

Multimodale Nutzerschnittstellen-Gestaltung: Wie wirkt sich linguistische Informationsredundanz auf die Gebrauchstauglichkeit aus?

Timo Sackmann
Universität Osnabrueck/
Fraunhofer IAO
49074 Osnabrück
timo.sackmann@gmail.com

Matthias Peissner
Fraunhofer IAO
Nobelstr. 12
70569 Stuttgart
matthias.peissner@iao.fraunhofer.de

Abstract

Die Kombination einer Sprach-Nutzerschnittstelle mit einer grafischen Nutzerschnittstelle, erlaubt eine Vielzahl von Darstellungsvarianten. Bisher gibt es wenige Gestaltungsrichtlinien, die auf empirischen Daten beruhen. Der vorliegende Beitrag beschäftigt sich mit der Auswirkung von linguistischer Informationsredundanz auf die

Gebrauchstauglichkeit einer Anwendung anhand von empirischen Ergebnissen. Es wurden drei mögliche Designvarianten als Prototyp implementiert und Nutzertests unterzogen. Auf deskriptiver Ebene führte ein mittlerer Redundanzgrad zu den besten Ergebnissen.

Keywords

information redundancy; cognitive load, multimodal interface, user experience

1.0 Einleitung

Multimodale Nutzerschnittstellen (User-Interfaces) kombinieren unterschiedliche Ein- und Ausgabemöglichkeiten, wie zum Beispiel Spracheingabe mit einem Touchbildschirm. Multimodale Nutzerschnittstellen ermöglichen (im Vergleich zu unimodalen) eine verbesserte Anpassung an die sensorischen bzw. kognitiven Fähigkeiten des Menschen und die unterschiedlichen Aufgaben der Mensch-Maschine Interaktion (Oviatt, 2002; Dybkjaer, 2004; van Erp et al. 2006). In Zukunft werden Mensch-Computer Schnittstellen vermehrt unterschiedliche Modalitäten zur Interaktion nutzen, die an das jeweilige Teilziel einer Aufgabe angepasst sind. (Oviatt, 2002; Sutcliffe 2002; van Erp et al., 2006). Einige Studien berichten, dass multimodale Benutzerschnittstellen für spezielle Aufgaben zu höherer Zufriedenheit der Nutzer, als auch zu höherer Effizienz gegenüber unimodalen Benutzer-Schnittstellen führen (z.B. Hauptmann 1989, Oviatt 1997; Lamel et al. 2002).

2.0 Theorie

2.1 Redundante oder komplementäre Informationsdarstellung in multimodalen Interfaces

Eine multimodale Nutzerschnittstelle stellt besondere Anforderungen an die Informationsgestaltung. Es ergeben sich viele unterschiedliche Kombinationsmöglichkeiten, wobei die Entscheidung für die Darstellung einer Information mittels einer Modalität durch physikalische und psychologische Aspekte bestimmt ist. Dazu zählen die Eigenschaften der Modalitäten des Nutzers und der Umgebung (Dybkjaer et al, 2004). Im Besonderen muss zwischen den Möglichkeiten die Informationen in den einzelnen Modalitäten hoch redundant oder eher komplementär (sich ergänzend) darzustellen, entschieden werden. Beispielsweise kann bei einer Nutzerschnittstelle, die visuelle und sprachliche Ausgabe kombiniert, die Aufforderung „Bitte geben Sie das Passwort ein.“ visuell als Text auf einem Bildschirm erscheinen und re-

dundant auditiv (Sprachausgabe) ausgegeben werden. Ein anderer Ansatz ist die über verschiedene Modalitäten komplementäre (complementary) bzw. verteilte (distributed) Darstellung. Hier ergänzen sich die Modalitäten und Informationsredundanz wird vermieden.

2.2 Begrenzte Aufmerksamkeitsressourcen und Informationsredundanz

Nahezu alle Modelle menschlicher Informationsverarbeitung gehen von begrenzten kognitiven Ressourcen des Arbeitsgedächtnisses aus. In der Literatur wird übereinstimmend vertreten, dass die Art der Informationsdarstellung, redundant oder komplementär, starke Auswirkung auf die Verarbeitung der Informationen und somit auf den Verlauf der Interaktion hat (Oviatt, 2002; Sutcliffe, 2002, Oviatt et al., 2004). Meist handelt es sich um Empfehlungen aufgrund theoretischer Überlegungen (Oviatt et al., 2004). So gibt es unterschiedliche Meinungen über die Auswirkung der redundanten Darstellung von Informationen über mehr als eine Modalität. Im Zusammenhang mit multimedialen Nut-

zerschnittstellen wird von einem positiven Effekt ausgegangen, der das Arbeitsgedächtnis unterstützt. Andererseits wird auch vermutet, dass die redundante Darstellung von ähnlichen Informationen (z.B. Sprachausgabe und Text) zu einem „Wettstreit“ (competition) um kognitive Ressourcen führen kann (Wickens et al., 1983; Lee 200). Daher sprechen sich einige Experten gegen redundante und für eine komplementäre Darstellung aus (Suttcliffe and Farrady, 1993; van Erp et al., 2006). Eine Empfehlung für das komplementäre Design von multimodaler Information stellt das „Modality Appropriateness Framework“ (van Erp et al., 2006) dar. Die Autoren empfehlen auf Grund verschiedener Studien, die den belastenden Effekt von redundanten Informationen gezeigt haben, eine komplementäre Informationsgestaltung. Sie weisen aber ausdrücklich auf die Notwendigkeit weiterer empirischer Daten hin.

3.0 Methode

3.1 Fragestellung: Wie wirkt sich linguistische Informationsredundanz auf die Gebrauchstauglichkeit aus?

Die oben genannte Fragestellung wurde anhand von sechs (Unterschieds-)

Hypothesen experimentell untersucht. Die Hypothesen umfassen verschiedene Usability Parameter.

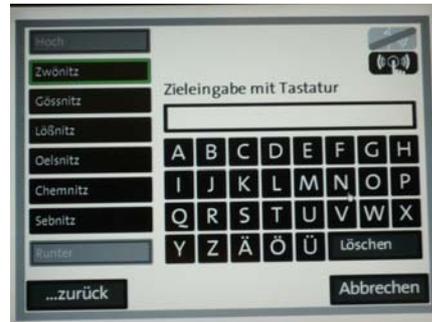


Abb. 1: Das Bild zeigt einen exemplarischen Dialogschritt unter dem Microsoft Internet Explorer 6.0 mit mittlerem Redundanzgrad

3.2 Drei Stufen der Informationsredundanz (unabhängige Variable)

Der Redundanzgrad zwischen visueller Textdarstellung und Sprachausgabe wurde in drei Stufen variiert (Tabelle 1). Bei maximaler Redundanz ist die Sprachausgabe identisch mit dem auf dem Bildschirm dargestellten Text. Bei mittlerer Redundanz werden die möglichen Nutzeraktionen als substantivierte Schlüsselwörter visuell präsentiert. Schließlich wird in der minimal redundanten Bedingung linguistische

Tabelle 1: Eine Systemausgabe in maximaler, mittlerer und minimaler linguistischer Redundanz

Redundanzgrad	Visuelle Ausgabe auf dem Bildschirm	Sprachausgabe
Maximal	„Wenn München Hauptbahnhof Ihr Ziel ist BESTÄTIGEN Sie, bitte. Andernfalls können Sie ein Ziel aus der Liste wählen. Falls Ihr Ziel nicht in der Liste ist, wählen Sie ÄNDERN.“	„Wenn München Hauptbahnhof Ihr Ziel ist bestätigen Sie, bitte. Andernfalls können Sie ein Ziel aus der Liste wählen. Falls Ihr Ziel nicht in der Liste ist, wählen Sie ändern.“
Mittel	Ziel bestätigen, Zielwahl aus Liste oder Ziel ändern	
Minimal	-	

Information nahezu¹ ausschließlich auditiv ausgeben.

3.3 Erhobene Usability-Parameter (abhängige Variablen)

Folgende Usability-Parameter wurden erhoben:

- Attraktivität (AttrakDiff 2.0)
- Pragmatische Qualität (AttrakDiff 2.0)
- Hedonische Qualität Stimulation (AttrakDiff 2.0)
- Geistige Beanspruchung (NASA Task Load Index)
- Aufgabebearbeitungszeit
- Anzahl der Nutzeraktionen zum Ziel

3.4 Das Untersuchungsobjekt: Ein multimodales Interface zur Zieleingabe

Die Fragestellung wurde anhand einer konkreten Aufgabe, der Suche in großen Listen eines Informationssystems, untersucht. Es handelt sich hierbei um ein Standardproblem, wie es häufig an öffentlichen oder anderen Informationssystemen (z.B. Zieleingabe an einem Fahrscheinautomat) auftritt. Die Zielsuche in großen Listen wurde ausgewählt, da es sich um eine praxisrelevante Aufgabe handelt, die von der Kombination einer grafischen- und einer Sprachschnittstelle profitiert (Sturm & Boves, 2005). Es wurde eine grafische Nutzerschnittstelle mit Sprachausgabe in Flash 9 implementiert und mit Microsoft Internet Explorer 6.0 auf einem 15 Zoll Touchbildschirm dargestellt (Abb. 1 & 2). Die Spracherkennung wurde mithilfe der

¹ Einige Schaltflächen (Buttons) und die Zielwörter beinhalten visuelle linguistische Information.

Wizard-of-Oz Technik (Fraser & Gilbert, 1991) durch den Versuchsleiter simuliert (Abb. 2 und 3). Die Datenerhebung fand im Usability-Labor der Universität Osnabrück statt.



Abb. 2: Die Zieleingabe wurde multimodales Interface auf einem 15 Zoll Touchbildschirm umgesetzt



Abb. 3: Die Spracherkennung wurde durch die Wizard-of-Oz Technik realisiert und aus dem Regieraum durch den Versuchsleiter kontrolliert.

3.5 Die Nutzeraufgabe: Zielsuche und Zieleingabe in einer großen Liste

Die Nutzer bearbeiten drei Aufgaben, die jeweils das Auswählen eines Zielortes aus einer Liste verlangten. Dabei variierte die Auswahl hinsichtlich der Anzahl der nötigen Interaktionschritte. In jeder Aufgabe gab es mindestens eine Spracheingabe und mindestens drei Eingaben über den Touchbildschirm. Falls es zu einem simulierten Erkennungsfehler kommt, wird dem Nutzer eine Liste von möglichen Zielen angeboten, aus der er ein Ziel wählen kann. Falls sich das sehr tief in der Liste befindet, kann der Nutzer auf ein Softkeyboard zurückgreifen und durch Eingabe

eines Buchstabens die Listengröße verringern. Schließlich wählt er sein Ziel aus der reduzierten Liste.

4.0 Ergebnisse

Auf deskriptiver Ebene lässt sich eindeutig feststellen, dass die Versuchspersonen in der Bedingung mit mittlerer Redundanz über alle sechs abhängigen Variablen die besten Werte erzielt haben. Dies betrifft sowohl die Performanzdaten als auch die Daten der Selbsteinschätzung. Allerdings sind alle aufgetretenen Effekte klein ($>0, 2$). Der deutlichste Effekt trat für die Skala Attraktivität auf, wobei der Mittelwert der Variante mit mittlerer Redundanz ($M = 5,14$) deutlich höher bewertet wurde, als die Mittelwerte der Variante mit maximaler ($M = 4,60$) und minimaler Redundanz ($M = 4,70$) (Tabelle 2). Somit bevorzugten die Versuchspersonen die Version mit mittlerer Redundanz. Die Effekte der anderen AttrakDiff 2.0 Skalen waren geringer.

Die Mittelwerte der Skala „Geistige Beanspruchung“ sind unter allen Bedingungen unter Wert 20 und damit weit unter der Skalenmitte. Die Variante mit minimaler Redundanz weist den höchsten Mittelwert ($M = 17,58$) auf. Die Mittelwerte der Varianten mit maximaler Redundanz ($M = 15,81$) und mittlerer Redundanz ($M = 12,81$) sind tendenziell niedriger. Also wurde auch hier die mittlere Variante am wenigsten kognitiv belastend gefunden.

Bei den Performanzdaten wiesen die Mittelwerte der Bearbeitungszeit sehr geringe Unterschiede auf. Den niedrigsten Mittelwert zeigt die Gruppe mit

mittlerer Redundanz ($M = 72$ Sekunden). Die Mittelwerte in der maximal redundanten ($M = 77$ Sekunden) und der minimal redundanten Bedingung ($M = 76$ Sekunden) sind nahezu identisch. Der Mittelwert der Bearbeitungszeit variiert zwischen etwa elf und zwölf Schritten. Die größte Anzahl an Schritten benötigten die Benutzer der minimal redundanten Variante ($M = 12,29$). Somit benötigte die Gruppe mit minimaler Redundanz in der Regel einen Schritt mehr als die Gruppe mit mittlerer Redundanz ($M = 11,19$). In den Bedingungen mit maximaler Redundanz ($M = 11,74$) und mittlerer Redundanz unterschieden sich die Mittelwerte um etwa einen halben Schritt. Die MANOVA war nicht signifikant (Wilks-Lambda = 0,179, $df = 12$, $F = 1,451$, $p = 0,147$) (Tabelle 2).

5.0 Schluss

Der Redundanzgrad linguistischer Information scheint Auswirkung auf verschiedene Usability Parameter einer Anwendung zu haben. Zwar ist das Ergebnis der MANOVA nicht signifikant, dennoch lässt sich feststellen, dass das mittlere Redundanzniveau über alle abhängigen Variablen zu den besten Ergebnissen geführt hat. Demnach sollte bei guten akustischen Bedingungen die linguistische Redundanz zwischen auditiver und visueller Ausgabe auf mittlerem Niveau gehalten werden. So kann visuelle Informationsredundanz reduziert und der Bildschirm effizienter genutzt werden. Möglicherweise ist der niedrige Schwierigkeitsgrad der zu bewältigende Aufgabe für die geringen Unterschiede zwischen den Gestaltungsvarianten verantwortlich. Es wäre denkbar, dass ein höherer Schwierigkeitsgrad sich gravierender in den Performanzmaßen aus-

Tabelle 2: Mittelwerte und Standardabweichungen der AttrakDiff 2.0

Skala „Attraktivität“ * Standardabweichung, ** Anzahl der Versuchspersonen

Skala	Redundanzgrad	Mittelwert	S*	N**
Attraktivität	Maximal	4,59	0,94	31
	Mittel	5,15	0,82	32
	Minimal	4,71	0,73	31

Tabelle 3: Abhängige Variablen und Signifikanz der ANOVAs

*Die MANOVA (Wilks-Lambda, $p = 0,147$) war nicht signifikant.

Abhängige Variable	F-Wert	Signifikanz
Attraktivität	3,91	0,02*
Pragmatische Qualität	6,6	0,00*
Hedonische Qualität Stimulation	0,88	0,42
Geistige Anforderungen	1,21	0,30
Aufgabenbearbeitungsschritte	1,41	0,32
Aufgabenbearbeitungszeit	1,36	0,26

wirkt. Da es sich bei den Versuchspersonen ausschließlich um Studenten gehandelt hat wäre es interessant, bei einer Folgeuntersuchung auch andere Bevölkerungsgruppen zu testen, die möglicherweise weniger Erfahrung mit Computeranwendungen besitzen.

6.0 Literatur

Dybkjaer, L. et al. (2004). Evaluation and usability of multimodal spoken language dialogue systems. *Speech Communication*, 43, 33-54.

Faraday, P. & Sutcliffe, A. (1993). A Method for multimedia interface design. *People and Computers, HCI'93*, 173-190.

Fraser & Gilbert, (1991). Simulating Speech Systems. *Computer Speech and Language*, 5, 81-99.

Hart, G.; Staveland, L. E. (1988). Development of NASA-TLX: Results of Empirical and Theoretical Research. In: P. A. Hancock & N. Meshkati (Eds.) *Human Mental Workload* (S. 239-250). Amsterdam: North Holland Press.

Hassenzahl, M., Burmester, M. Koller, F. (2003). AttrakDiff : Ein Fragebogen zur Messung pragmatischer und hedonischer Qualität In: G. Szwillus, J. Ziegler (Hrsg.), *Mensch & Computer 2003: in Bewegung* (S. 187-196). Stuttgart: B. G. Teubner.

Lamel, L. et al (2002). User evaluation of the Mask kiosk. *Speech Communication*, 38, 131-139. *Interaktion*

Oviatt, S. (2002). Multimodal Interfaces. In: Jacko, J. & Sears, A., *Handbook of Human Computer Interaction* (S. 286-304). New York: Lawrence Erlbaum.

Oviatt S. et al, (2004). When do we interact multimodally?: Cognitive load and multimodal communication patterns In: *Proceedings of the 6th international conference on Multimodal interfaces* (129 - 136). USA: State College, PA.

Sturm, J. & Boves L. (2005). Effective error recovery strategies for multimodal form-filling applications. *Speech Communication*, 45, 289-303.

Sutcliffe, A. (2002). Multimedia User Interface Design. In: Jacko, J. & Sears, A., *Handbook of Human Computer Interaction* (S. 246-261). New York: Lawrence Erlbaum.

Van Erp, J. B. et al. (2006). Multimodal Interfaces: A framework based on modality appropriateness. In: *Proceedings of the Human Factors and Ergonomics Society 50th Annual meeting-2006*, 1542-1549.

Wickens, C. et al., (1983). Compatibility and resource competition between modalities of input, central processing, and output. *Human Factors*, 25(2), 227-248.