

Benutzerstudien zur Bewertung multimodaler, interaktiver Anzeigetafeln in unterschiedlichen Entwicklungsstufen

Stefan Schaffer¹, Julia Seebode¹, Ina Wechsung², Florian Metze³, Sebastian Möller²

¹Graduiertenkollegs prometei, TU Berlin, 10589 Berlin

²Deutsche Telekom Laboratories, TU Berlin, 10589 Berlin

³interACT center, Carnegie Mellon University, Pittsburgh, PA

stefan.schaffer@zmms.tu-berlin.de

Abstract: Gegenstand des Beitrages sind zwei Evaluationsstudien einer interaktiven, digitalen Anzeigetafel. Das ursprünglich unimodale, Touchscreen-basierte System wurde schrittweise erweitert. Dabei wurde zunächst eine Sprachsteuerung implementiert, in einer zweiten Entwicklungsstufe wurde anschließend eine Gesichtserkennung ergänzt. Nach jeder Systemmodifikation erfolgte eine Benutzerstudie zur Bewertung der vom Nutzer wahrgenommenen Qualität der Interaktion mittels subjektiver, direkter Fragebogendaten und technischer, indirekter Performanzdaten. Die Ergebnisse der Fragebogendaten zeigen, dass die multimodalen Systemvarianten hinsichtlich der hedonischen Qualitäten besser beurteilt wurden als die unimodale Variante. Dies war auch dann zu beobachten, wenn sich die Bewertungen bezüglich der Usability nicht erhöhten. Die Analyse der indirekten Daten zeigte eine Verbesserung der Usability, die sich jedoch nicht in den Nutzerurteilen widerspiegelte.

1 Einleitung

Immer häufiger sind interaktive Anzeigetafeln in öffentlichen Bereichen und im Umfeld des Arbeitsplatzes anzutreffen. Ein typischer Anwendungsfall ist die Darstellung digitaler Daten in gemeinschaftlichen Bereichen, wie Bürofluren oder Foyers (Hu06). Dabei können unterschiedlichste Mediendaten aus Datenbanken geladen und auf dem Display angezeigt werden (GR01). Insbesondere die Möglichkeit mittels digitaler, interaktiver Anzeigetafeln Informationen in Echtzeit und mit wenig Aufwand zu aktualisieren zeigt sich dabei von Vorteil (TGS06). Schon seit einigen Jahren werden interaktive Displays mit neuen Formen der Mensch-Maschine-Interaktion, wie z.B. die Steuerung durch Berührung, Gesten oder Sprache kombiniert (Ov96). Durch den Einsatz dieser neuen Interaktionsmodalitäten wird die Nutzung mehrerer Kommunikationskanäle ermöglicht. Da auch zwischenmenschliche Kommunikation für gewöhnlich multimodal abläuft, wird dementsprechend angenommen, dass der Einsatz von Multimodalität die Systeminteraktion intuitiver gestaltet (Hö05). Sprache ist für den Menschen die natürlichste Form der Kommunikation (Fe03). Folglich ist es wünschenswert, Sprache auch zur Steuerung von interaktiven Anzeigetafeln effizient einzusetzen (WWL00).

In der vorliegenden Studie wurden dementsprechend Benutzerstudien zur Bewertung der wahrgenommenen Qualität der Interaktion mit einer interaktiven Anzeigetafel in drei unterschiedlichen Entwicklungsstufen verglichen. Das an einer Wand montierte Display unterstützt den Nutzer bei der Orientierung innerhalb einer Büroumgebung und hilft Mitarbeitern und Besuchern Personen und Besprechungsräume zu finden. In Stufe eins verfügte das System nur über eine unimodale Steuerung per Touchscreen. In Stufe zwei wurde das System um eine Sprachsteuerung erweitert. In Stufe drei wurde zusätzlich eine Kamera zur kontrollierten Aktivierung der Spracherkennung integriert.

2 Experiment 1

Ein ursprünglich unimodales, mit Touchscreen zu bedienendes System wurde zunächst um eine Sprachsteuerung erweitert. Anschließend erfolgte eine erste Evaluation, im Zuge derer die unimodale Systemvariante mit der multimodalen Systemversion (Multimodal 1) verglichen wurden.

2.1 Methode

Teilnehmer und System: An der ersten Evaluation nahmen insgesamt 36 deutschsprachige Personen (17 männlich; 19 weiblich) im Alter zwischen 21 und 39 Jahren teil. Keiner der Teilnehmer besaß Vorerfahrung mit dem getesteten System. Evaluiert wurde ein Raummanagement- und Informationssystem, das die Möglichkeit bietet, Veranstaltungen, Räume und Mitarbeiter zu suchen, deren Ort anzuzeigen und einzelne Räume zu buchen. Ursprünglich konnte das System unimodal über eine grafische Nutzerfläche mit Touchinput gesteuert werden. Im Experiment wurde ebenfalls eine unimodale Systemversion mit Sprachinput sowie eine multimodale Version mit der Möglichkeit der Kombination von beiden Eingabemodalitäten getestet.

Ablauf: Die Probanden bearbeiteten während des Versuchs 6 verschiedene Aufgaben mit dem System, darunter Navigation, Suchen und Anzeigen von Räumen, Anzeigen von Veranstaltungen und Suchen von Mitarbeitern. Dabei wurden die Probanden gebeten die 6 Aufgaben zunächst per Touchscreen zu bearbeiten und danach die Interaktion mit dem System zu bewerten.

Anschließend wurden die Aufgaben erneut präsentiert. Nun konnten die Probanden für jede Aufgabe die Modalität zur Bearbeitung frei wählen, dabei war auch eine Kombination der Modalitäten (Sprache und Touch) möglich. Abschließend erfolgte wieder eine Bewertung der Interaktion mit dem System im letzten Versuchsblock.

Erhobene Daten: Zur Erfassung der subjektiven Bewertungen der drei verschiedenen Systemversionen wurde der AttrakDiff-Fragebogen [HBK03] verwendet. Dieser Fragebogen umfasst 28 Items auf 4 Skalen. Die Skalen sind: *Pragmatische Qualität* als ein klassisches Maß der Usability, *Attraktivität* als globales Gesamturteil sowie *Hedonische Qualität - Stimulation* und *Hedonische Qualität - Identität* zur Messung hedonischer Qualitäten.

Zusätzlich wurden zur Messung der Güte des verwendeten Spracherkenners die Konzeptfehlerrate und die Rate der Äußerungen die nicht von der Grammatik abgedeckt werden (Out-of-Grammar-Rate) bestimmt.

2.2 Ergebnisse

Fragebogendaten: Auf der AttrakDiff Skala *Hedonische Qualität-Stimulation* wurde die multimodale Systemversion besser beurteilt als die Version mit Touchscreen. Auf den anderen Skalen zeigten sich keine signifikanten Veränderungen (s. Tabelle 1).

Performanzdaten: Für den Spracherkennung wurden eine Konzeptfehlerate von ca. 67 Prozent und eine Out-of-Grammar-Rate von 58 Prozent ermittelt.

Skala	System	MW	SD	t (df)	p
Pragmatische Qualität	Touch	1.20	0.75	1.96 (68)	.054
	Multimodal 1	0.76	1.10		
Hedonische Qualität - Stimulation	Touch	0.54	0.93	3.21 (69)	.002
	Multimodal 1	1.19	0.79		
Hedonische Qualität - Identität	Touch	0.82	0.70	0.46 (69)	.963
	Multimodal 1	0.83	0.91		
Attraktivität	Touch	1.16	0.80	1.05 (69)	.297
	Multimodal 1	0.93	0.97		

Tabelle 1. Ergebnisse der AttrakDiff-Skalen (MW = Mittelwert, SD = Standardabweichung)

2.3 Diskussion

Die Benutzer bewerteten auf der AttrakDiff Skala *Hedonische Qualität-Stimulation* die multimodale Systemversion besser, obwohl sich die *Pragmatische Qualität* (Usability) nicht verbesserte.

Ein Grund hierfür könnte sein, dass die Nutzer die neue Interaktionsmöglichkeit per Sprache prinzipiell als Bereicherung ansehen, gleichzeitig jedoch Konzeptfehlerrate und Out-of-Grammar-Rate sehr hoch waren. Dies kann sich negativ auf die Usability auswirken, da aufgrund der Erkennungsfehler häufig nicht die vom Nutzer erwünschte Systemreaktion erfolgt. Viele Erkennungsfehler traten auf, wenn im Bereich des Mikrofons von einer Person akustische Ereignisse ausgelöst wurden, die nicht an das System gerichtet sind. Eine Möglichkeit mit dieser Situation umzugehen ist die Integration einer Gesichtserkennung, welche die Spracherkennung nur aktiviert, wenn der Anzeige ein Gesicht zugewandt ist. Diese und weitere Veränderungen des Systems sollen daher im zweiten Experiment getestet werden.

3 Experiment 2

Im zweiten Entwicklungsschritt wurde die Sprachsteuerung mit Hilfe der Ergebnisse aus der ersten Evaluation verbessert und das System um eine Komponente zur Gesichtserkennung erweitert. Diese dient der Aktivierung der Spracherkennung und zwar nur dann, wenn sich ein Gesicht im Kamerabild befindet. Auch nach dieser Weiterentwicklung erfolgte eine Evaluation des Systems.

3.1 Methode

Teilnehmer und System: Insgesamt nahmen an der zweiten Evaluation 31 ebenfalls deutsche Muttersprachler (16 männlich; 17 weiblich) im Alter von 22 bis 29 Jahren teil. Von diesen Probanden hatten 14 schon an der ersten Evaluation teilgenommen und daher Erfahrung mit dem System, wohingegen die übrigen 17 Probanden keinerlei Vorerfahrung mit dem getesteten System besaßen. Getestet wurde wieder das Raummanagement- und Informationssystem mit multimodaler Eingabemöglichkeit und der Erweiterung um die Gesichtserkennung.

Ablauf: Mit allen Probanden wurde ein Versuchsblock durchgeführt, in dem dieselben Aufgaben, wie in der ersten Evaluation mit dem erweiterten, multimodalen System bearbeitet wurden. Danach bewerteten die Probanden wieder die Interaktion.

Erhobene Daten: Auch in diesem Experiment wurde der AttrakDiff-Fragebogen verwendet, um die subjektive Bewertung des erweiterten Systems zu erfassen. Zusätzlich wurden wieder Konzeptfehlerrate und Out-of-Grammar-Rate bestimmt.

3.2 Ergebnisse

Fragebogendaten: Die Ergebnisse der nach dem Test erhobenen subjektiven Beurteilungen, zeigen Unterschiede auf den Skalen *Pragmatische Qualität* und *Hedonische Qualität-Stimulation*. Post-hoc Einzelvergleiche zeigten dabei im Vergleich zwischen erstem multimodalen System und zweitem multimodalen System keine Verbesserung.

Im Vergleich zum unimodalen System war für die Skala *Pragmatische Qualität* eine Verschlechterung zu beobachten, für die Skala *Hedonische Qualität-Stimulation* dagegen eine Steigerung (s. Tabelle 2).

Performanzdaten: Die Analyse der Performanzdaten ergab eine Verbesserung beider Parameter. Sowohl die Out-of-Grammar-Rate ($MW_{Multimodal1} = 58,00$; $SD_{Multimodal1} = 12,17$ $MW_{Multimodal2} = 45,53$; $SD_{Multimodal2} = 21,20$; $t(df) = 4,92$; $p = .000$) als auch die Konzeptfehlerrate ($MW_{Multimodal1} = 66,75$; $SD_{Multimodal1} = 25,18$; $MW_{Multimodal2} = 33,07$; $SD_{Multimodal2} = 15,90$; $t(29) = 2,53$; $p = .017$) war nach der Erweiterung deutlich geringer.

Skala	System	MW	SD	F(df)	p
Pragmatische Qualität	Touch	1.20	0.75	3.77 (2, 98)	.026
	Multimodal 1	0.76	1.10		
	Multimodal 2	0.58	1.01		
Hedonische Qualität - Stimulation	Touch	0.54	0.93	6.75 (2, 99)	.002
	Multimodal 1	1.19	0.79		
	Multimodal 2	1.11	0.68		
Hedonische Qualität - Identität	Touch	0.82	0.70	.09 (2, 99)	.910
	Multimodal 1	0.83	0.91		
	Multimodal 2	0.76	0.57		
Attraktivität	Touch	1.16	0.80	.62 (2, 99)	.538
	Multimodal 1	0.93	0.97		
	Multimodal 2	1.00	0.76		

Tabelle 2. Ergebnisse der AttrakDiff-Skalen für alle Systeme. (MW = Mittelwert, SD = Standardabweichung)

3.3 Diskussion

Die Analyse der indirekten Daten zeigte eine Verbesserung der Usability, die sich jedoch nicht in den Nutzerurteilen widerspiegelte. Ein möglicher Grund könnte die immer noch hohe Konzeptfehlerrate sein, da aufgrund von Fehlern der Spracherkennung häufig Systemreaktionen erfolgten, die nicht vom Nutzer beabsichtigt waren. Die Veränderungen des Systems und daraus folgende Verringerung der Fehlerraten reichten augenscheinlich nicht aus, um die subjektiv wahrgenommene Usability (Skala: *Pragmatische Qualität*) des Systems zu erhöhen.

Zudem ist fraglich, ob die Bewertungen problemlos vergleichbar sind, da die Probanden in der ersten Evaluation das System als letztes beurteilten, also Erfahrungen mit der unimodalen Systemversion hatten, wohingegen sie in der zweiten Evaluation ohne vorherige Durchläufe das multimodale System testeten. Diese Reihenfolge könnte die Bewertungen beeinflusst haben.

4 Abschlussdiskussion und Schlussfolgerungen

Es zeigte sich, dass Multimodalität zu einer Verbesserung der subjektiv wahrgenommenen hedonischen Qualität führt, auch wenn die subjektiv wahrgenommene Usability nicht steigt, oder sich sogar verschlechtert. So wurden beide multimodale Systeme auf der Skala *Hedonische Qualität - Stimulation* deutlich besser bewertet als das unimodale System. Gleichzeitig gab es jedoch keine Verbesserungen der wahrgenommenen Usability (Skala: *Pragmatische Qualitäten*) und des Gesamturteils.

Der Vergleich beider multimodaler Systeme ergab für die ermittelten Performanzdaten (Out-of-Grammar-Rate und Konzeptfehlerrate) eine Verbesserung vom ersten zum zweiten multimodalen System. Dies spricht zwar für eine höhere Usability des zweiten multimodalen Systems, doch findet sich in den subjektiven Bewertungen (Pragmatische Qualität) keine Bestätigung: Beide System wurde als gleich gut beurteilt.

Insgesamt deuten die Ergebnisse darauf hin, dass ein großer Vorteil multimodaler Systeme in der Verbesserung der hedonischen Aspekte, insbesondere der Fähigkeit zur Stimulation, liegt. Doch ist zu erwarten, dass Nutzer mit einer weiteren Verbreitung multimodaler Systeme vertrauter im Umgang mit diesen werden. Dadurch verliert die Multimodalität den Aspekt des Neuen und Aufregenden, was das Potential der Stimulation verringern könnte. Um die Änderung der subjektiven Nutzerbewertungen im Laufe der Zeit beobachten zu können, ist eine Folgestudie zur Untersuchung der Langzeitnutzung eines solchen multimodalen Systems notwendig. Hinsichtlich der Usability zeigten sich dagegen kaum Verbesserungen, die mit Multimodalität assoziierten Vorteile ließen sich für dieses System nicht bestätigen.

Literaturverzeichnis

- [Fe03] Fellbaum, K.R. Speech Input and Output Technology - State of the Art and Selected Applications. In A. Düsterhöft, & B. Thalheim (Hrsg.): Natural Language Processing and Information Systems, 8th International Conference on Applications of Natural Language to Information Systems, 2003, S. 7-13.
- [GR01] Greenberg, S.; Rounding, M. The notification collage: posting information to public and personal displays. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, New York, 2001, S. 514-521.
- [HBK03] Hassenzahl, M.; Burmester, M.; Koller F..AttrakDiff: Ein Fragebogen zur Messung wahrgenommener hedonischer und pragmatischer Qualität. In J. Ziegler.; G. Szwillus (Hrsg.) Mensch & Computer 2003. Interaktion in Bewegung, B.G. Teubner: Stuttgart, 2003. S. 187-196.
- [Hu06] Huang, E. M. The Design and Analysis of Large Display Groupware Applications, Dissertation, Georgia Institute of Technology, 2006.
- [Hö05] Höysniemi, J.; Hämäläinen, P.; Turkki, L.; Rouvi, T. Children's intuitive gestures in vision-based action games. *Communications of the ACM*, 48 (1), 2005, S.45-50
- [Ov96] Oviatt, S.. Multimodal interfaces for dynamic interactive maps. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems: Common Ground* ACM: New York, 1996, S. 95-102.
- [TGS06] Tse, E., Greenberg, S., and Shen, C. 2006. GSI demo: multiuser gesture/speech interaction over digital tables by wrapping single user applications. In *Proceedings of the 8th international Conference on Multimodal interfaces*. ACM: New York, 2006, S. 76-83.
- [WWL00] Walker, Marilyn ; Wright, Jerry ; Langkilde, Irene: Using Natural Language Processing and Discourse Features to Identify Understanding Errors in a Spoken Dialogue System. *Proceedings of the 17th International Conf. on Machine Learning*, Morgan Kaufmann: San Francisco, 2000, S. 1111 – 1118.