Vorerfahrungen mit Computern touristische Informationen über die Historie und Öffnungszeiten der Sehenswürdigkeiten, Führungen, Gastronomie u. ä. abrufen sowie Routen planen können, die an ein Navigationssystem übergeben werden. Zu diesem Zweck verfügt das E2V über einen Touchscreen in der Mittelkonsole, mit dem Fahrer und Beifahrer das Informationssystem einhändig bedienen können.

Dieser Beitrag beschäftigt sich mit der Gestaltung und Evaluation dieses Informationssystems. Dabei wird einerseits auf die systematische, projektspezifische Auswahl, Kombination und Verwendung von Evaluationsmethoden eingegangen, andererseits auf besondere Erkenntnisse zur Optimierung des untersuchten Systems in Hinblick auf die speziellen Bedürfnisse der Nutzergruppe älterer und körperlich eingeschränkter Personen.

2. Methodenauswahl und -einsatz

Bei der Entwicklung des Informationssystems wurde ein benutzerorientierter Gestaltungsprozess in Anlehnung an DIN EN ISO 9241-210 verfolgt (vgl. Braun & Schmidt 2013). Dementsprechend wurden als Vorarbeiten der Interaktionsgestaltung zunächst der Nutzungskontext analysiert, Nutzeranforderungen identifiziert und die geplanten Interaktionsverläufe beschrieben. Um anschließend die in der jeweiligen Projektphase geeigneten Evaluationsmethoden zu bestimmen, wurde ein systematisches Verfahren auf Basis von ISO/TR 16982 verwendet, das Ansätze von Fischer et al. (2013) aufgreift. In ISO/TR 16982 sind insgesamt 18 unabhängige Einzelkriterien gelistet, die eine große Bandbreite an Faktoren abdecken, die für die Methodenauswahl relevant sind. Für jedes in einem Projektkontext zutreffende Kriterium ist eine Tendenz angegeben, ob dies für oder gegen den Einsatz bestimmter Methodentypen spricht bzw. ob diese Methoden überhaupt anwendbar sind. Die fünfstufige textuelle bzw. symbolische Ratingskala des ISO/TR von „Not applicable (NA)“ bis „Recommended (++)“ wurde zur Quantifizierung auf die Menge {„NA“; 0; 0,5; 0,75; 1} abgebildet. Dies ermöglicht eine teilweise (nicht binäre) Zustimmung zu einzelnen Auswahlkriterien und die Berechnung projektspezifischer Gesamtwertungen für die Methodentypen bzgl. der einzelnen Entwicklungsphasen aus dem Intervall [0, 1]. Das Werkzeug für diese Eingaben und Berechnungen wurde in Form eines Office-Open-XML-Tabellenkalkulationsdokuments implementiert (s. Abbildung 1).

	Observation of users	Performance-related measurements	Critical Incidents analysis	Questionnaires	Interviews	Thinking aloud	Collaborative design and evaluation	Creativity methods	Document-based methods	Model-based methods	Expert evaluation	Automated evaluation
Acquisition - Supply	0,87	0,60	-NA	0,75	0,71	0,60	0,88	0,69	0,87	0,46	0,79	0,50
Development - Requirements analysis	0,87	0,60	-NA	0,79	0,75	0,67	0,88	0,73	0,83	0,50	0,79	0,50
Development - Architectural design	0,83	0,63	-NA	0,75	0,71	0,67	0,88	0,77	0,87	0,50	0,79	0,54
Development - Qualification testing	0,83	0,63	-NA	0,79	0,75	0,63	0,88	0,69	0,83	0,50	0,79	0,54
Maintenance - Operation	0,83	0,60	-NA	0,75	0,71	0,60	0,88	0,69	0,79	0,46	0,79	0,50

Abbildung 1: Methodenauswahl auf Basis von ISO/TR 16982

Die Methodentypen konnten auf diese Weise schnell nach ihrer Brauchbarkeit sortiert und entsprechende Methoden ausgewählt werden. Die Methodenauswahl ist jedoch nicht vollständig durch die Berechnungen dieses Hilfsmittels determiniert: Wechselwirkungen und Abhängigkeiten zwischen verschiedenen konkreten Methoden müssen manuell von den verantwortlichen Usability-Ingenieuren berücksichtigt werden. So kann auf pragmatische Weise sichergestellt werden, dass trotz begrenzter Projektressourcen alle relevanten Facetten des komplexen Konstrukts der Gebrauchstauglichkeit angemessen berücksichtigt werden.

2.1 Anforderungsanalysephase

Als finale Aktivität innerhalb der Anforderungsanalysephase wurden Papierprototypen genutzt, um die Nutzeranforderungen bzgl. der Benutzungsschnittstelle zu evaluieren und konkretisieren. Die am höchsten bewerteten Methodentypen für diese Phase waren Nutzerbeobachtung (0,87), kollaborative Gestaltung und Evaluation (0,88) sowie dokumentenbasierte Methoden (0,83). Bei Evaluationen mit direkter Nutzereinbeziehung bietet es sich darüber hinaus an,

ergänzend entweder Leistungsmessungen (0,6) vorzunehmen oder die Methode Thinking Aloud (0,67) zu verwenden. Dem wurde folgendermaßen Rechnung getragen: Zunächst wurden verschiedene Szenarien definiert, die typische Aufgaben umfassen. Um den Aufwand in dieser Phase überschaubar zu halten, wurde nicht auf Testteilnehmer aus der primären Zielgruppe zurückgegriffen. Um dies teilweise zu kompensieren, wurden in den Testszenarien nicht nur die jeweiligen Aufgaben beschrieben, sondern explizit auch sensorische Einschränkungen und Vorerfahrungen bzgl. der Fahrzeugführung vorgegeben, sodass studentische Testteilnehmer mental diese Rolle einnehmen konnten (dokumentenbasierte Methode). Während des Tests wurden Nutzeraktionen und auftretende Probleme protokolliert (Nutzerbeobachtung) und die Teilnehmer sollten ihre Gedanken laut aussprechen, um Einsicht in die Ursachen der Interaktionsprobleme zu gewinnen (Thinking Aloud). Darüber hinaus wurden Rückmeldungen und Hinweise der Testteilnehmer bezüglich benötigter Funktionen und der Interaktionsgestaltung erfasst und in die Anforderungsliste überführt (kollaborative Gestaltung und Evaluation).

2.2 Gestaltungsphase

Auf Basis der Anforderungsanalyse und der Ergebnisse der Paper-Prototyping-Tests wurde ein Medium-Fidelity-Prototyp des Informationssystems in Form eines Storyboards aus pixelgenauen Darstellungen der Benutzungsschnittstelle erstellt. Zugunsten der Agilität des Prozesses wurde in dieser Phase auf Nutzerbeobachtungen (0,83) und kollaborative Gestaltung und Evaluation (0,88) verzichtet. Stattdessen wurde eine Kombination aus dokumentenbasierten Methoden (0,87) und Experten-Evaluation (0,79) verwendet. Die konkreten Methoden, Cognitive Walkthrough auf Basis einer primären Persona und Heuristische Evaluation, ermöglichten eine Analyse hinsichtlich leichter Erlernbarkeit (Wharton et al. 1994) und sowohl allgemeiner (Nielsen 2005) als auch Kiosksystem-spezifischer (Sandnes et al. 2010) Usability-Heuristiken. Nach einer ersten Überarbeitung anhand der Ergebnisse dieser Methoden wurden die geänderten Entwürfe jeweils iterativ durch informelle Experten-Reviews bewertet und ggf. direkt überarbeitet. Dieses Vorgehen kann als eine leichtgewichtige Adaption der RITE-Methode nach Medlock et al. (2002) betrachtet werden. Der ursprüngliche Entwurf der Benutzungsschnittstelle wurde dabei mehrfach fundamental verändert, bis die beteiligten Experten mit dem Resultat zufrieden waren.

2.3 Testphase

Die Entwürfe dienten im nächsten Schritt als Detailspezifikation für die Implementierung der Software. Um die Gebrauchstauglichkeit des Informationssystems schon vor der Integration des Gesamtfahrzeugs mit Vertretern der anvisierten Nutzergruppe und in Annäherung an den realen Nutzungskontext untersuchen und optimieren zu können, wurde anschließend ein teilfunktionaler Laborprototyp in Form eines stationären Fahrsimulators aufgebaut. Neben dem Fahrersitz und einer Mittelkonsole mit Touchscreens für das Informations- und Navigationssystem umfasst dieser auch einen Side Stick zur Steuerung des Fahrzeugs in einer speziell für diesen Zweck erstellten 3D-Simulationsumgebung des Bergparks Wilhelmshöhe. Unter Verwendung dieses Simulators wurden Nutzertests zur formativen Evaluation der Gebrauchstauglichkeit des Informationssystems

durchgeführt, an denen Vertreter der Zielgruppe älterer oder körperlich eingeschränkter Menschen teilgenommen haben (N=20). Der primäre Fokus der Untersuchungen lag darauf, Probleme hinsichtlich leichter Erlernbarkeit und Benutzbarkeit des Systems zu identifizieren. Die Nutzertests umfassten die Methodentypen Nutzerbeobachtung (0,83), Interviews (0,75) in opportunistischer, unstrukturierter Form sowie Fragebogen (0,79) zur Gebrauchstauglichkeit. Bezüglich letzteren fiel die Wahl auf den System Usability Scale (SUS; Brooke 1996), da dieser vergleichsweise kurz ist und bereits bei geringem Stichprobenumfang rasch konvergiert (vgl. Tullis & Stetson 2004). Ergänzend wurde anstelle von objektiven Leistungsmessungen (0,63) erneut Thinking Aloud (0,63) verwendet, da die hierdurch gewonnen Einsichten in kausale Zusammenhänge bzgl. der Nutzungsschwierigkeiten als wichtiger eingeschätzt wurden. Da das System für die Freizeitnutzung ausgelegt ist, war die objektive Effizienz der Benutzung ohnehin von eher nachrangiger Bedeutung.

3. Resultate

Die Evaluationsergebnisse spiegeln einige sehr spezielle Bedürfnisse und Anforderungen der primären Zielgruppe wider. Dies kam bereits bei den Expertenbasierten Evaluationen während der Gestaltungsphase zum Tragen. So sah der ursprüngliche Entwurf vor, dass touristische Informationen ähnlich wie in einer Online-Enzyklopädie auf einer einzelnen scrollbaren Seite mit entsprechenden Hyperlinks präsentiert werden sollten (vgl. Abbildung 2 links). Dieser Ansatz wurde aufgrund negativer Expertenbewertungen verworfen, da Vorerfahrungen aus anderen Projekten gezeigt hatten, dass ältere Nutzer oft Schwierigkeiten mit dieser Art der Navigation haben.



Abbildung 2: Erster Entwurf (links) und finale Version (rechts) der Informationsseite zu einer Sehenswürdigkeit im touristischen Informationssystem

Weitere interessante Resultate lieferten die Nutzertests. Wie links auf Abbildung 3 dargestellt, zeigte das System an einer Stelle einen Benutzungshinweis in Form eines „Popup“-Fensters an, das durch Berühren einer Schaltfläche mit der Beschriftung „Hinweis schließen“ ausgeblendet werden musste, um fortzufahren. Dies wurde von vier der fünf ältesten Testteilnehmer (67 bis 78 Jahre) nicht verstanden und stattdessen versucht, die ausgegrauten Schaltflächen im Hintergrund zu verwenden. Vermutlich ist dies darauf zurückzuführen, dass das von Desktop-Betriebssystemen etablierte Konzept, dass das „vorderste“ von mehreren

überlagerten Fenstern den Fokus hat und Eingaben entgegen nimmt, diesen Personen unbekannt war; jüngere Personen hatten an dieser Stelle keine Schwierigkeiten. Allgemein war die Bedeutung ausgegrauter Schaltflächen einigen Testnutzern offenbar unbekannt, da mehrfach versucht wurde, diese zu verwenden. Es erscheint daher zumindest für diese Zielgruppe empfehlenswert, aktuell nicht benutzbare Schaltflächen nicht nur auszugrauen, sondern komplett auszublenden.



Abbildung 3: Hinweis-„Popup“ (links) und Fahrzielauswahl (rechts)

Ähnliches gilt für die rechts auf Abbildung 3 dargestellte Auswahl von Fahrzielen. Die zunächst dunkelblau dargestellten Fahrziele wurden vom System hellblau hervorgehoben, wenn sie vom Nutzer ausgewählt wurden. Für einige Testnutzer war dieses visuelle Feedback offenbar nicht kontrastreich genug, sodass diese schlecht erkennen konnten, welche Fahrziele bereits ausgewählt wurden. Durchweg positiv wurde von den Testnutzern dagegen die Funktion rezipiert, sich die auf dem Bildschirm erscheinenden Hinweistexte und touristische Informationen automatisch mittels synthetischer Sprachausgabe vorlesen zu lassen.

Ergänzend zur Erhebung dieser qualitativen Ergebnisse wurde mithilfe des SUS-Fragebogens die subjektive Gebrauchstauglichkeit des Systems gemessen. Das arithmetische Mittel des SUS-Gesamtwerts aus N=15 verwertbaren Datensätzen lag bei 61,5 Punkten (min=32,5, max=90) mit einer Standardabweichung von ≈ 18 . Nach Bangor et al. (2009) entspricht dies einem Wert zwischen „OK“ (50,9) und „Good“ (71,4). Interessanterweise lagen die Werte des SUS-Teilfaktors „Learnability“ (nach Lewis & Sauro 2009) mit durchschnittlich 48,3 Punkten signifikant ($p < 0,01$) unter den Werten des „Usability“-Teilfaktors von durchschnittlich 64,8 Punkten, obwohl bei der Systemgestaltung gerade dem Aspekt der Erlernbarkeit durch die Auswahl entsprechender Evaluationsmethoden besondere Beachtung geschenkt wurde. Dies unterstreicht die Annahme, dass aufgrund mangelnder Vorerfahrungen vieler Personen aus der primären Zielgruppe ein besonders weitreichender Einführungs- und Unterstützungsbedarf besteht. Bei der Systemgestaltung wurde diesem Aspekt Rechnung getragen, indem ein umfangreiches optionales Einführungsvideo integriert wurde und über das System bei Bedarf jederzeit Kontakt mit einem Mitarbeiter des Fuhrparkbetreibers aufgenommen werden konnte. Letzteres wurde in den Nutzertests bei Bedarf durch Mitarbeiter der Versuchsleitung simuliert, indem diese dann durch eine Trennwand „blind“ mit den Testpersonen kommuniziert haben. Die Bewertungen der entsprechenden SUS-Items („Ich musste eine Menge Dinge lernen, bevor ich mit dem System arbeiten werde.“ und „Ich denke, dass ich technischen Support brauchen würde, um das System zu nutzen.“) könnten demnach als Indiz

interpretiert werden, dass viele Testnutzer das Gefühl hatten, diese vom System angebotenen Unterstützungsfunktionen seien für die erfolgreiche Nutzung obligatorisch gewesen.

4. Fazit

Das verwendete Verfahren auf Basis von ISO/TR 16982 hat die Auswahl der Evaluationsmethoden erleichtert, erfordert jedoch Expertise bezüglich der Wechselwirkungen mehrerer Methoden. Die Ergebnisse der Evaluation des touristischen Informationssystems legen nahe, dass Experten-basierte Methoden insbesondere bei einer Zielgruppe aus überwiegend älteren Nutzern mit geringer Computererfahrung „echte“ Nutzertests keinesfalls vollständig ersetzen können. Die ermittelten Werte des SUS-Teilfaktors Erlernbarkeit werfen die Frage auf, ob diese Skala bei sehr unerfahrenen Nutzern tatsächlich eine valide Einschätzung der diesbezüglichen Qualität eines Systems erlaubt.

5. Literatur

- Bangor A, Kortum P, Miller J (2009) Determining what individual SUS scores mean: Adding an adjective rating scale. *Journal of usability studies* 4(3): 114-123.
- Braun R, Schmidt L (2013) Benutzerorientierte Entwicklung eines für spezifische Nutzergruppen adaptierbaren, einachsigen, teilautonomen Elektrofahrzeugs. In: Verband der Elektrotechnik Elektronik Informationstechnik e. V. (Hrsg.): *Lebensqualität im Wandel von Demografie und Technik: 6. Deutscher AAL-Kongress (Berlin 2013)*. Berlin: VDE-Verlag, 321–325.
- Brooke J (1996) SUS-A quick and dirty usability scale. In: *Usability evaluation in industry*, 189-194.
- DIN EN ISO 9241-210 (2011) *Ergonomie der Mensch-System-Interaktion - Teil 210: Prozess zur Gestaltung gebrauchstauglicher interaktiver Systeme (ISO 9241-210:2010)*.
- Fischer H, Strenge B, Nebe K (2013) Towards a Holistic Tool for the Selection and Validation of Usability Method Sets Supporting Human-Centered Design. In: Marcus A (Hrsg.): *Design, User Experience, and Usability. Design Philosophy, Methods, and Tools*. Springer 2013 Lecture Notes in Computer Science Bd. 8012, 252–261.
- ISO/TR 16982 (2002) *Ergonomics of human-system interaction - Usability methods supporting human-centred design*.
- Lewis J, Sauro J (2009) The Factor Structure of the System Usability Scale. In: Kurosu M (Hrsg.) *Human Centered Design*. Springer 2009 Lecture Notes in Computer Science Bd. 5619, 94–103.
- Medlock MC, Wixon D, Terrano M, Romero R, Fulton B (2002) Using the RITE method to improve products: A definition and a case study. Usability Professionals Association.
- Nielsen J (2005) 10 Usability Heuristics for User Interface Design. Accessed December 09, 2014. <http://www.nngroup.com/articles/ten-usability-heuristics> - ISSN 1548-5552.
- Sandnes FE, Jian HL, Huang YP, Huang YM (2010). User interface design for public kiosks: an evaluation of the Taiwan high speed rail ticket vending machine. *Journal of information science and engineering* 26(1): 307-321.
- Tullis TS, Stetson JN (2004) A comparison of questionnaires for assessing website usability. Usability Professional Association Conference.
- Wharton C, Rieman J, Lewis C, Polson P (1994) The Cognitive Walkthrough Method: A Practitioner's Guide. In: Nielsen J, Mack RL (Hrsg.) *Usability Inspection Methods*. New York: John Wiley & Sons, 105–140.