



Hochschule für Angewandte Wissenschaften Hamburg
Hamburg University of Applied Sciences

Studienarbeit

Lennard Hamann

Eyetracker als Eingabemedien in der
multimodalen Interaktion

Lennard Hamann
Eyetracker als Eingabemedien in der multimodalen
Interaktion

Studienarbeit eingereicht im Rahmen der Veranstaltung
Seminar/Ringvorlesung
von Prof. Dr. Kai von Luck und Prof. Dr.-Ing. Bernd Schwarz
im Studiengang Master Of Science Informatik
am Fachbereich Elektrotechnik und Informatik
der Hochschule für Angewandte Wissenschaften Hamburg

Abgegeben am 28.02.2008

Inhaltsverzeichnis

Abbildungsverzeichnis	4
1 Einführung	5
2 Eyetracker als Eingabemedien	7
2.1 Grundlagen	7
2.1.1 Arten von Augenbewegungen	7
2.1.2 Eyetracking	7
2.2 Vorteile der Blickeingabe	9
2.3 Anwendungen der Blickeingabe	9
2.4 Probleme der Blickeingabe	10
2.4.1 Erkennung von Fixationen	10
2.4.2 Räumliche Auflösung	10
2.4.3 Zeitliche Auflösung	11
2.4.4 Midas-Touch Problem	11
2.4.5 Unbewußte Augenbewegungen	11
2.4.6 Aufmerksamkeit	11
2.4.7 Umwelteinflüsse	11
3 Multimodale Interaktion	12
3.1 Grundlagen	12
3.2 SmartKom	13
3.3 QuickSet	14
3.4 MATCH	14
4 Ansätze und Beobachtungen	16
4.1 Beobachtungen	16
4.2 Offene Fragen	17
5 Zusammenfassung	19
Literaturverzeichnis	20

Abbildungsverzeichnis

3.1	Interpretation von mimischen Gesten (aus [Wahlster 2002])	13
3.2	Multimodale Interaktion mit SmartKom (aus [Wahlster 2002])	14
3.3	Unimodale Eingabe von Handschrift beim MATCH-System (aus [Johnston u. a. 2005])	15

1 Einführung

Maschinen und Computer halten verstärkt Einzug in unseren modernen Alltag. Zum Kauf eines Bahntickets muss ein Fahrkarten-Automat bedient werden, Unterhaltungsmedien und technische Haushaltsgeräte werden komplexer und sogar an der Pflege älterer Menschen werden zukünftig Roboter beteiligt sein. Die Herausforderung ist, neue Systeme besonders benutzerfreundlich zu gestalten, damit die Menschen mit diesen Entwicklungen Schritt halten können. Dies bedeutet, daß die zukünftigen Systeme mehr auf die natürlichen Interaktionsmöglichkeiten und -gewohnheiten¹ des Menschen zugeschnitten werden müssen. Die heutigen Systeme sind textbasiert und mit ihren zahlreichen Optionen für den Computerlaien schwierig zu bedienen. Um die Interaktion menschenähnlicher und somit intuitiver zu gestalten, muss über Sprache, Gestik und Mimik, also multimodal, mit dem System interagiert werden können. Der Begriff Modalität bezeichnet hier zum einen die Sinnesorgane, mit denen der Benutzer die Ausgaben vom Computer wahrnimmt und zum anderen die Sensoren, mit denen ein Computer die Eingaben vom Benutzer empfangen kann. Bei der multimodalen Interaktion werden mehrere Modalitäten auf Anwender- wie auf Computerseite kombiniert. Der Begriff multimodale Interaktion beinhaltet multimodale Ein- und Ausgabe. In dieser Arbeit geht es vor allem um die multimodale Eingabe speziell unter Verwendung von Eyetrackern.

Ein bereits verbreitetes Anwendungsgebiet von Blickeingabe-Techniken² in der Mensch-Computer Interaktion³ sind Bereiche, in denen es schwierig oder unmöglich ist, die Hände zu benutzen, wie dies z.B. bei Menschen mit entsprechenden Behinderungen der Fall ist [Norris u. a. 1997]. In solchen Fällen kann Eyetracking als Augenmaus benutzt werden, um unterschiedliche elektronische Medien, wie z.B. Computer, Telefon, Fax oder Sprachgeneratoren zu steuern. Die Entwicklung der Blickeingabe zu einer Standard-Eingabetechnik ist noch in ihren Anfängen. Die immer leistungsfähigere und günstigere (Kamera-)Hardware, die

¹Mit dem Begriff der natürlichen Interaktion ist hier die Mensch-Mensch Kommunikation gemeint.

²In Anlehnung an den in der englischsprachigen Literatur gebräuchlichen Begriff 'gaze-input' wird in dieser Arbeit der Begriff Blickeingabe verwendet, wenn die Daten der Blickbewegungsmessung als Eingabe in ein IT-System interpretiert werden.

³In der Literatur werden die Begriffe Mensch-Computer Interaktion, Mensch-Maschine Interaktion, Mensch-Maschine Kommunikation und Mensch-Computer Kommunikation verwendet. Der Begriff Kommunikation deutet im Gegensatz zu Interaktion auf Kommunikationspartner auf gleicher Ebene (z.B. Mensch-Mensch Interaktion, Kommunikationsprotokolle in der Informatik) hin. Die Verwendung von Maschine oder Computer hängt vom Kontext ab. In Anlehnung an den Begriff 'Human Computer Interaction' aus der englischsprachigen Literatur wird hier der Begriff Mensch-Computer Interaktion verwendet.

Verbreitung von Webcams und die Integration von Kameras in Computer/Monitore sprechen jedoch für die Entwicklung von Techniken zur Steuerung von Computern per Eyetracking. Der Erfolg von interaktiven Spielen wie z.B. der Nintendo Wii Konsole zeigt die Chancen von neuen Eingabemedien.

In aktuellen Arbeiten (siehe Abschnitt 2.3) werden Eyetracker erfolgreich als Eingabegerät eingesetzt. Es geht dort vor allem darum, durch bewusst durchgeführte Augenbewegungen Anwendungen zu steuern. Häufig werden Blickeingabe-Techniken mit anderen Eingabemedien zu einem multimodalen System kombiniert. Es werden zum Beispiel deiktische Ausdrücke, die per Spracheingabe getätigt wurden, durch eine Einbeziehung der Blickposition aufgelöst.

Multimodale Systeme bieten ihren Benutzern die Möglichkeit, in verschiedenen Situationen die passende Eingabemethode zu wählen. Komplexe Sachverhalte können besser mit Sprache übermittelt werden, räumliche Beziehungen eher mit (Zeige-) Gesten. Durch die gleichzeitige Nutzung mehrerer Kanäle kann die Interaktion effizienter und robuster ablaufen.

In [Hamann 2007] werden die Grundlagen von Eyetracking-Techniken zusammengefaßt. In [Hamann 2008c] werden Arbeiten zur multimodalen Interaktion unter besondere Berücksichtigung von Systemen, die Eyetracker als Eingabemedium verwenden, ausgewertet. In [Hamann 2008b] wird auf Basis der vorliegenden Arbeit eine Anwendung zur Untersuchung von Fragen und Problemen zum Thema Eyetracking als Eingabegerät erstellt. Dort findet sich auch eine Beschreibung des im UbiComp-Labor verwendeten Eyetracking-Systems. Im SmartKom-Projekt⁴ wird ein umfangreiches Konzept zur multimodalen Interaktion vorgestellt. In [Oviatt 1999] werden Probleme der multimodalen Interaktion beschrieben.

Im folgenden Kapitel 2 über Eyetracker als Eingabemedien werden zunächst die Grundlagen dieser Technik beschrieben und Anwendungsmöglichkeiten und Probleme dargestellt. Im daran anschließenden Kapitel 3 über multimodale Interaktion wird die Mensch-Mensch Kommunikation als Grundlage und Vorbild für multimodale Systeme dargestellt. Dann werden Projekte vorgestellt, die das Konzept der multimodalen Interaktion bestätigen. Im anschließenden Kapitel 4 werden Beobachtungen aus bestehenden Arbeiten zu den Themen Blickeingabe bzw. multimodale Eingabe berichtet. Außerdem werden offene Fragen und Ansätze aus den beiden erwähnten Bereichen benannt.

⁴<http://www.smartkom.org>

2 Eyetracker als Eingabemedien

In diesem Kapitel werden zunächst die Grundlagen von Augenbewegungen und Eyetracking-Techniken beschrieben. Es werden die Vorteile der Blickeingabe genannt und Anwendungsmöglichkeiten gezeigt. Abschließend wird dann auf die Probleme der Blickeingabe eingegangen.

2.1 Grundlagen

2.1.1 Arten von Augenbewegungen

Während der Fixationen, bei denen sich das Auge in relativer Ruhe zu einem Sehobjekt befindet und nicht bewußt bewegt wird, werden visuelle Informationen aufgenommen. Sakkaden sind sehr schnelle und ruckhafte Augenbewegungen, mit denen die Augen bewusst von einem Fixationspunkt zum nächsten bewegt werden. Während einer Sakkade ist das visuelle Wahrnehmungsvermögen drastisch eingeschränkt. Auch während einer Fixation führt das Auge geringe Bewegungen (Mikrosakkaden) durch, um z.B. ein 'Einbrennen' des aktuell betrachteten Bildes auf der Netzhaut zu vermeiden. In [Hamann 2007] wird die Physiologie der Augen als Grundlage der Augenbewegungsmessung ausführlicher geschildert.

2.1.2 Eyetracking

Das Eyetracking ist das Beobachten von Augenbewegungen oder von Blickbewegungen. Mit dem Begriff Augenbewegungen werden sowohl Bewegungen des Augapfels als auch Lid-schluß und Pupillenmotorik zusammengefaßt [Galley 2001]. Beim Messen von Blickbewegungen werden die Augenbewegungen in Verbindung mit den vom Auge aufgenommenen Informationen interpretiert. Hierfür ist es also notwendig, die Augenbewegungen und den Zielort der Augen zu erfassen. In dieser Arbeit werden nur videobasierte berührungslose Eyetracker betrachtet, weil andere wie z.B. Head-Mounted-Systeme den Anwender bei der Steuerung zu sehr einschränken und nicht benutzerfreundlich sind. Die videobasierten Systeme bestehen aus mindestens einer Kamera, welche das Bild der Augen aufzeichnet. Aus

der Aufzeichnung wird dann die Blickposition berechnet. Die videobasierten Systeme müssen kalibriert werden. Eine umfangreichere Zusammenstellung von Techniken zur Messung von Augenbewegungen ist in [Hamann 2007] zu finden.

Messbereich

Der örtliche Messbereich ist der Winkelbereich, in dem die Augenbewegungen gemessen werden. Der zeitliche Messbereich ist der Aufzeichnungsdauer. Diese hängt ab von der Speicherkapazität des Datenträgers.

Auflösung

Die örtliche Auflösung ist die kleinste messbare Veränderung der Blickrichtung. Die zeitliche Auflösung ist die Anzahl der Messwerte pro Zeiteinheit.

Messgenauigkeit

Die örtliche Genauigkeit ist die Abweichung des vom Messsystem ermittelten Wertes und der tatsächlichen Augenposition bzw. Blickposition. Die örtliche Genauigkeit kann nicht größer als die örtliche Auflösung sein. Die zeitliche Genauigkeit entspricht dem Zeitabschnitt zwischen der Erfassung einer Augenposition und der Bereitstellung der Messgröße durch das System.

Messdaten

Die Eyetracking-Systeme können verschiedene Messdaten liefern, welche unterschiedlich interpretiert werden können:

- die Blickposition gibt Auskunft über aktuelle Aufmerksamkeit des Benutzers und kann so helfen z.B. deiktische Spracheingaben zu interpretieren
- die Fixationsdauer kann als Indikator für die Tiefe der Verarbeitung dienen
- die Pupillengröße kann als Indikator für die mentale Beanspruchung gelten
- die Lidschlagamplitude kann als Indikator für emotionale Erregung dienen
- der 'Scan path' zeigt die Auffindbarkeit von Objekten an
- die Anwesenheit, Position und Anzahl der Augen im Kamerabild

- die Anwesenheit, Position und Anzahl der Gesichter im Kamerabild

In dieser Arbeit geht es vor allem um die Blickpositionsmessung. Der Benutzer soll durch bewusst durchgeführte Augenbewegungen eine Anwendung steuern.

2.2 Vorteile der Blickeingabe

Die Vorteile von Blickeingabe-Techniken sind:

- Augenbewegungen sind schnell und natürlich
- Blicke zeigen den Fokus der visuellen Aufmerksamkeit an
- Augenbewegungen müssen nicht erlernt werden im Gegensatz zur Auge-Hand Koordination bei der klassischen Computer-Maus
- die Auswahl von Objekten auf einem Monitor kann mit den Augen viel schneller erfolgen als mit einer Maus [Kumar u. a. 2007b]; so können z.B. deiktisch Äußerungen aufgelöst werden
- die Hände müssen nicht verwendet werden
- Blicke sind eine natürliche Geste: Menschen sehen das Objekt, mit dem sie interagieren möchten, an [Jacob 1991], [Salvucci und Anderson 2000]
- im Vergleich mit anderen Eingabemedien kann mit dem Eyetracker ein schnelleres Finden und Auswählen eines Ziels möglich sein [Ware und Mikaelian 1987], [Sibert und Jacob 2000]

2.3 Anwendungen der Blickeingabe

In diesem Abschnitt werden die Anwendungsmöglichkeiten der Blickeingabe anhand von Beispielen dargestellt. Diese bestätigen das Konzept der Blickeingabe, werfen aber auch Fragen auf, die in zukünftigen Arbeiten geklärt werden sollen.

In [Smith und Graham 2006] wird Blickeingabe genutzt, um Videospiele zu bedienen.

In [Dickie u. a. 2006] werden Eyetracker verwendet, um ein Eingabemedium für mehrere Computer zu benutzen. Die Eingaben werden jeweils auf den gerade angeblickten Monitor bezogen. [Benko und Feiner 2005] verfolgen Kopfbewegungen, um ein ähnliches System zu implementieren.

Die Anwendung iDict [Hyrskykari u. a. 2000] verwendet Eyetracking-Informationen, um den Benutzer beim Lesen fremdsprachiger Texte zu unterstützen.

Das TableTop-System [Holman 2007] bietet eine virtuelle Arbeitsfläche mit Multiuser-Eyetracking.

Die HCI-Group der Stanford University¹ arbeitet unter anderem an Grundlagen der Blickeingabe. In diesem Rahmen wurden verschiedene Anwendungen der Blickeingabe implementiert. Beispielhaft sei hier die Applikation EyeExpose [Kumar u. a. 2007b] genannt, mit der der Benutzer unter Verwendung von Blicken und einer Taste zwischen laufenden Programmen wechseln kann.

2.4 Probleme der Blickeingabe

In diesem Abschnitt werden Probleme der Blickeingabe betrachtet. Es bestehen folgende Probleme beim videobasiertem Eyetracking mit einer berührungslosen Kamera.

2.4.1 Erkennung von Fixationen

Der Blick ruht beim Betrachten der graphischen Benutzerschnittstelle nie fest auf einer Position oder gar einem Pixel (dagegen spricht auch die räumliche Auflösung). Der Grund hierfür liegt in der Physiologie der Augen (Mikrosakkaden, siehe Abschnitt 2.1.1). Das Problem ist also, einen räumlichen Bereich zu definieren, in dem sich der Blick für eine bestimmte Zeitspanne aufhalten muss, um als Fixation zu gelten [Kumar 2007].

2.4.2 Räumliche Auflösung

Die räumliche Auflösung der Eyetracking-Systeme ist begrenzt. Die Gründe dafür liegen beim Video-Sensor, der Kalibrierung und der System-Konfiguration (Abstände und Winkelverhältnisse zwischen Monitor, Benutzer und Eyetracker). Das bedeutet, dass die Elemente, die per Blick selektiert werden sollen, bestimmte Abmessungen und Abstände haben müssen, um eindeutig ausgewählt werden zu können [Miniotas u. a. 2006].

¹<http://hci.stanford.edu>

2.4.3 Zeitliche Auflösung

Die Messdaten der Eyetracking-Systeme sind diskret (z.B. 60 - 120 Blickpunkte pro Sekunde für Tobii X120). Außerdem ergibt sich eine Latenzzeit durch die Verarbeitung der aufgezeichneten Videodaten zu Blickpositionen.

2.4.4 Midas-Touch Problem

Die Anwender sind es nicht gewöhnt, nur durch Augenbewegungen Geräte zu bedienen. Sie erwarten, Objekte ansehen zu können, ohne sofort mit ihnen zu interagieren [Jacob 1991], [Velichkovsky u. a. 1997]. Der Einsatz einer weiteren Modalität kann dieses Problem lösen [Kumar u. a. 2007b], [Kumar u. a. 2007c]. Zum Beispiel durch das Drücken einer Taste oder einen Sprachbefehl kann der Anwender dem System mitteilen, dass die Blicke als Eingabe interpretiert werden können. Alle anderen Blickbewegungen werden dann vom System nicht als Eingabe gewertet.

2.4.5 Unbewußte Augenbewegungen

Nicht alle Augenbewegungen werden bewusst durchgeführt. Zu den unbewussten Augenbewegungen gehören z.B. die Mikrosakkaden. Durch Glättung der Tracking-Daten und durch Schwellwerte für Fixationsdauer und Fixationsort können die Probleme unbewußter Augenbewegungen für die Blickeingabe reduziert werden.

2.4.6 Aufmerksamkeit

Die Tatsache, das ein Anwender ein Element betrachtet bedeutet nur, dass sich die visuelle Aufmerksamkeit auf das Element bezieht. Ob die kognitive Aufmerksamkeit ebenfalls mit dem Element beschäftigt ist, ist mit einem Eyetracking-System alleine nicht festzustellen.

2.4.7 Umwelteinflüsse

Wenn durch Kopfbewegungen, Blinzeln oder Veränderungen der Lichtverhältnisse die Augen für einen Moment nicht von der Eyetracking-Kamera erfasst werden können, muss mit dem Tracking neu begonnen werden.

3 Multimodale Interaktion

Die Mensch-Mensch Kommunikation ist Grundlage für ein System, daß eine natürliche Interaktion zwischen Mensch und Maschine unterstützen soll. Deshalb werden hier die Grundlagen der Mensch-Mensch Kommunikation dargestellt. Des weiteren werden in diesem Kapitel mit SmartKom, QuickSet und MATCH drei multimodale Konzepte vorgestellt.

3.1 Grundlagen

In diesem Abschnitt wird die Mensch-Mensch Kommunikation als Basis für die Mensch-Computer Interaktion beschrieben. Es wird herausgestellt, dass die Mensch-Mensch Kommunikation multimodal verläuft.

Mensch-Mensch Kommunikation

Die zwischenmenschliche Kommunikation hat verbale und nonverbale Anteile. Der verbale Anteil ist die Sprache. Die nonverbale Kommunikation beinhaltet alle übertragenen Informationen, die nicht sprachlicher Natur sind. Dazu gehören die Kommunikationskanäle der Mimik, Gestik und Körpersprache. Sie tragen zum großen Teil die Emotionen und unterstützen die Verständigung.

Die menschliche Kommunikation ist multimodal. Der Mensch ist in der Lage, die Sprache und die nonverbalen Kommunikationskanäle intuitiv zu kombinieren, um seine Anliegen den Interaktionspartnern verständlich zu machen.

Die nonverbale Kommunikation übermittelt durch eine Kombination vieler Modalitäten Informationen, die die Kommunikation unterstützen. Der Informationsfluss ist dabei parallel, d.h. über mehrere Modalitäten werden gleichzeitig Signale gesendet. Die Mensch-Mensch Kommunikation läuft über Sprache aber auch bewusst und unbewusst zeitgleich über Mimik, Gestik, Körperhaltung und Augenkontakt.

Die Effizienz der Mensch-Mensch Kommunikation ergibt sich aus der Robustheit und der Adaptivität paralleler Modalitäten. Dadurch werden Fehler in der Kommunikation verringert (durch Redundanz) und die Menge an Informationen gesteigert (durch Parallelität).

Eine ausführlichere Beschreibung der Mensch-Mensch Kommunikation als Grundlage für die multimodale Mensch-Computer Interaktion findet sich in [Hamann 2008].

Im folgenden werden mit SmartKom, QuickSet und MATCH drei multimodale Systeme vorgestellt.

3.2 SmartKom

Das SmartKom-Projekt hat ein Konzept für ein multimodales Dialogsystem erarbeitet. Das System nutzt Sprache, Gesten und Mimik zur Ein- und Ausgabe. Zur Ausgabe und zum Dialog mit dem Benutzer wird ein animierter Agent verwendet. Bei der Interpretation der Eingaben werden Informationen über den Benutzer, die Aufgabe und den Kontext berücksichtigt.

- (1) **Smartakus: Hier sehen Sie die Übersicht zum heutigen ZDF-Programm.**
- (2) **Benutzer: Echt toll.**
- 
- (3) **Smartakus: Ich zeige Ihnen alternativ das Programm eines anderen Senders.**
- (2') **Benutzer: Echt toll.**
- 
- (3') **Smartakus: Welche Sendungen wollen Sie aus dem ZDF-Programm sehen oder aufzeichnen?**

Abbildung 3.1: Interpretation von mimischen Gesten (aus [Wahlster 2002])

In Abbildung 3.1 ist ein Dialogverlauf abgebildet, bei dem das System mimische Gesten in die Dialog-Planung miteinbezieht.

Die Abbildung 3.2 zeigt die Einbindung von Zeigegesten in SmartKom.

Im Rahmen des Smartkom-Projektes wurden mit SmartKom-Mobile, SmartKom-Home/Office und SmartKom-Public drei prototypische Anwendungen erstellt, die das Konzept multimodaler Interaktion bestätigen. Mit diesen Anwendungen kann z.B. auf Kinoprogramme und Restaurantführer zugegriffen werden.

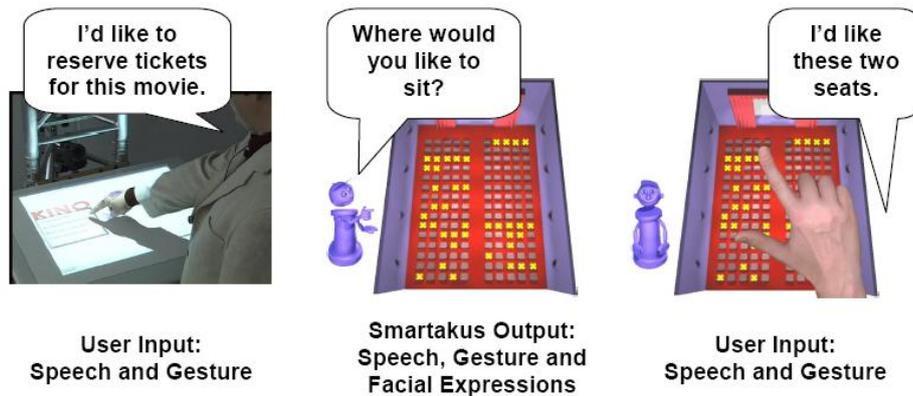


Abbildung 3.2: Multimodale Interaktion mit SmartKom (aus [Wahlster 2002])

3.3 QuickSet

Das QuickSet-System [Cohen u. a. 1997a], [Cohen u. a. 1997b] ist eine multimodale Anwendung für PDAs und Desktop-PCs, die das Konfigurieren und Steuern von militärischen Simulationen ermöglicht. Es verwendet eine Kombination von Spracheingabe und Gesten, die mit einem Stift eingegeben werden. Die beiden Modalitäten sollen ihre jeweiligen Schwächen gegenseitig kompensieren. Die Spracheingabe ist auf festgelegte, anwendungsbezogene Begriffe reduziert. Bei der Integration der Modalitäten werden zuerst für jede einzelnen die möglichen Bedeutungen ermittelt. Dann wird aus den am besten zusammen passenden Interpretationen (welche nicht die höchsten Einzelbewertungen sein müssen) das Ergebnis ermittelt. Beim QuickSet-Einsatz wurde eine Verringerung von Eingabebefehlen durch die gestische Eingabemöglichkeit (z.B. bei Lokalisierungsangaben) gegenüber unimodaler Spracheingabe beobachtet.

3.4 MATCH

In [Johnston u. a. 2005] wird das MATCH-Projekt (Multimodal Access To City Help) vorgestellt. Das MATCH-System ermöglicht es dem Anwender, per PDA auf Gastronomie- und U-Bahn Informationen zuzugreifen. Es nutzt wie das QuickSet-System eine Kombination von Sprach- und Stifteingabe, wobei mit dem Stift auch Handschrift-Eingabe möglich ist (siehe Abbildung 3.3). Dabei ist die Schrifterkennung auf 285 anwendungsbezogene Vokabeln eingeschränkt. Die multimodalen Eingaben werden mit einem Best-Match-Verfahren zusammengeführt. Da Anwender eines multimodalen Systems aber nicht zwangsläufig jeden Befehl multimodal ausdrücken [Oviatt u. a. 1997], ist auch eine unimodale Eingabe möglich.

Dies ist durch einen Timeout umgesetzt. Erfolgt innerhalb einer Sekunde nach einer Sprach- oder Stifteingabe keine weitere Eingabe, wird unimodal interpretiert.

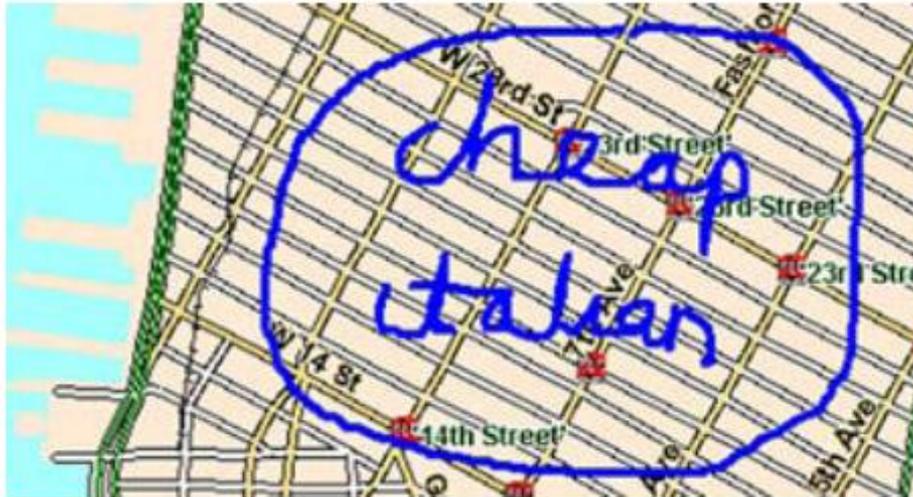


Abbildung 3.3: Unimodale Eingabe von Handschrift beim MATCH-System (aus [Johnston u. a. 2005])

4 Ansätze und Beobachtungen

In diesem Kapitel werden Ansätze, Beobachtungen und offene Fragen aus den Bereichen Blickeingabe und multimodale Interaktion zusammengefasst, um im Projekt oder einer Masterarbeit bearbeitet zu werden.

4.1 Beobachtungen

In [Miniotas u. a. 2006] werden die selektierbaren Objekte einer Standard-Benutzerschnittstelle auf ihre Blickeingabe-Tauglichkeit hin untersucht. Hierzu wird ein System, das Sprach- und Blickeingabe kombiniert verwendet. Dort wird eine Abmessung von 30 x 30 Pixel für zu selektierende Objekte und ein Abstand von 10 Pixeln zwischen den Objekten vorgeschlagen. Ebendort wird eine Fixations-Dauer von 1500ms zur Selektion eines Elements vorgeschlagen.

In [Ashmore u. a. 2005] wird der fixierte Bereich optisch vergrößert. So werden die Probleme Midas-Touch und räumliche Auflösung verringert. In Versuchen konnte eine Steigerung der Genauigkeit und Schnelligkeit beim Auswählen von Objekten festgestellt werden.

In [Spakov und Miniotas 2005] wird die Auswahl von Standard-Menueinträgen verbessert, in dem der fixierte Eintrag vergrößert dargestellt wird, um dann selektiert zu werden.

[Bates und Istance 2002] beschreiben ebenfalls eine Technik, bei der fixierte Objekte vergrößert werden, um eine genauere Auswahl treffen zu können.

In [Huckauf u. a. 2005] wird sich mit der Selektion von Objekten durch Blicke beschäftigt. Hierzu wird die Anti-Sakkaden Technik vorgestellt. Wird ein Objekt fixiert, erscheint eine Kopie des Objekts neben dem Objekt. Die Kopie muß dann zur fixiert werden, um eine Selektion auszulösen. So soll die visuelle Aufmerksamkeit 'gelenkt' und ein Fokus-Verlust verhindert werden.

In [Kaur u. a. 2003] wird mit einer Kombination aus Blick- und Spracheingabe gearbeitet. Bei einer 'Move it there'-Aufgabe wurde beobachtet, dass die Fixation, die das 'it' auflöst, vor dem gesprochenen 'move' stattfindet. Es wurde eine 95% Genauigkeit¹ erreicht, wenn die

¹Genauigkeit bedeutet hier, dass das richtige Objekt selektiert wurde.

Fixation vor dem 'move' zur Auswahl verwendet wurde. Wenn die Fixation zum Zeitpunkt der Aussprache des 'it' benutzt wurde, konnte nur eine Genauigkeit von 60% erreicht werden.

In [Kumar 2007] wird ein Algorithmus zur Glättung der Eyetracker-Messdaten vorgestellt. Dieser Algorithmus ermittelt gewichtete Durchschnittswerte der gemessenen Blickpositions-Koordinaten.

Die heutigen Standard-Widget-Bibliotheken sind auf Maus-Eingabe optimiert. Die Abmessungen der Objekte und ihre Navigierbarkeit sind auf eine Bedienung mit der Maus ausgelegt. In [Beinhauer 2006] wird deshalb die Entwicklung einer Widget-Bibliothek, die für Blickeingabe optimiert ist, vorgeschlagen.

4.2 Offene Fragen

In diesem Abschnitt werden offene Fragen aus den Bereichen Blickeingabe und multimodale Interaktion zusammengefasst. Diese beziehen sich auf die Beobachtungen anderer Arbeiten und deren Anwendbarkeit in der UbiComp-Labor Umgebung. Im Projekt und einer Master-Thesis könnten diese Fragen aufgegriffen und weiter bearbeitet werden.

Eyetracking

- Wie lange muss eine Fixation andauern, um als Selektion zu gelten (unimodal)?
- Wie groß müssen Objekte der Bedienoberfläche sein, um durch Blickeingabe selektiert werden zu können?
- Welcher Mindestabstand zwischen den Objekten ist nötig?
- Soll der Blickpunkt graphisch dargestellt werden? Mit welchen Abmessungen?
- Ist es hilfreich, das fixierte Objekt grafisch hervorzuheben, zu markieren, zu zoomen? Nach welcher Fixationsdauer wird damit begonnen?

Multimodal

- Welche zweite Modalität eignet sich, um mit einem Eyetracker eine multimodale Eingabe zu ergeben?
- Welche Verbesserung ergibt sich durch die zweite Modalität?
- Wie ist mit zwei widersprüchlichen Eingaben umzugehen?

-
- Wie ist der zeitliche Zusammenhang zwischen einer Spracheingabe und der dazugehörigen Geste (Blickeingabe)? Läßt sich dieser in einem Algorithmus fassen?
 - Gibt es benutzerspezifische Muster in der Verwendung mehrerer Modalitäten? Ist das System trainierbar?

5 Zusammenfassung

In dieser Arbeit wurde beschrieben, warum in Zukunft benutzerfreundlichere, an menschlichen Kommunikationsgewohnheiten orientierte Benutzerschnittstellen benötigt werden. Die Grundlagen von Eyetracking-Techniken und multimodaler Interaktion wurden dargestellt. Es wurden Anwendungen vorgestellt, die bestätigen, daß die Blickeingabe-Technik eine vielversprechende Ergänzung zu anderen Eingabemedien ist. Ebenso wurde die Benutzerfreundlichkeit und Effizienz von multimodaler Eingabe gezeigt. Schließlich wurden Ergebnisse anderer Arbeiten herangezogen, um Ansätze und Fragestellungen für das Projekt zu erarbeiten.

Literaturverzeichnis

- [Ashmore u. a. 2005] ASHMORE, Michael ; DUCHOWSKI, Andrew T. ; SHOEMAKER, Garth: Efficient eye pointing with a fisheye lens. In: *GI '05: Proceedings of Graphics Interface 2005*. School of Computer Science, University of Waterloo, Waterloo, Ontario, Canada : Canadian Human-Computer Communications Society, 2005, S. 203–210. – ISBN 1-56881-265-5
- [Bates und Istance 2002] BATES, Richard ; ISTANCE, Howell: Zooming interfaces!: enhancing the performance of eye controlled pointing devices. In: *Assets '02: Proceedings of the fifth international ACM conference on Assistive technologies*. New York, NY, USA : ACM, 2002, S. 119–126. – ISBN 1-58113-464-9
- [Beinhauer 2006] BEINHAUER, Wolfgang: A widget library for gaze-based interaction elements. In: *ETRA '06: Proceedings of the 2006 symposium on Eye tracking research & applications*. New York, NY, USA : ACM, 2006, S. 53–53. – ISBN 1-59593-305-0
- [Benko und Feiner 2005] BENKO, Hrvoje ; FEINER, Steven: Multi-monitor mouse. In: *CHI '05: CHI '05 extended abstracts on Human factors in computing systems*. New York, NY, USA : ACM, 2005, S. 1208–1211. – ISBN 1-59593-002-7
- [Cohen u. a. 1997a] COHEN, Philip R. ; JOHNSTON, Michael ; MCGEE, David ; OVIATT, Sharon ; PITTMAN, Jay ; SMITH, Ira ; CHEN, Liang ; CLOW, Josh: QuickSet: multimodal interaction for distributed applications. In: *MULTIMEDIA '97: Proceedings of the fifth ACM international conference on Multimedia*. New York, NY, USA : ACM, 1997, S. 31–40. – ISBN 0-89791-991-2
- [Cohen u. a. 1997b] COHEN, Philip R. ; JOHNSTON, Michael ; MCGEE, David ; OVIATT, Sharon ; PITTMAN, Jay ; SMITH, Ira ; CHEN, Liang ; CLOW, Josh: QuickSet: multimodal interaction for simulation set-up and control. In: *Proceedings of the fifth conference on Applied natural language processing*. San Francisco, CA, USA : Morgan Kaufmann Publishers Inc., 1997, S. 20–24
- [Dickie u. a. 2006] DICKIE, Connor ; HART, Jamie ; VERTEGAAL, Roel ; EISER, Alex: Look-Point: an evaluation of eye input for hands-free switching of input devices between multiple computers. In: *OZCHI '06: Proceedings of the 20th conference of the computer-human interaction special interest group (CHISIG) of Australia on Computer-human interaction:*

- design: activities, artefacts and environments*. New York, NY, USA : ACM, 2006, S. 119–126. – ISBN 1-59593-545-2
- [Galley 2001] GALLEY, N.: Physiologische Grundlagen und Meßmethoden der okulomotorischen Aktivität. In: RÖSLER, Frank (Hrsg.): *Grundlagen und Methoden der Psychophysiologie*. Göttingen, Bern [u.a.] : Hogrefe, Verl. für Psychologie, 2001. – ISBN 3-8017-0551-X
- [Hamann 2007] HAMANN, Lennard: *Grundlagen des Eyetrackings*. HAW Hamburg. 2007. – URL <http://users.informatik.haw-hamburg.de/~ubicomp/projekte/master2007/hamann/bericht.pdf>
- [Hamann 2008a] HAMANN, Lennard: *Eyetracker als Eingabemedien in der multimodalen Interaktion*. HAW Hamburg. 2008. – URL <http://users.informatik.haw-hamburg.de/~ubicomp/projekte/master07-08/hamann/bericht.pdf>
- [Hamann 2008b] HAMANN, Lennard: *Konzept für eine multimodale Test-Umgebung*. HAW Hamburg. 2008. – URL <http://users.informatik.haw-hamburg.de/~ubicomp/projekte/master07-08-proj/hamann/bericht.pdf>
- [Hamann 2008c] HAMANN, Lennard: *Multimodale Interaktion*. HAW Hamburg. 2008. – URL <http://users.informatik.haw-hamburg.de/~ubicomp/projekte/master07-08-aw/hamann/bericht.pdf>
- [Holman 2007] HOLMAN, David: Gazetop: interaction techniques for gaze-aware tabletops. In: *CHI '07: CHI '07 extended abstracts on Human factors in computing systems*. New York, NY, USA : ACM, 2007, S. 1657–1660. – ISBN 978-1-59593-642-4
- [Huckauf u. a. 2005] HUCKAUF, Anke ; GOETTEL, Timo ; HEINBOCKEL, Malte ; URBINA, Mario: What you don't look at is what you get: anti-saccades can reduce the midas touch-problem. In: *APGV '05: Proceedings of the 2nd symposium on Applied perception in graphics and visualization*. New York, NY, USA : ACM, 2005, S. 170–170. – ISBN 1-59593-139-2
- [Hyrskykari u. a. 2000] HYRSKYKARI, Aulikki ; MAJARANTA, Paivi ; AALTONEN, Antti ; RAIHA, Kari-Jouko: Design issues of iDICT: a gaze-assisted translation aid. In: *ETRA '00: Proceedings of the 2000 symposium on Eye tracking research & applications*. New York, NY, USA : ACM, 2000, S. 9–14. – ISBN 1-58113-280-8
- [Jacob 1991] JACOB, Robert J. K.: The use of eye movements in human-computer interaction techniques: what you look at is what you get. In: *ACM Trans. Inf. Syst.* 9 (1991), Nr. 2, S. 152–169. – ISSN 1046-8188

- [Johnston u. a. 2005] JOHNSTON, Michael ; BANGALORE, Srinivas ; VASIREDDY, Gunaranjan ; STENT, Amanda ; EHLEN, Patrick ; WALKER, Marilyn ; WHITTAKER, Steve ; MALOOR, Preetam: MATCH: An Architecture for Multimodal Dialogue Systems. In: *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*. Philadelphia, USA, 2005, S. 376–383
- [Kaur u. a. 2003] KAUR, Manpreet ; TREMAINE, Marilyn ; HUANG, Ning ; WILDER, Joseph ; GACOVSKI, Zoran ; FLIPPO, Frans ; MANTRAVADI, Chandra S.: Where is it? Event Synchronization in Gaze-Speech Input Systems. In: *ICMI '03: Proceedings of the 5th international conference on Multimodal interfaces*. New York, NY, USA : ACM, 2003, S. 151–158. – ISBN 1-58113-621-8
- [Kumar 2007] KUMAR, Manu: *GUIDe Saccade Detection and Smoothing Algorithm*. Technical Report CSTR 2007-03. Stanford Human-Computer Interaction Group. 2007. – URL <http://hci.stanford.edu/cstr/reports/2007-03.pdf>
- [Kumar u. a. 2007a] KUMAR, Manu ; KLINGNER, Jeff ; PURANIK, Rohan ; WINOGRAD, Terry ; PAEPCKE, Andreas: *Improving the Accuracy of Gaze Input*. Stanford Human-Computer Interaction Group. 2007. – URL <http://hci.stanford.edu/research/GUIDe/publications/StanfordCSTR2007-12-ImprovingtheAccuracyofGazeInput.pdf>
- [Kumar u. a. 2007b] KUMAR, Manu ; PAEPCKE, Andreas ; WINOGRAD, Terry: *EyeExposé: Switching Applications with Your Eyes*. Stanford Human-Computer Interaction Group. 2007. – URL hci.stanford.edu/cstr/reports/2007-02.pdf
- [Miniotas u. a. 2006] MINIOTAS, Darius ; SPAKOV, Oleg ; TUGOY, Ivan ; MACKENZIE, I. S.: Speech-augmented eye gaze interaction with small closely spaced targets. In: *ETRA '06: Proceedings of the 2006 symposium on Eye tracking research & applications*. New York, NY, USA : ACM, 2006, S. 67–72. – ISBN 1-59593-305-0
- [Norris u. a. 1997] NORRIS, Gregg ; ; WILSON, Eric: The Eye Mouse, an eye communication device. In: *Proceedings of the IEEE 1997 23rd Northeast*, 1997, S. 66 – 67. – ISBN 0-7803-3848-0
- [Oviatt 1999] OVIATT, Sharon: Ten myths of multimodal interaction. In: *Commun. ACM* 42 (1999), Nr. 11, S. 74–81. – ISSN 0001-0782
- [Oviatt u. a. 1997] OVIATT, Sharon ; DEANGELI, Antonella ; KUHN, Karen: Integration and synchronization of input modes during multimodal human-computer interaction. In: *CHI '97: Proceedings of the SIGCHI conference on Human factors in computing systems*. New York, NY, USA : ACM, 1997, S. 415–422. – ISBN 0-89791-802-9

- [Salvucci und Anderson 2000] SALVUCCI, Dario D. ; ANDERSON, John R.: Intelligent gaze-added interfaces. In: *CHI '00: Proceedings of the SIGCHI conference on Human factors in computing systems*. New York, NY, USA : ACM, 2000, S. 273–280. – ISBN 1-58113-216-6
- [Sibert und Jacob 2000] SIBERT, Linda E. ; JACOB, Robert J. K.: Evaluation of eye gaze interaction. In: *CHI '00: Proceedings of the SIGCHI conference on Human factors in computing systems*. New York, NY, USA : ACM, 2000, S. 281–288. – ISBN 1-58113-216-6
- [Smith und Graham 2006] SMITH, J. D. ; GRAHAM, T. C. N.: Use of eye movements for video game control. In: *ACE '06: Proceedings of the 2006 ACM SIGCHI international conference on Advances in computer entertainment technology*. New York, NY, USA : ACM, 2006, S. 20. – ISBN 1-59593-380-8
- [Spakov und Miniotas 2005] SPAKOV, Oleg ; MINIOTAS, Darius: Gaze-based selection of standard-size menu items. In: *ICMI '05: Proceedings of the 7th international conference on Multimodal interfaces*. New York, NY, USA : ACM, 2005, S. 124–128. – ISBN 1-59593-028-0
- [Velichkovsky u. a. 1997] VELICHKOVSKY, Boris ; SPRENGER, Andreas ; UNEMA, Pieter: Towards gaze-mediated interaction: Collecting solutions of the "Midas touch problem". In: *INTERACT '97: Proceedings of the IFIP TC13 Interantional Conference on Human-Computer Interaction*. London, UK, UK : Chapman & Hall, Ltd., 1997, S. 509–516. – ISBN 0-412-80950-8
- [Wahlster 2002] WAHLSTER, W.: SmartKom: Fusion and Fission of Speech, Gestures, and Facial Expressions. In: *Proceedings of the 1st International Workshop on Man-Machine Symbiotic Systems*, 2002, S. 213–225
- [Ware und Mikaelian 1987] WARE, Colin ; MIKAELIAN, Harutune H.: An evaluation of an eye tracker as a device for computer input. In: *SIGCHI Bull.* 17 (1987), Nr. SI, S. 183–188. – ISSN 0736-6906