Hochschule für Angewandte Wissenschaften Hamburg
Hamburg University of Applied Sciences

Projektbericht

Lennard Hamann

Validierung von Konzepten für die multimodale Eingabe

Lennard Hamann

Validierung von Konzepten für die multimodale Eingabe

Projektbericht eingereicht im Rahmen der Veranstaltung Projekt
von Prof. Dr. Kai von Luck
im Studiengang Master Of Science Informatik
am Fachbereich Elektrotechnik und Informatik
der Hochschule für Angewandte Wissenschaften Hamburg

Abgegeben am 29.02.2008

Inhaltsverzeichnis

Tabellenverzeichnis

Abbildungsverzeichnis

1 Einführung

Ein Forschungsbereich des UbiComp-Labors der HAW Hamburg ist die Entwicklung von innovativen Konzepten für die Mensch-Computer Interaktion. Es wird untersucht, wie neue Technologien wie Motion-Tracking, Multitouch-Displays und Eye-Tracking als Eingabemedien verwendet werden können. Des weiteren wird erforscht, wie diese Eingabegeräte zusammengestellt werden können, um dem Benutzer eine multimodale Eingabe zu ermöglichen.

Das Ziel dieser Arbeit ist es, eine Anwendung zu entwickeln, die Blick- und Spracheingabe unterstützt. Mit dieser Anwendung sollen

- die Möglichkeiten, die im UbiComp-Labor verwendete Eyetracking-Hardware als Eingabemedium zu verwenden, untersucht werden.
- die in [Hamann 2008a] vorgestellten Konzepte zur Blickeingabe bzw. zur multimodalen Eingabe analysiert werden

In [Hamann 2007] werden die Grundlagen von Eyetracking-Systemen dargestellt. In [Hamann 2008c] findet sich ein Überblick von Basistechnologien und Konzepten zur multimodalen Interaktion.

In Kapitel 2 wird das Ziel des Projekts beschrieben. Im daran anschließden Kapitel 3 werden der System-Aufbau und die verwendete Software dargestellt. Im abschließenden Kapitel 4 wird die Anwendung beschrieben.

2 Ziel des Projekts

Das Ziel des Projekts ist es, ein System zu erstellen, mit dem Ansätze und Konzepte zur Blickeingabe und zur multimodalen Eingabe untersucht werden können. Dabei geht es vor allem darum, wie eine Standard-Benutzerschnittstelle per Blickeingabe bzw. per Kombination von Blick- und Spracheingabe bedient werden kann.

Um die Untersuchung durchzuführen, wird eine Anwendung entwickelt, mit der Buttons, Menueinträge und andere grafische Elemente per Blick selektiert und per Sprachbefehl manipuliert werden können.

Es wird der für die Nutzung des Eyetrackers als Eingabemedium grundlegende Algorithmus aus Abschnitt 4.1 implementiert und getestet. Mit ihm werden die Messdaten des Eyetrackers geglättet und die Unterscheidung von Fixationen und Sakkaden ermöglicht.

Dann wird eine Anwendung erstellt, die der Benutzer durch Blickeingabe steuern kann. Steuern bedeutet hier, das er unbewegliche Objekte wie Buttons und Menueinträge per Fixation auswählen kann. Hier kann untersucht werden

- welche Zeitspanne optimal ist für den Fixations-Schwellwert zur Selektion; der Benutzer muss die Möglichkeit haben, die Benutzerschnittstelle zu betrachten, ohne Aktionen auszulösen
- welche Abmessungen die zu selektierenden Objekte haben müssen, um per Fixation ausgewählt werden zu können
- wie groß die Abstände zwischen den zu selektierenden Objekte sein müssen, um Fixationen zu einem Objekt zuordnen zu können
- ob eine graphische Repräsention des aktuellen Blickpunkts bei der Selektion von Objekten hilfreich oder störend ist
- welche Abmessungen die graphische Repräsention haben muss
- ob es hilfreich ist, dass selektierte Objekt graphisch hervorzeheben, zu zoomen, zu markieren, bevor es aktiviert wird
- wie durch Glättung der Messdaten der Einfluss von Jittern verringert werden kann

- wie sich andere Einflüsse wie die Lichtverhältnisse, der Abstand zwischen Augen, Eytracker und Monitor und der Kalibrierung auswirken
- ob es benutzerspezifische Muster in den Blickbewegungen gibt, deren Kenntnis den Algorithmus verbessern könnte

Im nächsten Schritt soll die Anwendung um die Möglichkeit zur Spracheingabe erweitert werden, um die kombinierte Eingabe per Blick und Sprache untersuchen zu können. Hierbei wird nur ein begrenzes Vokabular verwendet, um die Befehlserkennung zu vereinfachen. Ein kleines Lexikon ist ausreichend für die zu untersuchenden Fragen, bei denen es vor allem um Probleme der multimodalen Kombination von Blick- und Spracheingabe geht und nicht um Sprachverarbeitung. Es werden nur einfache Befehle wie z.B. 'select' und 'drag and drop' benötigt. Mit dem erweiterten System kann untersucht werden:

- ob die zweite Modalität eine Verbesserung hinsichtlich der Bedienbarkeit bringt
- bei welchen Aufgaben der Benutzer unimodal und bei welchen Aufgaben multimodal interagiert
- wie die zeitlichen Verhältnisse zwischen den beiden Modalitäten sind
- ob die Eingaben über beide Kanäle sequentiell oder simultan erfolgen
- ob es benutzerspezifische Muster in der sequentiellen oder simultanen Benutzung der Kanäle gibt

3 Aufbau des Systems

Abbildung 3.1: Anwendungsszenarien Tobii X120 (Quelle: http://www.tobii.com)

Für diese Arbeit wird der mittlere desktop-ähnliche Versuchsaufbau aus Abbildung 3.1 verwendet. Das Eyetracking-System wird von dem Monitor platziert. Der Abstand zwischen Eyetracker und Anwender sollte etwa 60-70cm betragen. Der Monitor wird auf eine Auflösung von 1024 x 768 eingestellt.

Für weitere Arbeiten ist ein Aufbau mit der PowerWall des UbiComp-Labors wie in Abbildung 3.1 links zu sehen geplant.

3.1 Hardware

3.1.1 Eyetracking-Systeme

Im folgendem werden die im UbiComp-Labor der HAW Hamburg verwendeten Eyetracking-Systeme vorgestellt.

Es werden mit dem Xuuk eyebox2 Eyetracker und dem Tobii X120 Eyetracker zwei berührungslose, kamerabasierte Systeme verwendet. Diese System zeichnen Bilder der Augen auf und erhalten durch Bildverarbeitung (und vorherige Kalibrierung) ihre Daten.

Xuuk eyebox2

Die eyebox2 der Firma Xuuk besteht aus zwei Elementen. Die Kamera-Einheit enthält eine 1,3 Megapixel Bildsensor und LEDs, die LED-Einheit nur LEDs. Sie können per USB2 an einen PC angeschlossen werden.

Abbildung 3.2: xuuk eyebox2 Eyetracker (Quelle: http://www.xuuk.com)

LINSE	BEREICH	ENTFERNUNG
25 mm	25°	>10 m
12 mm	50°	>7 m
8 mm	67°	>4 m
2fach 12 mm	100°	>7 m

Tabelle 3.1: Meßfeld der eyebox2 (Quelle: http://www.xuuk.com)

Das von der LEDs ausgestrahlte Infrarotlicht verursacht eine helle Verfärbung der Pupillen in den Kamerabildern. Durch diese Markierung können die Pupillen während der Bildverarbeitung effektiver und genauer erkannt werden.

Das eyebox2-System liefert laut Hersteller folgende Meßdaten:

- Anwesenheit, Position und Anzahl der Augen im Kamerabild auf eine Entfernung bis zu 10 Metern und
- Anwesenheit, Position und Anzahl der Gesichter im Kamerabild auf eine Entfernung bis zu 10 Metern.

Dabei soll folgende Genauigkeit erzielt werden:

- Blickerkennung: bei einer Entfernung von 3m von der Kamera bis zu 8,5° zu jeder Seite
- ca. 95 %ige Gesichtserkennung
- ca. 85 %ige Blickerkennung

Der Hersteller merkt im Datenblatt an, daß die obigen Daten mit den Lichtbedingungen variieren.

Mit der eyebox2 kann also festgestellt werden, ob der Anwender in eine bestimmte Richtung (eyebox2 Kamera-einheit) sieht. Diese Daten könnten dazu verwendet werden, freihand die Bedienung verschiedener Geräten zu wechseln [Dickie u. a. 2006].

Tobii X120

Der Tobii X120 Eyetracker besteht aus einer Einheit, die zwischen Monitor und Anwender plaziert wird. Das System kann per LAN oder USB2 an einen Computer angeschlossen werden.

Abbildung 3.3: Tobii X120 Eyetracker (Quelle: http://www.tobii.com)

Das Eyetracking-System liefert folgende Messdaten:

- Zeitstempel
- Blickposition relativ zum Monitor (zwei-dimensional)
- Blickrichtung (drei-dimensional)
- Positon der Augen im Raum (in mm, drei-dimensional)
- Positon der Augen relativ zum Eyetracking-System (drei-dimensional)
- Pupillengrößen der Augen
- Gültigkeits-Code für jedes Auge

Die räumliche Genauigkeit beträgt laut Datenblatt 0.5° Abweichung vom Blickpunkt. Es können zwischen 60 und 120 Blickpunkte pro Sekunde erfaßt werden bei einer Latenzzeit von 33ms. Der Abstand zwischen Benutzer und Eyetracker kann bis zu 70cm betragen. Kopfbewegungen können im Raum von 30cm x 22cm x 30cm kompensiert werden.

Mit dem Tobii X120 Eyetracker kann also unter anderem die Blickposition eines Anwenders auf dem Monitor ermittelt werden. Diese Daten könnten zur Steuerung einer EyeMouse [Norris u. a. 1997] verwendet werden.

3.1.2 Mikrophon für die Spracheingabe

Das Mikrophon-Modell stand zum Zeitpunkt der Erstellung diesr Arbeit noch nicht zur Verfügung. Daher können hier keine technischen Details wiedergegeben werden. Für eine stabile Spracherkennung ist ein leistungsfähiges Mikrophon Voraussetzung.

3.1.3 Optisches Ausgabegerät

Es wird ein Standard VGA-Monitor verwendet. Ein weiterer möglicher Versuchsaufbau ist, die PowerWall des Ubicomp-Labors als Ausgabemedium zu verwenden.

3.1.4 Rechner

Die Eyetracking-Hardware, der Monitor und das Mikrophon für die Spracheingabe werden an ein Dell-Latitude Laptop mit 4GB Arbeitsspeicher angeschlossen.

3.2 Software

3.2.1 Betriebssystem

Das verwendete Betriebssystem ist Microsoft Windows XP. Die Entscheidung hierfür wurde getroffen, weil die Tobii-Eyetracker Software und das dazugehörige Software Development Kit (SDK) nur für dieses Betriebssystem zur Verfügung steht.

3.2.2 Entwicklungsumgebung

Zur Erstellung der Anwendung wird die Entwicklungsumgebung Microsoft Visual Studio 2005 mit der Programmiersprache Visual C++ verwendet. Die Kalibrierungs- und Tracking-Software-Komponenten des Tobii X120 Eyetrackers stehen als COM-Schnittstellen zur Verfügung, welche mit dieser Entwicklungsumgebung verwendet werden können.

Die Programmier-Schnittstelle für die xuuk-eyebox2 war zum Zeitpunkt der Berichtserstellung noch nicht bekannt.

3.2.3 Spracherkennug

Aufgrund der übrigen Entwicklungsumgebung bietet sich als Spracherkennugs-Software das Microsoft Speech SDK für die englische Sprache an. Diese stellt eine COM-Schnittstelle zur Verfügung.

3.2.4 Usability-Analyse

Zur Durchführung der Usability-Untersuchung wird die Anwendung Tobii-Studio verwendet. Hiermit lassen sich Verwendungs-Abläufe von Anwendungen aufzeichnen und auswerten.

4 Konzept für eine multimodale Test-Umgebung

Um einfache Operationen wie das Selektieren von Benutzerschnittstellen-Elementen per Blickeingabe durchführen zu können, müssen die Fixationen erkannt werden. Die Software muß unterscheiden können, ob der Blick des Benutzers zur Informationsaufnahme die Oberfläche durchsucht oder ob er bewußt ein Element anblickt, um mit ihm zu interagieren. Um diese Unterscheidung treffen zu können, wird der folgende Algorithmus implementiert.

4.1 Sakkaden-Erkennungs- und Glättungs-Algorithmus

Die Messdaten von Eyetrackern sind verrauscht aufgrund der Physiologie des Auges und Schwächen des Systems. Zudem liefert das im UbiComp-Labor verwendete System Tobii X120 60 bis 120 Blickpunkze pro Sekunde. Beides macht eine Glättung der Daten notwendig.

In [Kumar 2007] wird ein Algorithmus zur Sakkaden-Erkennung und zur Glättung der Messdaten vorgestellt. Der Algorithmus beinhaltet zwei Mengen: erstens die Menge von Blickpunkten, die zur aktuellen Fixation gehören und zweitens die Menge von Blickpunkten, die potentiell die nächste Fixation bilden könnten. Wenn sich ein Blickpunkt räumlich nahe (innerhalb eines Genzwertes) der aktuellen Fixation befindet, wird er zu dieser Menge hinzugefügt. Die neue aktuelle Fixation p_{mean} wird dann durch einen gewichteten Mittelwert berechnet, der die aktuellen Punkte stärker gewichtet als die älteren. Der älteste Punkt wird mit 1 gewichtet, der aktuellste Punkt mit n.

$$p_{mean} = \frac{1p_0 + 2p_1 + ... + np_{n-1}}{1 + 2 + ... + n}$$

Wenn sich der neue Blickpunkt außerhalb des Grenzwertes befindet, wird er zur Menge der Punkte der potentiell nächsten Fixation hinzugefügt und die aktuelle Fixation bleibt unverändert. Wenn sich der folgende Punkt näher an der aktuelle Fixation befindet, wird er zu dieser Menge hinzugefügt und die Menge der potentiell nächsten Fixation wird verworfen.

Wenn sich der folgende Punkt näher zur potentiellen Fixation befindet, wird er zu dieser hinzugefügt und diese wird dann die neue Fixation.

4.2 Anforderungen an die Software

Damit die in [Hamann 2008a] beschriebenen Konzepte mit der Anwendung untersucht werden können, muß diese folgende Anforderungen erfüllen:

- Implementation von Blick- und Spracheingabe
- Implementation der Fixations-Erkennung
- der zeitliche Fixations-Schwellwert muß konfigurierbar sein
- die Abmessungen des Fixations-Bereichs müssen konfigurierbar sein
- die Abmessungen von den zu selektierenden Objekten müssen konfigurierbar sein
- die Abstände zwischen den zu selektierenden Objekten müssen konfigurierbar sein
- die graphische Repräsentatin des Blickpunkts muß ein- und abschaltbar sein
- das Layout der graphischen Repräsentatin des Blickpunkts muss konfigurierbar sein (Abmesssungen)
- der Algorithmus zur Glättung der Messdaten muß auswählbar und parametrisierbar sein
- die zweite Modalität muß ein- und abschaltbar sein; unimodale Eingage muß möglich sein
- Timeout-Dauer, nachdem eine Eingabe über einen Kanal getätigt wurde und auf die Eingabe über den anderen Kanal gewartet wird

4.3 Methoden

Um Aussagen über benutzerspezifische Präferenzen bei der Verwendung von multimodaler Eingabe treffen zu können, müssen mehrere Teilnehmer den Versuch durchführen. Um die Blickeingabe allgemein auf ihre Tauglichkeit hin zu überprüfen, sollte es eine Probandengruppe geben, die die Aufgaben ausschließlich mit dem gewohnten Eingabemedium Maus durchführt.

Ein Versuch beginnt mit der Kalibrierung des Eyetrackers auf den Teilnehmer. Dann sollen verschiedene Auswahl- und Verschiebeoperationen nur mit der Maus, nur mit Blickeingabe und mit einer Kombination von Blick- und Spracheingabe durchgeführt werden. Eine Auswahl mit der Maus erfolgt wie gewohnt durch das Bewegen des Mauszeigers über das zu selektierende Objekt. Eine Auswahl per Blickeingabe erfolgt durch das Fixieren des zu selektierenden Objekts für einen bestimmten Zeitraum oder gegebenenfalls durch einen Sprachbefehl.

4.4 Probleme

Da nur einfache Auswahloperationen durchgeführt werden sollen, ist der Anwender eventuell nicht ausreichend motiviert. Dadurch kann das Versuchsergebnis verfälscht werden. Um eine ’natürlichere’ Testumgebung zu erhalten, ist es eventuell sinnvoll, eine Spiel-Anwendung zu erstellen, die auf Auswahl- und Verschiebeoperationen basiert.

Literaturverzeichnis

[Dickie u. a. 2006] DICKIE, Connor ; HART, Jamie ; VERTEGAAL, Roel ; EISER, Alex: LookPoint: 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

[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: *Multimodale Interaktion*. HAW Hamburg. 2008. – URL http://users.informatik.haw-hamburg.de/~ubicomp/projekte/master07-08-aw/hamann/bericht.pdf

[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

[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

[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