Karlsruher Schriften
zur Anthropomatik

Band 64

Jutta Hild

Nutzung von Blickbewegungen für die Mensch-Computer-Interaktion mit dynamischen Bildinhalten am Beispiel der Videobildauswertung

KIT Scientific Publishing

Jutta Hild

Nutzung von Blickbewegungen für die Mensch-Computer-Interaktion mit dynamischen Bildinhalten am Beispiel der Videobildauswertung

Karlsruher Schriften zur Anthropomatik
Band 64
Herausgeber: Prof. Dr.-Ing. habil. Jürgen Beyerer

Eine Übersicht aller bisher in dieser Schriftenreihe
erschienenen Bände finden Sie am Ende des Buchs.

Nutzung von Blickbewegungen für die Mensch-Computer-Interaktion mit dynamischen Bildinhalten am Beispiel der Videobildauswertung

von
Jutta Hild

KIT Scientific Publishing

Karlsruher Institut für Technologie
Institut für Anthropomatik und Robotik

Nutzung von Blickbewegungen für die Mensch-Computer-Interaktion mit dynamischen Bildinhalten am Beispiel der Videobildauswertung

Zur Erlangung des akademischen Grades eines Doktors der Naturwissenschaften von der KIT-Fakultät für Informatik des Karlsruher Instituts für Technologie (KIT) genehmigte Dissertation

von Jutta Hild

Tag der mündlichen Prüfung: 25. November 2022
Erster Gutachter: Prof. Dr.-Ing. habil. Jürgen Beyerer
Zweiter Gutachter: Prof. Dr.-Ing. Tanja Schultz

Impressum

Scientific Publishing

Karlsruher Institut für Technologie (KIT)
KIT Scientific Publishing
Straße am Forum 2
D-76131 Karlsruhe

www.ksp.kit.edu

Print on Demand 2024 – Gedruckt auf FSC-zertifiziertem Papier

ISSN 1863-6489
ISBN 978-3-7315-1330-8
DOI 10.5445/KSP/1000164414

Kurzfassung

Das Ziel der vorliegenden Arbeit ist zu erforschen, ob die Nutzung von Eyetracking die Benutzungsschnittstelle für Bildfolgenanalyse leistungsfähiger und belastungsärmer machen kann. Dieses Ziel wird vor dem Hintergrund des Anwendungsgebietes der Echtzeit-Verkehrs- und Umweltüberwachung für die zivile und militärische Sicherheit betrachtet. Hier sind die Anforderungen für den Benutzer besonders hoch und vielfältig. Sie sollten daher auch diejenigen anderer Anwendungsgebiete mit dynamischen Szenen wie Geowissenschaften, Medizin oder Videocomputerspiele abdecken. Ausgangspunkt ist das Videoauswertesystem ABUL[1] des Fraunhofer IOSB, das neben der Erfassung, Verarbeitung und Interaktion mit Bildfolgen auch zahlreiche Verfahren zur automatischen Bildanalyse wie Bewegungsdetektion und Einzelobjekttracking bietet.

Der Benutzer ist bei der Bildfolgenanalyse mit dynamischen Bildschirminhalten konfrontiert. Visuelle Wahrnehmung und Kognition sind hier aufgrund der Präsenz von Bewegung stark belastet, insbesondere, wenn sicherheitskritische Entscheidungen unter Zeitdruck getroffen werden müssen. Die Motorik ist stark belastet, wenn bewegte Objekte in der Szene selektiert werden müssen, was mit der traditionellen Mauseingabe manuell anstrengend, mühsam und fehlerträchtig ist.

Beim Eyetracking erfasst ein Eyetracker die Blickbewegungen des Benutzers und liefert so einen Hinweis auf den Fokus seiner visuellen Aufmerksamkeit. Blickbasierte Interaktion gilt daher als intuitiv für Zeigeoperationen, da der Mensch gewöhnlich an die Stelle blickt, an der eine Systemeingabe erfolgt.

[1] **A**utomatisierte **B**ildauswertung für **U**nbemannte **L**uftfahrzeuge

Mit dem Blick als Zeige"gerät" ist keine manuelle Zeigeoperation erforderlich, was blickbasierte Interaktion sehr schnell macht. Insbesondere für die Bewegtobjektselektion in dynamischen Szenen, wo Objekte oft nur begrenzte Zeit sichtbar sind, kann sich der Geschwindigkeitsvorteil blickbasierter Interaktion positiv auswirken. Zudem verliert die Mauseingabe mit zunehmender Objektgeschwindigkeit an Auflösungsvermögen, sodass das generell geringere Auflösungsvermögen von Eyetracking von ca. 2° konkurrenzfähiger wird.

Die vorliegende Arbeit liefert drei Beiträge.

Zur Untersuchung **blickbasierter Selektion bewegter Objekte** werden folgende Interaktionstechniken identifiziert und in Nutzerstudien evaluiert:

Blick+Taste nutzt die vom Eyetracker gelieferte Blickposition zum Zeigen und die ENTER-Taste zur Selektionsauslösung. Sie zeigt sich als Alternative zur Mauseingabe, denn sie erzielt in fünf Nutzerstudien mit insgesamt N=73 Versuchspersonen (darunter 26 Videoanalyseexperten) bei vergleichbar guter Selektionsfehlerquote und Zufriedenstellung eine bis zu 46% kürzere Selektionszeit und wird von einer deutlichen Mehrheit (79%) der Versuchspersonen bevorzugt. Eine Längsschnittstudie (N=4) zeigt, dass sich die Benutzerleistung durch Training verbessert und nach wenigen Monaten stabilisiert; nach 6 Monaten wird die Ermüdung der Augen signifikant geringer bewertet. Alternativ zur ENTER-Taste wird eine Fußtaste zur Selektionsauslösung getestet. Die Ergebnisse sind im Vergleich zu Blick+Taste etwas schlechter.

MAGIC pointing nutzt die vom Eyetracker gelieferte Blickposition zum groben Zeigen und nutzt die Mauseingabe zum feinen Zeigen und zur Selektionsauslösung. Die in der vorliegenden Arbeit neu vorgeschlagene Variante *MAGIC Button* zeigt für Selektionsfehlerquote, Selektionszeit und Zufriedenstellung keine signifikanten Unterschiede zur Mauseingabe und wird von zwei Dritteln der Versuchspersonen bevorzugt.

Blick+EEG nutzt den Blick zum Zeigen und das ereigniskorrelierte Potenzial P300 des EEG (Elektroenzephalogramm) zur Selektionsauslösung. Zusätzlich übernimmt das EEG den kognitiven Teil des menschlichen Entscheidungsprozesses, ob ein Objekt Zielobjekt ist oder nicht. Blick+EEG ist damit eine implizite, passive Eingabemethode, die keinerlei manuelle oder bewusste Aktion

des Benutzers erfordert. Für Blick+EEG wird in zwei Nutzerstudien (N=21) die prinzipielle Machbarkeit in einer Offline-Analyse gezeigt. Im Rahmen der vorliegenden Arbeit wird die Analyse der räumlichen Lokalisation anhand der Blickdaten durchgeführt; die zeitliche EEG-Lokalisation auf Basis des ereigniskorrelierten Potenzials P300 wird von EEG-Experten der Uni Bremen beigetragen.

Ein zweiter Beitrag betrachtet die **blickbasierte Interaktion bei automatischen Bildanalyseverfahren**. Solche Verfahren dienen dem Benutzer als Assistenz zur Aufmerksamkeitssteuerung. Zum einen wird die Auswirkung der Verfügbarkeit automatischer Verfahrensergebnisse eines Bewegungsdetektionsverfahrens (Independent Motion Detection, IMD) auf die Leistung und Belastung des Benutzers bei der Bewegtobjektmarkierung untersucht. Bei Informationsfusion aus Detektionsleistung des Verfahrens und Erkennungsleistung des menschlichen Experten ist die Effektivität deutlich höher und die Zufriedenstellung sehr gut. Zum anderen bestätigt sich in zwei Nutzerstudien (N=24, darunter 18 Videoanalyseexperten) zur Aufgabe der Initialisierung eines Einzelobjekttrackingverfahrens Blick+Taste als leistungsfähige, belastungsarme Alternative zur Mauseingabe.

Der dritte Beitrag untersucht die **blickbasierte Klassifikation der Benutzertätigkeit bei der Bildfolgenanalyse**, die bspw. dazu dienen könnte, Systemfunktionen automatisch aufzurufen, ohne dass der Benutzer bewusst interagieren muss. Wir untersuchen die Klassifikation der vier typischen Tätigkeiten Erkunden, Beobachten, Suchen und Verfolgen anhand von aufgezeichneten Blickdaten von 30 Versuchspersonen. Die Klassifikation wird für unterschiedlich lange Segmente (4 min bis 3 s) der Blickdatenprotokolle durchgeführt. In allen Fällen gelingt die Klassifikation deutlich über Zufallswahrscheinlichkeit, sie ist jedoch nicht gut genug, um darauf basierend Systemeingaben zu tätigen.

Danksagung

Diese Arbeit entstand am Fraunhofer-Institut für Optronik, Systemtechnik und Bildauswertung (IOSB) in Kooperation mit dem Lehrstuhl für Interaktive Echtzeitsysteme (IES) des Karlsruher Instituts für Technologie (KIT).

Mein besonderer Dank gilt Herrn Prof. Dr.-Ing. habil. Jürgen Beyerer für die Betreuung dieser Arbeit und seine zahlreichen wertvollen Anregungen dazu. Ebenfalls mein besonderer Dank gilt Frau Prof. Dr.-Ing. Tanja Schultz für die Übernahme des Korreferats.

Mein besonderer Dank geht zudem an meinen Gruppenleiter Dr.-Ing. Michael Voit sowie meine Abteilungsleiterin Dr. Elisabeth Peinsipp-Byma für das in mich gesetzte Vertrauen und dafür, dass sie mir stets mit Rat und Tat zur Seite standen. Ihnen und Dr. Jürgen Geisler verdanke ich die Inspiration zu diesem Thema sowie die Finanzierung durch fachlich verwandte Forschungsprojekte. In diesem Zusammenhang geht mein Dank auch an die Wehrtechnische Dienststelle für Informationstechnologie und Elektronik (WTD 81) der Bundeswehr, die diese Forschungsprojekte förderte.

Ich danke zudem der Fraunhofer-Gesellschaft e.V. für die Förderung im Rahmen des Personalentwicklungsprogramms TALENTA speed up, die mir zusätzlichen Freiraum für die Anfertigung dieser Arbeit verschaffte.

Besonders zu Dank verpflichtet bin ich etlichen Angehörigen der Bundeswehr für Anregungen aus ihrer Praxiserfahrung sowie für die Unterstützung zweier Nutzerstudien, sowohl organisatorisch als auch durch die Teilnahme als Versuchspersonen.

Mein besonderer Dank gilt zudem Prof. Dr.-Ing. Tanja Schultz und Dr.-Ing. Felix Putze von der Uni Bremen für die bereichernde und vertrauensvolle Zusammenarbeit im Themenbereich Blick+EEG.

Herzlichen Dank meinen Kolleg:innen am Fraunhofer IOSB und den zahlreichen Student:innen, die durch fachliche Diskussionen bzw. mit ihren Abschlussarbeiten oder als studentische Hilfskräfte zu dieser Arbeit beigetragen haben.

Herzlichen Dank auch an Dr.-Ing. Hermann Hild für das sorgfältige Korrekturlesen dieser Niederschrift, an Dr.-Ing. Mathias Anneken für die Unterstützung im Umgang mit der LaTeX-Vorlage sowie an Dr. Tim Zander und Corinna Claassen für die Unterstützung bei der finalen Durchsicht.

Zuletzt ein herzliches Dankeschön an meine Familie, alle Freund:innen und Kolleg:innen dafür, dass sie mein Leben bereichern, jede:r auf eine individuelle und besondere Weise.

Inhaltsverzeichnis

Notation

In diesem Kapitel werden die in dieser Arbeit verwendeten Notationen und Symbole eingeführt.

Allgemeine Notationen

Skalar	kursive lateinische und griechische Buchstaben	x, T, α
Vektor	fette kleine lateinische Buchstaben	$\mathbf{r}, \mathbf{f}$

Spezielle Notationen

$\hat{x}$	Schätzwert
$\bar{x}$	empirischer Mittelwert (arithmetisches Mittel) einer Stichprobe
$\bar{x} \pm s$	empirischer Mittelwert ± Standardabweichung einer Stichprobe
N	Anzahl Versuchspersonen
α	Signifikanzniveau eines statistischen Tests
t	t-Wert, Teststatistik eines t-Tests
F	F-Wert, Teststatistik einer Varianzanalyse (ANOVA)

p	p-Wert, Signifikanzwert eines statistischen Tests
x, y	Koordinaten

1 Einleitung

1.1 Ausgangssituation und Zielsetzung

Die vorliegende Arbeit fällt in den Informatik-Teilbereich Mensch-Maschine-Interaktion (kurz MMI, engl. *Human-computer interaction* oder kurz *HCI*). Sie gehört zur angewandten Informatik und stützt sich neben der Informatik auf weitere Disziplinen, darunter Psychologie, Kognitionswissenschaften und Ergonomie [Mac12].

Besonders wichtig ist die Ergonomie (engl. *Human Factors*). Ihr Ziel sind effiziente, sichere, komfortable und angenehme (technische) Systeme unter Berücksichtigung der Fähigkeiten, Grenzen und Leistung des Menschen. Die MMI verfolgt dasselbe Ziel für Computersysteme, indem zugeschnitten auf die Arbeitsaufgabe eine geeignete Benutzungsschnittstelle realisiert wird.

Gegenstand der Betrachtung in der vorliegenden Arbeit ist die MMI für die Aufgabe der *Bildfolgenanalyse*. Bildfolgenanalyse kommt in unterschiedlichen Anwendungsdisziplinen vor, etwa bei der Verkehrs- und Umweltüberwachung für die zivile und militärische Sicherheit, in den Geowissenschaften, der Medizin, der Produktion, aber auch bei Videocomputerspielen oder wissenschaftlichen Simulationen [Has11].

Philipson [Phi97] definiert *Bildanalyse* (engl. *Image Analysis*) als „(…) den Prozess, bei dem Menschen und/oder Maschinen fotografische Bilder und/oder digitale Daten untersuchen zum Zwecke der Identifikation von Objekten und

der Beurteilung ihrer Bedeutsamkeit."[1]. Die *Bildfolgenanalyse* unterscheidet sich davon dadurch, dass der Prozess nicht auf ein einzelnes (Stand-)Bild angewendet wird, sondern in einer Bildfolge erfolgt. Damit verbunden ist, dass die zu interpretierenden Bildinhalte nicht statisch und unbewegt, sondern dynamisch und bewegt sind.

Die Definition von Philipson stammt aus dem Umfeld der Geowissenschaften, trifft aber auch in anderen Anwendungsdisziplinen zu: In der Verkehrs- oder Umweltüberwachung, wenn ein Videobildauswerter eine Personen- oder Fahrzeugsuche durchführt; in der Medizin, wenn ein Arzt eine Ultraschallaufnahme interpretiert; in der Produktion, wenn ein Werker in der Qualitätskontrolle über einen Bildschirm Bauteile auf Fehler und Schäden inspiziert.

In allen genannten Anwendungsfällen nutzt der menschliche Experte heutzutage computerisierte Unterstützung. Der Systemaufbau ist üblicherweise ein Desktop-Computer-Einzelarbeitsplatz, der ortsfest an einem Tisch aufgebaut ist. Die Benutzungsschnittstelle ist eine *grafische Benutzungsoberfläche* (engl. *graphical user interface*, kurz *GUI*).

Grafische Benutzungsoberflächen bestehen aus zwei ineinander verwobenen Komponenten, der *Informationsausgabe* und der *Informationseingabe*. Die Informationsausgabe liefert dem Benutzer eine Sicht auf den aktuellen Systemzustand, der auf einem oder mehreren Bildschirmen visualisiert wird. Die Informationseingabe ermöglicht dem Benutzer, auf das System einzuwirken und den Systemzustand zu verändern. Dabei steuert der Benutzer den Systemzustand über Interaktionselemente auf dem Bildschirm mithilfe der typischen Eingabegeräte Computertastatur und Computermaus.

Im Vergleich zu Anwendungen mit ausschließlich statischem Bildschirminhalt ist der Benutzer bei der Interaktion mit dynamischen Bildschirminhalten mit besonderen Herausforderungen konfrontiert. Die visuelle Wahrnehmung ist aufgrund der Präsenz von Bewegung deutlich höher belastet. Dasselbe gilt für Kognition und Aufmerksamkeit, insbesondere, wenn sicherheitskritische

[1] „Image analysis: the process by which humans and/or machines examine photographic images and/or digital data for the purpose of identifying objects and judging their significance."

Entscheidungen unter Zeitdruck getroffen werden müssen. Auch die Motorik ist höher belastet, denn es ist mit einer manuellen Eingabetechnik wie der Mauseingabe deutlich schwieriger, ein bewegtes Objekt auf einem Bildschirm zu selektieren als ein unbewegtes [Has11].

Die technische Entwicklung ermöglicht heute auf die Anwendung zugeschnittene Software sowie leistungsfähige Rechner und Bildschirme, die die Bildfolgen in hoher Qualität präsentieren. Mehr und mehr stehen auch automatische Verfahren zur Bildfolgenanalyse zur Verfügung, die der Anwender als Assistenz zu Rate ziehen kann. Die Bandbreite, mit der die Informationsausgabe des Computersystems Information in Richtung des menschlichen Benutzers liefern kann, ist heutzutage also enorm.

Demgegenüber ist die Bandbreite der Informationseingabe in Richtung des Computersystems deutlich geringer, wenn der menschliche Benutzer nur die oben genannten traditionellen Eingabegeräte zur Verfügung hat. Zudem wurde die Computermaus für Zeige- und Selektionsoperationen an unbewegten GUI-Elementen entwickelt, wofür sie sich als pixelgenaue Eingabetechnik etabliert hat; Zeige- und Selektionsoperationen an bewegten Objekten sind mit einer indirekten Eingabetechnik wie der Mauseingabe jedoch mühsam und fehlerträchtig.

Diese Ausgangssituation ist der Anknüpfungspunkt der vorliegenden Arbeit.

Das Ziel ist zu erforschen, ob die Nutzung von Eyetracking auf Seiten der Informationseingabe einer Benutzungsschnittstelle für Bildfolgenanalyse leistungsfähiger und belastungsärmer für den Benutzer machen kann. Eyetracking bietet sich als Lösungsansatz deshalb an, weil die Bildfolgenanalyse im Kern eine visuelle Aufgabe ist. Daher liegt die Annahme nahe, dass die Benutzungsschnittstelle leistungsfähiger wird, wenn das Computersystem über Information bezüglich des aktuellen Aufmerksamkeitsfokus des Benutzers verfügt.

Im Folgenden wird in Abschnitt 1.2 die Problemstellung im Detail beschrieben, in Abschnitt 1.3 der Lösungsansatz.

1.2 Problemstellung

Alle oben genannten Anwendungsgebiete von Bildfolgenanalyse haben ihre spezifischen Herausforderungen für den jeweiligen menschlichen Experten. Besonders vielfältig sind die Anforderungen jedoch bei der *Echtzeit-Verkehrs- bzw. Umweltüberwachung*, sodass anzunehmen ist, dass diese auch die Anforderungen der anderen Anwendungsgebiete abdecken.

Zudem bestand während der Durchführung der vorliegenden Arbeit enger Kontakt zu Experten in militärischer Videobildauswertung. Mit ihnen konnte zum einen die Relevanz der geplanten Forschungsarbeiten diskutiert und abgeklärt werden. Zum anderen konnten für einige der Nutzerstudien Videoanalyseexperten als Versuchspersonen gewonnen werden.

Daher wird das Ziel einer leistungsfähigen und belastungsarmen Benutzungsschnittstelle für die Bildfolgenanalyse vor dem Hintergrund dieses Anwendungsgebiets betrachtet.

Ausgangspunkt für die Formulierung der Problemstellung ist das Videoauswertesystem ABUL des Fraunhofer IOSB [Hei08, Hei10, IOS21]. Die ABUL-Software ermöglicht die Erfassung, Aufzeichnung, Verarbeitung, Anzeige und Interaktion mit Bildfolgen. Zusätzlich bietet sie zahlreiche Verfahren zur automatischen Bildanalyse als Systemfunktionen an. Dazu gehören Verfahren zur Bildverbesserung wie Videostabilisierung und Superresolution, Mosaicking sowie Bewegungsdetektion und Einzelobjekttracking.

Zur Informationsdarstellung nutzt ABUL ein Doppelmonitorsystem, Eingabegeräte sind standardmäßig Computermaus und Computertastatur. Der erste Monitor zeigt das Videofenster mit allen verfügbaren Videoströmen an und ermöglicht Echtzeit-Videoanalyse, Abb. 1.1 zeigt die GUI. Der zweite Monitor zeigt ein Kartenfenster und ermöglicht die Anzeige von Bildverarbeitungsergebnissen sowie Nachbearbeitung, Exportfunktionalität und den Zugriff auf externe Software.

Relevant für die vorliegende Arbeit ist nur der erste Monitor, da nur dort mit dynamischen Bildinhalten interagiert wird. Groß und zentral ist das Fenster

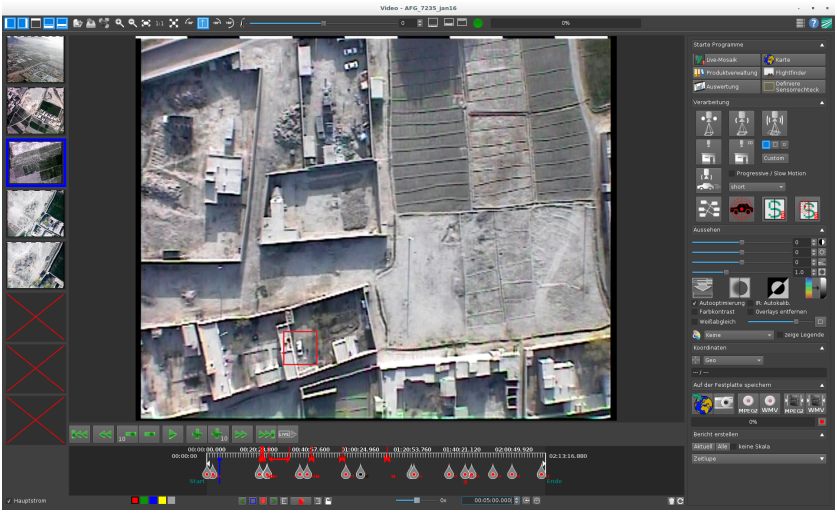


Abbildung 1.1: Benutzungsoberfläche des Videoauswertesystems ABUL. Im Hauptvideostrom wurde ein Fahrzeug mit einem roten Rahmen als bedeutsam markiert.

platziert, das den Hauptvideostrom anzeigt. In diesem Fenster führt der Videobildauswerter die Bildanalyseaufgabe durch: Er untersucht das Bildmaterial auf bedeutsame Objekte (vgl. Definition auf S. 2) und markiert diese mit einem roten Rahmen, indem er einen linken Mausklick auf dem Objekt ausführt.

Rechts des Hauptvideostrom-Fensters sind die Schaltflächen der Systemfunktionen platziert, einschließlich der automatischen Bildanalyseverfahren. Aktiviert der Videobildauswerter beispielsweise das Verfahren Einzelobjekttracking durch Selektion des entsprechenden Icons zur Nutzung, so bewirkt ein linker Mausklick auf ein Objekt im Hauptvideostrom-Fenster keine Markierung mit einem roten Rahmen, sondern die Initialisierung des Einzelobjekttracking-Verfahrens für dieses Objekt.

Unterhalb des Hauptvideostroms befinden sich weitere Schaltflächen zur Interaktion mit dem Hauptvideostrom-Fenster für den Fall, dass nicht in einer Echtzeit-Bildfolge, sondern mit einer archivierten Bildfolge interagiert wird.

Die Fähigkeit von ABUL, auch archivierte Videos abzuspielen, wurde in der vorliegenden Arbeit genutzt, indem die ABUL-Software die Basis für die Versuchssoftware verschiedener Experimente war.

Links des Hauptvideostrom-Fensters werden die Nebenvideoströme visualisiert. Sie liefern – je nach Sensorausstattung – Ausschnitte des Hauptvideostroms in verschiedenen Zoomstufen oder auch die Ansicht einer Frontkamera. Will der Videobildauswerter in einem der Nebenströme die Bildfolge analysieren, so selektiert er durch Mausklick (linke Maustaste) den gewünschten Nebenstrom, wodurch dessen Anzeige unmittelbar in das große zentrale Hauptvideostrom-Fenster wechselt.

Führt der Videoanalyseexperte einen Auftrag zur Bildfolgenauswertung durch, so ist er mit mehreren Herausforderungen konfrontiert, die im Folgenden beschrieben sind: die Art der visuellen Präsentation der Welt in der Szene, die Interaktion mit bewegten Objekten in der Szene sowie die psychische Belastung.

1.2.1 Art der visuellen Präsentation der Welt in der Szene

Die erste Herausforderung liegt darin, dass die Objekte in der Szene von ihrem bekannten Aussehen in der Realität in mehrerlei Hinsicht abweichen können. Folgende Charakteristika sind hierbei typisch:

- Je nach Ausrichtung des Sensors liefern die dargestellten Szenen eine Sicht auf die Welt, die sich von der normalen visuellen Alltagswahrnehmung unterscheidet. Objekte werden bspw. in ungewohnter Perspektive wie Schrägsicht oder Draufsicht präsentiert.
- Je nach Sensor werden die Bilder nur als Grauwertbilder oder nicht farbecht präsentiert, sodass die Objekte anders aussehen als in der Realität.
- Je nach Auflösung des Sensors und seinem Abstand vom abgebildeten Objekt weichen im Bild Objektgröße und Objektschärfe

von den Gegebenheiten der visuellen Alltagswahrnehmung ab. Sehr kleine Objekte oder solche, die sich vom Hintergrund nicht ausreichend kontrastreich abheben, sind schwierig zu detektieren.

- Zu geringe Übertragungsbandbreite kann zu unscharfen Bildern und Artefakten führen.

Aufgrund dieser Gegebenheiten sind Objektdetektion, Objektidentifikation, aber auch Objektverfolgung erschwert und erfordern die kontinuierliche, hohe Aufmerksamkeit des menschlichen Beobachters auf dem Objekt. Visuelle Wahrnehmung und Kognition werden stark belastet.

1.2.2 Interaktion mit bewegten Objekten in der Szene

Die zweite Herausforderung betrifft die Art und Weise, mit der der Videoanalyseexperte als bedeutsam identifizierte Objekte dokumentiert, sodass das Analyseergebnis als Basis für weitere Aktionen bezüglich des Objekts genutzt werden kann. Dazu muss er in der Bildfolge interagieren.

Ein Beispiel ist eine Verkehrsüberwachung, bei der ein bestimmtes Fahrzeug gefunden werden muss, bspw. ein Lieferwagen mit besonderem Aufbau, der mit Ansicht von oben besonders gut zu erkennen ist. Die Kontrolleure an den Straßen erhalten Unterstützung durch einen Videoanalyseexperten am Überwachungsmonitor, der das Kamerabild einer Drohne, die die Straße erfasst, nach potenziellen Zielfahrzeugen durchmustert. Entdeckt er eines, markiert er es in der Bildfolge und übermittelt das annotierte Bild an die Kontrolleure.

Die Markierung erfolgt als Selektionsoperation mit Mauseingabe: Der Mauszeiger wird auf das potenzielle Zielfahrzeug positioniert und beim Drücken der linken Maustaste setzt das Videoauswertesystem einen Rahmenmarker geeigneter Größe. Die Mauszeigerposition definiert dabei den Mittelpunkt des Rahmens; die Rahmengröße ist vorab eingestellt, denn aus der Flughöhe lässt sich die Größe des Lieferwagens im Bild ableiten.

Je nach Situation kann eine solche manuelle Selektionsoperation in einer Bildfolge aufgrund folgender Objekteigenschaften schwierig sein:

- Zeitlich begrenzte Sichtbarkeit des Objekts in der Bildfolge:

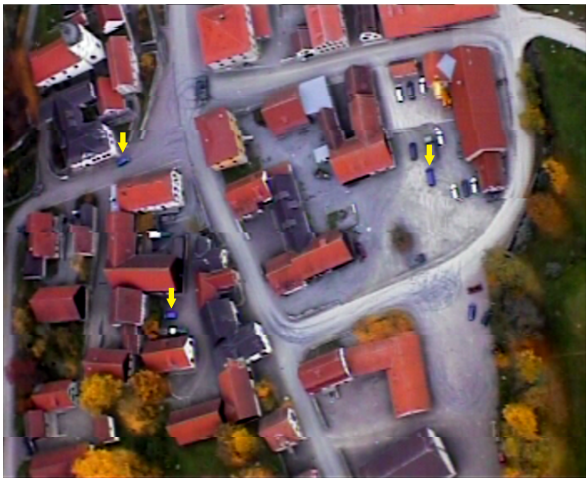
  Dies gilt für visuelle Objekte, die von einem bewegten Sensor (z. B. an ein Fluggerät montiert) erfasst werden und auch für bewegte Objekte wie Fahrzeuge und Personen, die sich durch die Szene eines ortsfest installierten Sensors bewegen. Als Konsequenz ergibt sich eine Begrenzung der Zeitdauer, innerhalb der der Objekt-Markierungsvorgang abgeschlossen sein muss.

- Objekt-Größe, -Geschwindigkeit und -Bewegungstrajektorie:

  Objekte sind oft klein, schnell und/oder ändern unvorhersehbar ruckartig schnell ihre Bewegungstrajektorie im Bild – aus Eigeninitiative oder wenn eine Windböe Fluggerät und Sensor erfasst. Daher gelingt die Markierung mit der Computermaus oft nicht beim ersten Versuch.

  Wiederholte Markierungsversuche können nicht nur manuell, sondern auch perzeptiv und kognitiv herausfordernd und anstrengend sein. Denn der Benutzer muss den Mauszeiger vor jedem Markierungsversuch neu korrekt auf das Objekt platzieren, was ein Aufteilen der visuellen Aufmerksamkeit zwischen Objekt und Mauszeiger erfordert. Sind in der Umgebung noch viele andere Objekte vorhanden, kann es passieren, dass die Markierung ein falsches Objekt trifft und rückgängig gemacht werden muss. Im schlechtesten Fall gelingt die Markierung des bedeutsamen Objekts gar nicht, weil es inzwischen die Szene bereits verlassen hat.

Abb. 1.2 zeigt ein Beispiel, in dem als Videoauswerteauftrag blaue Lieferwagen selektiert werden müssen. Im oberen Bild, Abb. 1.2a, sind drei Fahrzeuge zu sehen, die der Spezifikation entsprechen (in der Abbildung zur Verdeutlichung mit gelben Pfeilen versehen). Im unteren Bild, Abb. 1.2b, ist zu sehen, dass der Benutzer zwei Fahrzeuge mit einem Rahmen markiert hat. Die Markierung des dritten Fahrzeugs ist nicht mehr möglich, da der Sensor es zu diesem Zeitpunkt nicht mehr erfasst.



(a) Ausgangssituation mit drei Zielobjekten (gelbe Pfeile als Hinweis).

(b) Endsituation mit zwei markierten Zielobjekten. Das dritte ist nicht mehr selektierbar, da es vom Sensor nicht mehr erfasst wird.

Abbildung 1.2: Beispielhafter Videoauswerteauftrag „Finde (und markiere) blaue Lieferwagen“.

Neben den besonderen Objekteigenschaften besteht ein weiteres Problem in der generell großen manuellen Belastung, wenn der Benutzer mit der Computermaus über lange Zeit hinweg selektiert. Gesundheitliche Beschwerden sind unter dem Fachbegriff des RSI-Syndroms bekannt („Repetitive Strain Injury“) und weit verbreitet [Ric97, Zha99, Szu00, Tho14]. Betroffen von Schmerzen sind besonders die oberen Extremitäten, vor allem Handgelenk, Finger und Ellenbogen, aber auch der Nacken- und Schulterbereich.

Bei typischen Desktop-Anwendungen wie Büro- oder Programmierarbeiten, aber auch bei der Einzelbildanalyse mit Systemen wie ERDAS Imagine[1] bewirken viele kleine Zeigerverschiebebewegungen, viele Mausklicks, Mausziehbewegungen und häufige Mausrad-Rollbewegungen einseitige Bewegungsmuster, besonders bei unnatürlicher Körperhaltung, die bei Ermüdung häufig ist.

Das RSI-Syndrom wurde auch von den militärischen Videobildauswertern berichtet, zu denen während der Arbeiten an der vorliegenden Dissertation Kontakt bestand. Sie bestätigten auch die generelle Herausforderung der manuellen Selektion bewegter Zielobjekte in Bildfolgen.

1.2.3 Psychische Belastung

Eine dritte Herausforderung besteht in der hohen psychischen Belastung, die folgende Aspekte beinhaltet:

- Bei der Verkehrs- und Umweltüberwachung sind häufig Sicherheitsaspekte tangiert, sodass hier eine besondere Tragweite der Konsequenzen entsteht, falls ein Objekt fälschlicherweise nicht als bedeutsam erkannt, dokumentiert und behandelt wird. Dies bedeutet hohen psychischen Druck.

[1] Softwaresystem zur Analyse von Fernerkundungsdaten für Anwendungsfelder wie Bodenanalyse, Geologie, Land- und Forstwirtschaft, Archäologie, Stadtplanung, Umweltüberwachung oder Überwachung der zivilen und militärischen Sicherheit,
siehe https://www.hexagongeospatial.com/products/power-portfolio/erdas-imagine

- Über- oder Unterforderung der Aufmerksamkeit können zu gedanklichem Abschweifen führen. Überforderung kann beispielsweise bei besonders reichhaltigem visuellem Input vorkommen. Sind bspw. sehr viele bewegte Objekte in der Szene zu sehen, von denen die meisten nicht relevant sind, so ist es mühsam und aufwändig, das Zielobjekt in der Masse zu identifizieren. Unterforderung entsteht, wenn das Ereignisaufkommen über einen längeren Zeitraum zu niedrig war. Die Konsequenz ist dieselbe: Ist die Aufmerksamkeit zu lange abgewichen, kann das bedeutsame Objekt inzwischen die Szene verlassen haben und seine Dokumentation wird verpasst.

1.2.4 Fazit

Die Aufgabe der Bildfolgenanalyse bei der Echtzeit-Verkehrs- und Umweltüberwachung ist sowohl perzeptiv als auch kognitiv, motorisch und psychisch herausfordernd.

Allen beschriebenen Herausforderungen kann dadurch begegnet werden, dass die Benutzungsschnittstelle dem Benutzer ermöglicht, seine Aufmerksamkeit möglichst kontinuierlich und ungeteilt auf die Bildfolge und die als relevant erkannten Objekte gerichtet zu lassen.

Die Präsenz von Bewegung als zusätzlichem visuellem Merkmal und die begrenzte zeitliche Sichtbarkeit von Objekten in der Szene erfordern zudem eine Benutzungsschnittstelle, die die Selektion bewegter Objekte schneller und belastungsärmer erlaubt als die manuelle Mauseingabe.

1.3 Lösungsansatz

Der Lösungsansatz für das Ziel einer leistungsfähigeren und belastungsärmeren Benutzungsschnittstelle für die Aufgabe der Bildfolgenanalyse besteht in der vorliegenden Arbeit darin, Eyetracking auf Seiten der Informationseingabe zu nutzen. Eyetracking erfasst die Blickbewegungen des Benutzers und

liefert damit einen Hinweis, worauf der Benutzer seine visuelle Aufmerksamkeit richtet.

Die Beiträge, die die vorliegende Arbeit dazu leistet, sind in den folgenden drei Abschnitten beschrieben. Der umfangreichste Beitrag adressiert das Thema „Blickbasierte Interaktion für Selektionsoperationen an bewegten Objekten“ (Abschnitt 1.3.1). Der zweite Beitrag adressiert das Thema „Blickbasierte Interaktion bei automatischen Bildanalyseverfahren“ (Abschnitt 1.3.2). Der dritte Beitrag adressiert das Thema „Blickbasierte Klassifikation der Benutzertätigkeit bei der Bildfolgenanalyse“ (Abschnitt 1.3.3).

1.3.1 Blickbasierte Interaktion für Selektionsoperationen an bewegten Objekten

Dieser Beitrag zielt darauf ab, die Selektion bewegter Objekte im Vergleich zur traditionellen Mauseingabe bezüglich folgender, in der Problemstellung adressierter Aspekte zu verbessern:

(A) Die Aufmerksamkeit des Benutzers soll kontinuierlich auf der Bildfolge und den zu analysierenden Objekten liegen können.

(B) Die Selektion soll schneller durchführbar sein.

(C) Die Selektion soll mit geringerer Belastung – sowohl kognitiv als auch manuell – durchführbar sein.

Dabei wird im ersten Schritt in einem Konzept (Kapitel 3) erarbeitet, welche Ausprägungen blickbasierter Interaktion für die betrachtete Anwendung der Bewegtobjektselektion angemessen erscheinen.

Im zweiten Schritt werden diese Ausprägungen dann in Nutzerstudien anhand abstrakter, teilweise standardisierter Selektionsaufgaben im Vergleich zur Mauseingabe evaluiert (Kapitel 5). In Abschnitt 5.1 werden explizite, aktive Blickinteraktionstechniken betrachtet, in Abschnitt 5.2 eine implizite, passive. Die Nutzung abstrakter Versuchsaufgaben erlaubt gut kontrollierte Experimente mit hoher intrinsischer Validität.

Im dritten Schritt wird die leistungsfähigste Ausprägung anhand praxisnaher Selektionsaufgaben bewegter Objekte in Full Motion Video-Datenmaterial im Vergleich zur Mauseingabe evaluiert (Kapitel 6, Abschnitt 6.1). Der höhere Realitätsbezug (Datenmaterial, Zielobjekte) dieser Versuchsaufgaben zur Bildfolgenanalyse bzw. Videobildauswertung, wie sie in der Praxis vorkommt, macht die Experimente weniger gut kontrollierbar, liefert dafür jedoch höhere externe Validität.

1.3.2 Blickbasierte Interaktion bei automatischen Bildanalyseverfahren

Dieser Beitrag betrachtet die Nutzung automatischer Bildanalyseverfahren (Kapitel 6, Abschnitt 6.2) in Kombination mit blickbasierter Interaktion.

Die zentrale Rolle bei der Bildfolgenanalyse hat der menschliche Experte. Denn da der Mensch zusätzlich zu seinen visuellen Fähigkeiten über Erfahrungswissen bezüglich der Aufgabe verfügt, trifft er stets die finale Entscheidung bezüglich der Relevanz von Objekten oder Ereignissen. Automatische Bildanalyseverfahren, wie das ABUL-System sie bietet, haben die Rolle einer Assistenz und sind dank ihrer heutigen Zuverlässigkeit eine wertvolle Unterstützung [Hof13].

Im Rahmen dieses Beitrags werden automatische Verfahren zur Bewegungsdetektion und zum Einzelobjekttracking betrachtet. Sie können den Benutzer dabei unterstützen, bewegte Objekte zu finden bzw. sie nicht aus den Augen zu verlieren. Sie dienen dadurch der Aufmerksamkeitssteuerung und reduzieren die perzeptive und kognitive Belastung des Benutzers.

Gleichzeitig produzieren die Verfahrensergebnisse jedoch zusätzlichen visuellen Input, der vom Benutzer wahrgenommen werden muss, was die visuelle Perzeption wiederum zusätzlich belastet. Außerdem entsteht durch die Nutzung der automatischen Verfahren zusätzlicher Interaktionsaufwand.

Um die Wirkung automatischer Verfahrensergebnisse zu erforschen, werden Nutzerstudien anhand von Full Motion Video-Datenmaterial durchgeführt, die wie die Untersuchungen des Beitrags aus Abschnitt 1.3.1 die Aufgabe der

Bewegtobjektselektion beinhalten. Wie dort im dritten Schritt werden jeweils die leistungsfähigste blickbasierte Interaktionstechnik und die Mauseingabe verglichen.

Für ein automatisches Verfahren zur *Bewegungsdetektion* werden Leistungsfähigkeit und Belastung des Benutzers für die Aufgabe der Bewegtobjektmarkierung *mit* und *ohne* automatische Verfahrensergebnisse untersucht. Die Bedingung *mit* automatischen Verfahrensergebnissen beschreibt einen Fall von Informationsfusion aus Detektionsleistung des Verfahrens und Erkennungsleistung des menschlichen Experten.

Für ein automatisches Verfahren zum *Einzelobjekttracking* wird die Leistungsfähigkeit des Benutzers für die Selektionsaufgabe der Initialisierung des Verfahrens für ein bewegtes Objekt untersucht.

1.3.3 Blickbasierte Klassifikation der Benutzertätigkeit bei der Bildfolgenanalyse

Betrachtet man die GUI des ABUL-Systems, so gibt es zahlreiche Schaltflächen rechts vom Hauptvideostrom, mit denen Systemfunktionen wie bspw. die genannten automatischen Bildanalyseverfahren aufgerufen werden können. Je nachdem welches ausgewählt ist, bewirkt eine Selektion im Hauptvideostrom-Fenster eine andere Systemreaktion:

- Ist keine Systemfunktion ausgewählt, wird durch eine Selektion eines Objekts das Objekt wie oben beschrieben durch einen roten Rahmen markiert. Ist das Verfahren zur Bewegungsdetektion aktiviert, bewirkt eine Selektionsoperation ebenfalls eine Rahmenmarkierung.
- Ist das Verfahren für Einzelobjekttracking selektiert, bewirkt die Selektion eines Objekts, dass ein automatisches Trackingverfahren auf das Objekt aufgesetzt wird.

Um eine Systemfunktion zu aktivieren, muss der Benutzer also seine Aufmerksamkeit zumindest kurz auf die Schaltfläche richten und diese selektieren. Diese explizite Aktion wäre unnötig, wenn es möglich wäre, die Intention

des Benutzers zu schätzen. Will er ein Objekt markieren oder ein Einzelobjekttracking initialisieren? Die Intention ist wiederum abhängig von der Aufgabe bzw. der Tätigkeit, die der Benutzer aktuell durchführt (z. B. Objekt suchen, um es zu markieren, oder Objekt verfolgen, um einen Einzelobjekttracker aufzusetzen).

Gelänge eine Klassifikation der Tätigkeit, könnte man daraus auf die Intention schließen und bei einer Selektion entsprechend entweder einen Rahmen platzieren oder einen Objekttracker aufsetzen, ohne dass zuvor die entsprechende Schaltfläche zu selektieren wäre.

Blickbasierte Intentionsschätzung wird in der MMI im Zusammenhang mit der Realisierung passiver Systemeingaben betrachtet. Meist geht es darum, genau an der Blickposition eine bestimmte Funktion auszuführen. Die Klassifikation komplexerer Tätigkeiten auf der Basis komplexeren Blickverhaltens ist herausfordernd und wird aktuell vor allem von Forschungsarbeiten der Kognitionspsychologie betrachtet. Diese Untersuchungen nutzen überwiegend statische Szenen, die von Versuchspersonen unter verschiedenen Aufgabenstellungen betrachtet werden, bspw. einmal mit dem Ziel, die Szeneninhalte zu memorieren, und einmal mit dem Ziel, bestimmte Objekte im Bild zu finden. Auf Basis des Blickverhaltens wird dann versucht, die Aufgaben „Memorieren“ bzw. „Suchen“ zu klassifizieren.

Im Rahmen der vorliegenden Arbeit wird ein erster Schritt unternommen, blickbasiert Aufgaben bzw. Tätigkeiten zu klassifizieren, wie sie bei der Bildfolgenanalyse vorkommen können.

Dafür werden zunächst typische Tätigkeiten identifiziert und voneinander abgegrenzt. Für jede dieser Tätigkeiten wird dann eine Versuchsaufgabe gestaltet, die im Rahmen einer Datenerhebung von Versuchspersonen durchgeführt wird. Die dabei aufgezeichneten Blickbewegungen werden dann algorithmisch zu einer Menge Fixationen aggregiert, auf deren Basis wiederum Fixations- und Sakkadenparameter extrahiert werden. Diese dienen als Merkmale eines Merkmalsvektors, auf dessen Basis die Klassifikation der Tätigkeiten erfolgt.

1.3.4 Abgrenzung

Wie jede typische GUI weist auch die ABUL-GUI zahlreiche statische Bedienelemente sowie Dialoge mit Texteingabe auf. Die blickbasierte Interaktion für die Selektion statischer Objekte sowie für Texteingabe wird in der vorliegenden Arbeit nicht betrachtet, da hierzu Forschungsarbeiten in großer Menge vorliegen; sie sind in Abschnitt 2.4 als Grundlagen der vorliegenden Arbeit dokumentiert. Es ist erwiesen, dass blickbasierte Interaktion hier der pixelgenauen Mauseingabe unterlegen ist, da Objekte erst ab einer Größe von ca. 2 cm auf dem Monitor robust selektierbar sind. In der ABUL-GUI wären daher die Nebenvideoströme robust blickbasiert selektierbar [Hil13c], die kleineren Schaltflächen nicht.

1.4 Aufbau der Arbeit

Im Folgenden beschreibt Kapitel 2 die Grundlagen und verwandten Forschungsarbeiten zur vorliegenden Arbeit.

Kapitel 3 beschreibt die Konzepte, die für die drei Beiträge erarbeitet wurden.

Kapitel 4 beschreibt die für Kapitel 5 und Kapitel 6 realisierten Versuchssysteme mit den eingesetzten Eyetrackern und ihren Eigenschaften.

Kapitel 5 und Kapitel 6 beschreiben die Untersuchungen, die anhand von Nutzerstudien zu den drei Beiträgen durchgeführt wurden.

Kapitel 5 umfasst dabei Nutzerstudien, die im Rahmen von Beitrag 1 (Blickbasierte Interaktion für Selektionsoperationen an bewegten Objekten, Abschnitt 1.3.1) an abstrakten Versuchsaufgaben geeignete Blickinteraktionstechniken für die Bewegtobjektselektion erforschen. Kapitel 5 trägt daher den Titel „Identifikation geeigneter Blickinteraktionstechniken für die Bewegtobjektselektion“.

Kapitel 6 umfasst alle Untersuchungen, bei denen die Versuchsaufgaben anhand von Full Motion Video-Datenmaterial gestaltet wurden.

Abschnitt 6.1 umfasst dabei die Untersuchungen zur Bewegtobjektselektion, die den Beitrag 1 komplettieren (vgl. Abschnitt 1.3.1, S. 13 „Im dritten Schritt ...“).

Abschnitt 6.2 umfasst die Untersuchungen im Rahmen von Beitrag 2, der blickbasierten Interaktion bei automatischen Bildanalyseverfahren (vgl. Abschnitt 1.3.2).

Abschnitt 6.4 beschreibt die Untersuchung zu Beitrag 3, der blickbasierten Klassifikation der Benutzertätigkeit bei der Bildfolgenanalyse (vgl. Abschnitt 1.3.3).

Kapitel 7 bringt eine Zusammenfassung und einen Ausblick.

2 Stand von Forschung und Technik

Wie in Kapitel 1 ausgeführt, will die vorliegende Arbeit einen Betrag dazu leisten, Benutzungsschnittstellen für Desktop-Anwendungen mit dynamischen Szenen und bewegten Objekten, wie sie bei der Bildfolgenanalyse vorkommen, zu verbessern. Der vorgeschlagene Lösungsansatz ist, die Benutzungsschnittstelle auf Seiten der Informationseingabe um Eyetracking zu ergänzen.

Eyetracking wird dabei zum einen untersucht als Basis für blickbasierte Interaktion zur Selektion bewegter Objekte, zum anderen als Basis für die Tätigkeitsschätzung des Benutzers.

Die Aufgabe der Selektion bewegter Objekte, im Folgenden meist mit Bewegtobjektselektion bezeichnet, wird dabei im Vergleich mit dem Status Quo der Mauseingabe betrachtet.

Dieses Kapitel gibt einen Einblick in den Stand der Forschung und Technik. Abschnitt 2.1 beschreibt die Eigenschaften und Fähigkeiten des Menschen mit Schwerpunkt auf dem Auge und der visuellen Wahrnehmung. Abschnitt 2.2 beschreibt Eigenschaften und Fähigkeiten von Eyetracking und für die MMI geeignete Eyetracking-Geräte, die die technische Grundlage für blickbasierte Interaktion sind.

Die Abschnitte 2.3 und 2.4 bringen verwandte Arbeiten aus der MMI. Der Abschnitt 2.3 behandelt die manuelle Selektionsoperation mit separaten Abschnitten zu Eingabegeräten mit direkter (Abschnitt 2.3.1) bzw. indirekter Manipulation (Abschnitt 2.3.2) sowie zur Selektion statischer (Abschnitt 2.3.3) bzw. bewegter Objekte (Abschnitt 2.3.4).

Der Abschnitt 2.4 behandelt die blickbasierte Informationseingabe mit separaten Abschnitten zu Blickdatenfilterung (Abschnitt 2.4.1), blickbasierter Selektionsoperation (Abschnitt 2.4.2), blickbasierter Selektion bewegter Objekte (Abschnitt 2.4.3), Interaktion mit Blick und Brain-Computer-Interfaces (Abschnitt 2.4.4) sowie blickbasierter Tätigkeitsschätzung (Abschnitt 2.4.5).

Der Abschnitt 2.5 beschließt das Kapitel mit einer Beschreibung gängiger Normen, Metriken und Fragebögen der MMI, die in dieser Arbeit verwendet werden.

2.1 Der Mensch bei der Bildfolgenanalyse

Die Ausführungen in diesem Unterkapitel basieren vorwiegend auf

- Philipson [Phi97] *Manual of Photographic Interpretation*,
- Goldstein [Gol15] *Wahrnehmungspsychologie*,
- Lippert u. a. [Lip17] *Anatomie*,
- Brandes u. a. [Bra19] *Physiologie des Menschen*,
- Kaufmann u. a. [Kau20] *Strabismus*.

Wie jede Handlung des Menschen folgt auch eine Arbeitsaufgabe wie die Bildfolgenanalyse dem typischen Prozess aus Perzeption – Kognition – Aktion. Nach dem vereinfachten Modell von Goldstein [Gol15] kann man den Prozess in sieben Schritte unterteilen (Abb. 2.1). Die Schritte 1 bis 4 beschreiben die neuronalen Prozesse der visuellen Wahrnehmung und ihre grundlegenden physiologischen Prinzipien, die Schritte 5 bis 7 die Verhaltensreaktion des Benutzers.

Für die Aufgabe, in der der Benutzer ein bestimmtes Fahrzeug in der Bildfolge finden und markieren soll, sind die sieben Schritte wie folgt.

Am Beginn steht der Umgebungsreiz (1), im wissenschaftlichen Umfeld auch kurz mit *Reiz* oder *Stimulus* bezeichnet. Im Beispiel besteht der Reiz, der aktuell für den Benutzer von Interesse ist, aus dem Fahrzeug in der Bildmitte.

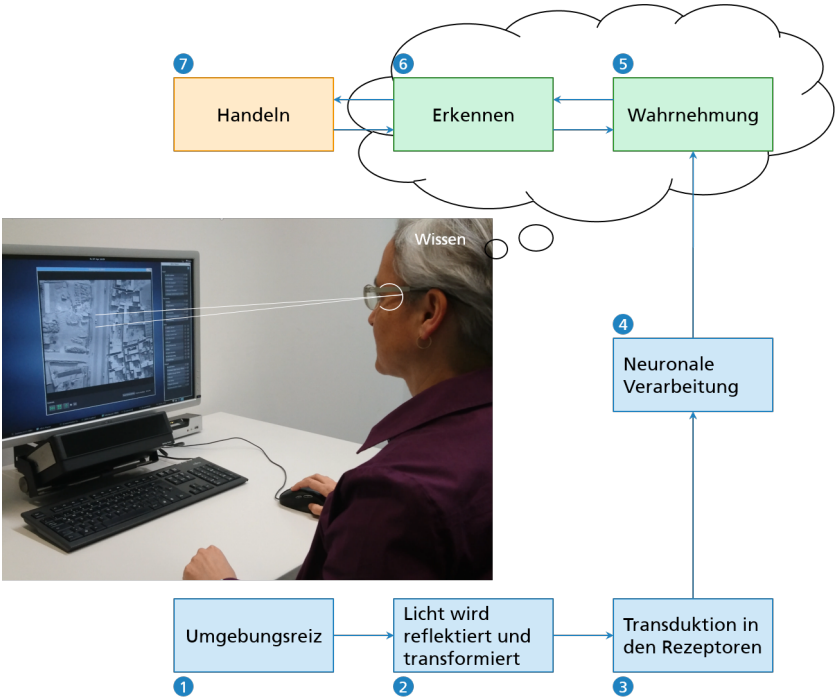


Abbildung 2.1: Wahrnehmungsprozess nach Goldstein [Gol15] am Beispiel der Aufgabe Bildfolgenanalyse.

Das Licht, das von diesem Fahrzeug ausgeht, wird ins Auge des Benutzers reflektiert. Augenhornhaut und Augenlinse transformieren das Licht und erzeugen ein invertiertes Abbild des Fahrzeuges auf der Netzhaut des Benutzers (2). Die Netzhaut enthält sensorische Fotorezeptoren, sodass der Reiz jetzt auf Rezeptorebene vorliegt.

Im nächsten Schritt (3) wandeln die Sehpigmente der Fotorezeptoren die eingefallene Lichtenergie in elektrische Energie um (Transduktion). Jetzt liegt das Abbild des Fahrzeugs als elektrische Repräsentation vor.

Danach werden diese elektrischen Signale neuronaler Verarbeitung unterzogen (4). Dabei leiten die Neuronen der Netzhaut ihre Signale über den Sehnerv

ins Gehirn, wo sie weiter verarbeitet werden und schließlich den primären visuellen Kortex (V1) erreichen. Während der Schritte 2 bis 4 werden die Signale wiederholt transformiert, repräsentieren jedoch stets das Fahrzeug.

In den folgenden Schritten wird aus den elektrischen Signalen die bewusste Wahrnehmung des Fahrzeugs (5), das Erkennen ordnet dem Fahrzeug eine Bedeutung zu (6). Die Pfeile in beide Richtungen zwischen Schritt 5 und 6 sollen verdeutlichen, dass Wahrnehmung und Erkennen auch gleichzeitig oder in umgekehrter Reihenfolge geschehen können.

Am Ende des Gesamtprozesses steht das Handeln (7). Wenn der Benutzer das Fahrzeug als das gesuchte erkannt hat, markiert er es. Bevor er markiert, wird der Benutzer in vielen Fällen das Fahrzeug erst genau ansehen müssen, um sicher entscheiden zu können, dass es das gesuchte Fahrzeug ist. Handeln besteht nach Goldstein also zum einen aus der motorischen Aktion des Markierens. Zum anderen gehören zum Handeln eine Anzahl Entscheidungen, Teile des Fahrzeugs zu betrachten, um zu erkennen, ob sie die gesuchte Beschreibung der Aufgabenstellung erfüllen. Daher enthält Abb. 2.1 auch zwischen Erkennen und Handeln Pfeile in beide Richtungen.

Neben diesen sieben Schritten beeinflusst das Wissen des Benutzers den Wahrnehmungsprozess. Goldstein beschreibt Wissen als „(…) jegliche Information, die der Wahrnehmende in eine Situation einbringt“. Wissen kann dabei unmittelbar zuvor oder Jahre zuvor erworben worden sein. Die eingebrachte Information wirkt dabei in unterschiedlicher Weise auf den Wahrnehmungsprozess, je nachdem, ob sie „Bottom-up“ oder „Top-down“ verarbeitet wird.

Angestoßen wird jeder Wahrnehmungsprozess durch Bottom-up-Verarbeitung (auch daten- oder reizgesteuerte Verarbeitung genannt). Im Beispiel bildet das Abbild des Fahrzeugs auf der Netzhaut die Grundlage dafür.

Basis für die Top-down-Verarbeitung (auch wissensbasierte Verarbeitung genannt) ist das Wissen des Benutzers, im Beispiel Wissen über Fahrzeuge allgemein und speziell über das gesuchte Fahrzeug. Laut Goldstein gilt „(…) je komplexer die Reize (…), desto mehr Einfluss gewinnt die Top-down-Verarbeitung“.

Abb. 2.1 zeigt die Situation, in der der Benutzer das Fahrzeug als Zielobjekt erkannt und die Entscheidung zu seiner Markierung getroffen hat. Wenn der Benutzer nach Stand der Technik die Markierung mit einem linken Mausklick auf das Fahrzeug markieren will, muss er als nächstes den Mauszeiger finden und auf das Fahrzeug platzieren. Dazu muss er seine visuelle Aufmerksamkeit zwischen Mauszeiger und Fahrzeug aufteilen und die sieben Schritte so lange für beide getrennt durchführen, bis sie sich nahe genug sind, sodass er sie gleichzeitig visuell erfassen kann. Hat er den Mauszeiger über dem Fahrzeug platziert, kann er den Mausklick zur Markierung durchführen.

Für den Fall, dass der Benutzer seine Aufgabe mit Unterstützung eines automatischen Verfahrens zur Bildfolgenanalyse durchführt, muss er auch dessen visualisierte Verfahrensergebnisse visuell wahrnehmen und ihnen Aufmerksamkeit widmen. Bevor er die automatischen Verfahren nutzen kann, muss er seine Aufmerksamkeit von der Bildfolge weg hin zu den entsprechenden Schaltflächen auf der GUI wenden, um die Verfahren durch Selektion der Schaltflächen zu aktivieren.

2.1.1 Das Auge

Die visuelle Wahrnehmung beginnt im Auge. Dabei beeinflusst das Auge die visuelle Wahrnehmung in dreierlei Weise:

- Durch den *optischen Apparat* auf der Vorderseite des Auges,
- durch die Sehpigmente in den *Fotorezeptoren* der Netzhaut
- sowie durch die *Verschaltung der Neuronen* der Netzhaut.

Abb. 2.2 verbildlicht dies; in schwarzer Schrift auf blauem Grund die Beschreibung der physikalischen Prozesse der Schritte (1) bis (4), in oranger Schrift darüber die zugehörige Wahrnehmungsleistung.

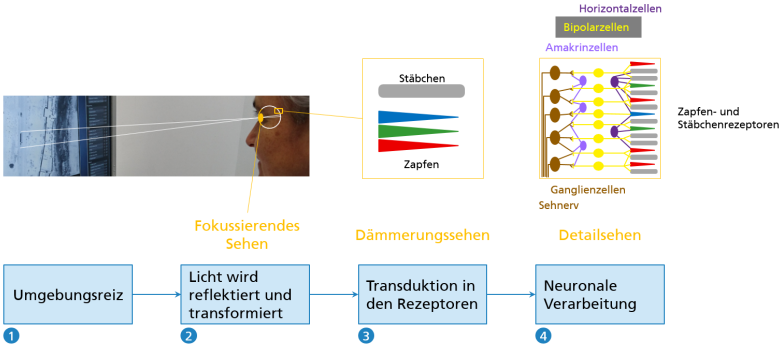


Abbildung 2.2: Schritte 1 bis 4 im Wahrnehmungsprozess nach Goldstein [Gol15] im Detail.

Der Augapfel selbst ist ein nahezu kugelförmiger Körper[1]. Die Wand des Augapfels besteht aus drei Hauptschichten:

- Die äußere, faserreiche Augenhaut besteht aus der weißen, undurchsichtigen Lederhaut (Sklera) und der glasklaren, durchsichtigen Hornhaut (Cornea).
- Die mittlere, gefäßreiche Augenhaut besteht aus Aderhaut (Chorioidea), Strahlenkörper und Regenbogenhaut (Iris).
- Die innere, nervenzellreiche Augenhaut ist die Netzhaut (Retina).

Den *optischen Apparat* bilden Hornhaut (Cornea), vordere Augenkammer, Regenbogenhaut (Iris), Linse und Glaskörper. Einfallendes Licht wird mehrmals gebrochen, zunächst an der Hornhaut (80% der Brechkraft). Das Licht tritt dann über die Öffnung der Iris – die Pupille (Augenapertur) – weiter ins Auge vor und wird durch die Linse erneut gebrochen (20% der Brechkraft).

Die Linse ist ein glasklares, linsenförmiges Gebilde im Innern des Auges (Durchmesser ca. 1 cm, größte Dicke 4 mm) [Lip17]. Ihre Aufgabe ist, das

[1] Vgl. z. B. Brandes u. a. [Bra19]: Abb. 56.1 Auge und optische Abbildung anhand eines Horizontalschnitts durch ein linkes Auge.

Licht so zu fokussieren, dass auf die Netzhaut ein scharfes Abbild des betrachteten Objektes projiziert wird. Dazu wirkt ein Ziliarmuskel auf die Linse, der ihre Brennweite je nach Entfernung des Reizes verändert.

Die Iris regelt den Lichteintritt zur Netzhaut. Änderungen der Pupillenweite erfolgen durch zwei Muskeln, die vom autonomen Nervensystem gesteuert werden. Die Pupillenweite ist variabel zwischen 1,5 mm und 8 mm und kann auf schnelle Veränderungen der Intensität des Umgebungslichts reagieren.

Erreicht das Licht die Netzhaut (Retina), wird es von den Sehpigmenten der *Fotorezeptoren* absorbiert, wodurch elektrische Signale ausgelöst werden[1]. Diese Signale werden über die Neuronen der Netzhaut weitergeleitet, dabei verarbeitet und über den Sehnerv an das Gehirn geliefert. Sie signalisieren dem Gehirn die Eigenschaften des betrachteten Objekts und sind deshalb entscheidend für die Wahrnehmung des Objekts.

Die Netzhaut kleidet die Innenseite der Augenwand bis vor zur Pupille aus. Der „sehende Teil der Netzhaut“ [Lip17] befindet sich an der hinteren Augenwand und enthält die Fotorezeptoren sowie weitere Nervenzellen[2].

Der Mensch besitzt zwei Arten von Fotorezeptoren. Die *Stäbchen* unterscheiden Hell und Dunkel („skotopisches Sehen“ oder „Nachtsehen“), die *Zapfen* Farben („photopisches Sehen“ oder „Tagsehen“).

Das Auge ist empfindlich für elektromagnetische Strahlung zwischen ca. 400 und 750 nm. Alle Sehpigmente besitzen eine etwa gleich große effektive spektrale Bandbreite, aber unterschiedliche spektrale Maxima. Stäbchen absorbieren maximal viel Licht bei etwa 496 nm (blau-grün). Bei den Zapfen muss man unterscheiden zwischen Zapfenpigmenten für kurzwelliges, mittelwelliges und langwelliges Licht mit Absorptionsmaxima bei etwa 419 nm (purpurblau), 535 nm (grün) bzw. 558 nm (grün-gelb).

[1] Fällt Licht auf einen Fotorezeptor, verändert sich die chemische Form des Sehpigments, was eine chemische Kettenreaktion auslöst, die letztlich den Rezeptor aktiviert.

[2] Abb. 2.2 mit vereinfachter Darstellung, für eine detaillierte Abbildung siehe z. B. Lippert u. a. [Lip17, S. 323].

Die menschliche Netzhaut enthält etwa 120 Mio. Stäbchen und 6 Mio. Zapfen. Da sie unterschiedliche Funktionen haben, sind sie unterschiedlich dicht auf der Netzhaut verteilt. Im Bereich der Fovea (auch Sehgrube oder Gelber Fleck) sind nur Zapfen vorhanden, etwa 50.000 auf 1,5 mm Durchmesser. Die restlichen Zapfen sind gleichmäßig in der peripheren Netzhaut verteilt.

Die Fovea ist der Teil der Netzhaut, mit dem das scharfe, detailreiche Sehen möglich ist. In Abb. 2.1 und Abb. 2.2 kann man sich als Ort der Fovea den Bereich an der Augenhinterwand vorstellen, den die beiden weiß eingezeichneten Linien abgrenzen. Denn um das Fahrzeug detailreich sehen zu können, richtet der Benutzer seine Augen so aus, dass das Abbild des Fahrzeugs auf den Bereich der Fovea abgebildet wird.

Der Bereich, den die Fovea umfasst, beträgt ca. 1,5-2° Sehwinkel [Hol11]. Bei einem für Desktopsysteme typischen Augenabstand vom Bildschirm von 65 cm erfasst die Fovea einen Bereich von ca. 17,0 bis 22,7 mm auf dem Bildschirm. Der geometrische Zusammenhang[1] zwischen Reizgröße x (in mm) und Sehwinkel θ (in °) berechnet sich bei gegebenem Abstand d (in mm) der Augen zum Bildschirm als

$$tan(\theta/2) = x/d \tag{2.1}$$

Wie die Wahrnehmung am Ende aussieht, wird bereits in der Netzhaut durch die Verbindung der Zapfen und Stäbchen bestimmt. Ihre Verschaltung erfolgt mithilfe von vier weiteren Neuronentypen (vgl. Abb. 2.2, ganz rechts) und ist sehr komplex[2].

Von entscheidender Bedeutung für die Wahrnehmung ist das Organisationsprinzip der Konvergenz. Denn den 126 Mio. Fotorezeptoren in der Netzhaut

[1] Vgl. [Hol11, S. 23].

[2] Die Fotorezeptoren (in der Abbildung grau die Stäbchen, blau, grün und rot die drei Zapfentypen) senden ihre Signale an die Bipolarzellen (gelb). Die Bipolarzellen senden weiter and die Ganglienzellen (braun), deren Axone den Sehnerv bilden, über den die Signale an das Gehirn weitergeleitet werden. Horizontalzellen (dunkelviolett) erlauben, dass sich Signale der Fotorezeptoren horizontal durch die Netzhaut verbreiten. Amakrinzellen beeinflussen, wie die Signale von den Bipolarzellen zu den Ganglienzellen verlaufen.

stehen nur 1 Mio. Ganglienzellen gegenüber. Die Konvergenz beträgt für die Stäbchen durchschnittlich 126 pro Ganglienzelle, für die Zapfen 6 pro Ganglienzelle.

In der Fovea sind viele der Zapfen jeweils alleine mit genau einer Ganglienzelle verbunden. Dies begründet die hohe Detailwahrnehmung der Fovea: Ohne Konvergenz wird ein Hinweis darauf geliefert, wie weit zwei Lichtreize voneinander entfernt sind. Die Auflösungssehschärfe (minimum separabile), d.h. der Abstand, unter dem zwei Punkte gerade noch getrennt erscheinen, beträgt 30 Winkelsekunden[1] (= 1/120 Grad).

Einfluss auf die Sehschärfe haben Helligkeit und Kontrast des visuellen Reizes sowie die Darbietungszeit (bei mehr als 500 ms wird die Sehschärfe für viele Personen besser). Außerdem spielt die Pupillenweite eine Rolle[2].

Die Sehschärfe ist abhängig vom Netzhautort. Ausgehend von der Fovea nimmt die Sehschärfe mit der Exzentrizität im Gesichtsfeld zunächst steil, dann reizabhängig weniger steil ab. Summationseffekte des Binokularsehens verbessern die Sehschärfe im Vergleich zum Monokularsehen (Faktor 1,1). Sehen ist nicht möglich an der Stelle, an der die Sehnervenfasern gebündelt als Sehnerv (Durchmesser 4 mm) die Netzhaut verlassen (sog. „Blinder Fleck“). Dies fällt jedoch nicht auf, weil die blinden Flecke der beiden Augen nicht zusammenfallen bzw. beim einäugigen Sehen das Großhirn den Ausfall ergänzt.

Der Augapfel liegt in der *Augenhöhle* (Volumen ca. 27 cm³ (Tiefe ca. 42 mm, Breite ca. 40 mm, Höhe ca. 35 mm). Zwischen Augapfel und Augenhöhle liegt ein Fettkörper, der die Bindegewebe, Augenmuskeln, Nerven und Gefäße einbettet und bei Augenbewegungen die Verlagerung der Augen ermöglicht.

[1] Dies entspricht genau dem Durchmesser eines Zapfens, was einleuchtet, da mindestens ein nicht stimulierter Zapfen zwischen den beiden durch die Punkte stimulierten Zapfen liegen muss.

[2] Sehschärfen-Bestwert bei 3 mm Durchmesser.

Die *Augenlider* bilden die vordere Begrenzung der Augenhöhle (Breite 25 mm, Dicke 1 mm, Höhe Oberlid 10 mm, Höhe Unterlid 5 mm). Sie dienen als Lichtschutz, d.h. sie ermöglichen, das Sehorgan „abzuschalten", als Blendschutz (Zukneifen der Lider wirkt wie Abblenden) sowie als mechanischer Schutz, wenn bei Annäherung einer Gefahr die Lider reflektorisch geschlossen werden. Außerdem verteilt der Lidschlag die Tränenflüssigkeit über die Augenvorderfläche und schützt so die Hornhaut vor Austrocknung.

Lidschluss tritt also einerseits reflektorisch auf, andererseits willentlich. Reflektorischer Lidschluss tritt häufig im Zusammenhang mit Sakkaden auf und erfolgt ca. 20-mal pro Minute [Lei15] mit einer Dauer von ca. 300 ms. In diesem Fall nimmt der Mensch die Dunkelphase nicht bewusst wahr, da das Gehirn die visuelle Wahrnehmung kurz vor dem Lidschlag unterdrückt.

2.1.2 Binokularsehen

Visuelle Wahrnehmung erfordert, dass die Lichtreize nach der Verarbeitung im Auge von höheren Schichten des Gehirns weiterverarbeitet werden. Nach den Schichten der Netzhaut leitet der Sehnerv die Nervenerregung vom Augapfel weiter ins Gehirn[1].

An der Sehnervenkreuzung (Chiasma) wechseln die Sehnervenfasern aus den nasalen Netzhauthälften jeweils zur Gegenseite und verlaufen danach zusammen mit den nicht kreuzenden Sehnervenfasern in verschiedene Hirnareale. Dadurch wird die linke Hälfte des Gesichtsfeldes in der rechten Gehirnhälfte verarbeitet und die rechte in der linken Gehirnhälfte.

90% der Sehnervenfasern enden im CGL (Corpus Geniculatum Laterale, im Zwischenhirn), der zentralen Schaltstelle zur Vermittlung der sensorischen Information zur visuellen Wahrnehmung in der Sehrinde, die wiederum am Hinterhauptlappen liegt.

[1] Vgl. z. B. Brandes u. a. [Bra19], Abb. 58.1 Schema der Sehbahn im Gehirn des Menschen.

Im CGL werden zum einen die Informationen aus den Sehnerven weiter analysiert, zum anderen interagieren die Sehnerven mit rückprojizierenden Neuronen aus der Sehrinde und nicht-visuellen Neuronen aus dem Hirnstamm (Steuerung überlebenswichtiger Funktionen wie Reflexe und Atmung) und dem Zwischenhirn. Dies bewirkt, dass z. B. der Wachheitsgrad (Hypothalamus) und die räumlich gerichtete Aufmerksamkeit die visuelle Signalverarbeitung beeinflussen. Signale ins Mittelhirn dienen der Steuerung der Augenbewegungen und des Pupillenlichtreflexes.

Grundsätzlich ist die neuronale Abbildung der Netzhaut in den Seharealen (CGL, Sehrinde) retinotop organisiert[1]. Informationen aus zwei benachbarten Netzhautregionen bleiben während der visuellen Analyse also benachbart. Die Abbildung ist jedoch nicht maßstabsgetreu, sondern verzerrt zugunsten der Netzhautmitte inklusive der Fovea, die durch eine relativ große Fläche repräsentiert wird. Die Fovea umfasst nur 0,01% der Netzhaut, belegt aber 8-10% der retinotopen Karte. Dadurch steht mehr Verarbeitungskapazität für Aufgaben zur Verfügung, die große Sehschärfe verlangen.

Die Sehrinde hat mehrere Teile mit jeweils unterschiedlichen Funktionen. Die primäre Sehrinde V1 prüft die Signale auf Linien, Bewegung und Farben und leitet sie zur weiteren Analyse in spezialisierte Subsysteme weiter[2]. Die Sinneseindrücke der beiden Gesichtshälften werden dann im Corpus callosum miteinander verbunden.

2.1.3 Augenbewegungen

Um Handlungen auf Basis visueller Wahrnehmung zu ermöglichen, muss ein kontinuierlicher Seheindruck gewährleistet werden. Dazu sind Augenbewegungen erforderlich, da detailreiche Seheindrücke nur auf der Fovea möglich sind. Augenbewegungen geschehen mithilfe von Augenmuskeln.

[1] „Die Abbildung der Netzhaut wird als räumliches Erregungsmuster auf die zentralen Sehareale wie auf eine feine Landkarte projiziert.“ [Bra19, S. 745]

[2] V2: Gestalt- und Konturerkennung, V3: Analyse bewegter Formen, V4: Analyse von Farbe, V5: Analyse von Bewegung.

Die drei *inneren Augenmuskeln* (in der Regenbogenhaut/Strahlenkörper) regeln Pupillenweite und Entfernungseinstellung der Linse. Diese beiden Vorgänge werden von autonomen Nerven gesteuert: Der Sympathikus erweitert die Pupille, der Parasympathikus schließt die Pupille und steuert den Linsenmuskel.

Im Ruhezustand fokussiert das Auge Objekte im Abstand von 80 cm bis unendlich. Maximale Akkommodation liefert den Nahpunkt (altersabhängig ca. 10 cm mit 20 Jahren, ca. 100 cm mit 60 Jahren). Der Akkommodationsprozess beinhaltet einen Regelkreis, der kontinuierlich versucht, die Qualität des Netzhautbildes zu optimieren.

Die sechs *äußeren Augenmuskeln* ermöglichen dem Menschen eine willentliche Augenbewegung, denn sie sind von sogenannten Willkürnerven innerviert. Bei fast allen Augenbewegungen sind Kontraktionen mehrerer Augenmuskeln beteiligt.

Bewegt sich das Auge nicht, so besteht ein Gleichgewicht zwischen allen am Auge angreifenden Drehmomenten. Als Primärstellung des Auges bezeichnet man den bei gerader Kopf- und Körperhaltung geradeaus gerichteten Blick. Ein Auge kann aus der Primärstellung in horizontaler Richtung (nach links und rechts) sowie in vertikaler Richtung nach unten um 9-10 mm bewegt werden, in vertikaler Richtung nach oben 5-7 mm (vgl. Abb. 1.30 in [Kau20]).

Als Gebrauchsblickfeld im täglichen Leben sind Blickwendungen nach links und rechts um 20°, Blickhebungen um 10° und Blicksenkungen um 30° typisch. Jede Blickbewegung wird von einer Kopfbewegung begleitet. Denn Kopfbewegungen setzen nicht erst mit Erreichen der Grenzen des Blickfeldes ein, sondern sind auch bei kleinen Augenbewegungen nachweisbar.

Das Blickfeld, „(...) der Bereich der visuellen Umwelt, der bei unbewegtem Kopf, aber frei umherblickenden Augen wahrgenommen werden kann“

[Bra19, S. 745] beträgt für jedes Auge knapp 100° in horizontaler Richtung und etwas über 100° in vertikaler Richtung[1].

Beim Binokularsehen klassifiziert man die Augenbewegungen in Versionen und Vergenzen[2]. Versionen werden wie folgt unterschieden:

- Rasche Versionen
  - Sakkaden (auch Augenrucke): Blickzielbewegungen, willkürliche Blickbewegungen, spontane Blickbewegungen.
  - Rasche Phasen des optokinetischen Nystagmus[3]
  - Rasche Phasen des vestibulären Nystagmus
- Langsame Versionen
  - Langsame Phasen des optokinetischen Nystagmus
  - (Gleitende) Folgebewegungen (*engl. smooth pursuits*)
  - Vestibuläre Kompensationsbewegung (auch: vestibulärer Reflex; Entsprechung: langsame Phasen des vestibulären Nystagmus)

Stellt man sich die Okulomotorik als ein System mit mehreren Regelkreisen vor, so hat die Netzhaut die Rolle des Fühlers, das Gehirn umfasst verschiedene Regler und die Augenmuskeln dienen als Stellglied. Ändert das Auge seine Stellung, wirkt dies auf die Netzhaut zurück, sodass sich der Informationsfluss zum Kreis schließt.

Je nach Augenbewegung fühlt die Netzhaut eine andere Regelgröße. Bei Sakkaden ist dies die Position des Netzhautbildes, bei Folgebewegungen zusätzlich die Geschwindigkeit der Bildverschiebung. Die Netzhaut übermittelt die jeweilige(n) Regelgröße(n) an das Gehirn, wo die gelieferten IST-Werte mit

[1] Da die beiden monokularen Blickfelder sich nicht komplett überlappen, ist das binokulare Blickfeld (der Blickbereich, in dem beide Augen gemeinsam mit den beiden Foveas fixieren können) horizontal etwas kleiner als die Summe der beiden monokularen Blickfelder.

[2] Vergenzen sind weiter unterscheidbar in akkommodative und fusionale [Kau20], beide sind für die vorliegende Arbeit nicht von Interesse.

[3] Nystagmus: natürlicher Bewegungsreflex der Augen.

den von der Aufmerksamkeitszuwendung abhängigen SOLL-Werten verglichen und darauf basierend neue Stellbefehle für die Augen errechnet werden. Die Stellbefehle werden über Hirnnerven an die Augenmuskeln übermittelt und aus verschiedenen Hirnarealen gesteuert [Kau20].

Die *raschen Augenbewegungen* (Sakkaden und rasche Nystagmusphasen) dienen dazu, die Fovea möglichst schnell auf Objekte besonderen Interesses zu richten. Während Sakkaden selbst ist keine visuelle Wahrnehmung möglich (sakkadische Suppression). Ein Blickpositionswechsel erfolgt etwa 5-mal pro Sekunde. Dabei sind einzelne Blickbewegungen willkürlich steuerbar.

Oft beeinflusst ausschließlich die allgemeine Aufmerksamkeitszuwendung eine ganze Folge von Sakkaden. Die Entscheidung, ob ein Ziel von Bedeutung ist und eine Sakkade dorthin erfolgen soll, wird von vier Hirnarealen gesteuert, Hirnstamm, Großhirn, Mittelhirn und Kleinhirn. Sakkaden können willkürlich sein, aber auch durch einen Reflex bedingt.

Die Amplitude von Sakkaden liegt zwischen wenigen Winkelminuten bis über 90° [Bra19]. Die Winkelgeschwindigkeit von Sakkaden ist abhängig von ihrer Amplitude und beträgt zwischen 30 und 500°/s [Hol11]; Brandes u. a. [Bra19] nennen für große Sakkaden (> 57°) Winkelgeschwindigkeiten über 500°/s.

Die Sakkadendauer ist ebenfalls abhängig von der Sakkadenamplitude und wird mit 10-80 ms [Bra19] bzw. 30-80 ms [Hol11] angegeben. Carpenter [Car88] modellierte den Zusammenhang zwischen Sakkadendauer (in ms) und Sakkadenamplitude (in °) mithilfe der Formel

$$Sakkadendauer = 2{,}2 * Sakkadenamplitude + 21 \tag{2.2}$$

Gleitende Folgebewegungen dienen dazu, kleine bewegte Sehobjekte kontinuierlich auf der Fovea abzubilden. Sie sind willkürlich und konjugiert. Die Reizgeschwindigkeiten, bei denen sie auftreten, werden von Brandes u. a. [Bra19] mit bis zu 100°/s angegeben, Holmqvist u. a. [Hol11] nennen einen Wert von bis zu 30°/s.

Bei höheren Geschwindigkeiten ist das Folgesystem überfordert und es werden Sakkaden eingestreut, damit das Abbild des Fixierobjekts trotzdem nahe

der Fovea gehalten werden kann. Es resultiert dann eine sogenannte „sakkadierte Folgebewegung“, bei der die horizontalen Folgebewegungen zwischen den Sakkaden bis zu 150°/s erreichen können. Folgebewegungen sind in erheblichem Maße abhängig von Aufmerksamkeitszuwendung.

Ausgelöst werden Folgebewegungen, sobald ein Objekt sich zu bewegen beginnt und sich sein Bild über die Netzhaut verschiebt. Je mehr sich die Winkelgeschwindigkeit der Augen der des Objekts anpasst, umso weniger verschiebt sich das Bild über die Netzhaut. Trotzdem folgen die Augen weiter dem Objekt, denn das Gehirn ist über die aktuelle Augenbewegung informiert (es gibt selbst den Befehl dazu). Fehlt die retinale Bildverschiebung, interpretiert das Gehirn dies als Bestätigung, die Bewegung fortzusetzen. Folgen die Augen zu langsam oder zu schnell, verschiebt sich das Netzhautbild erneut und veranlasst das Gehirn, die Geschwindigkeit der Augenbewegung zu korrigieren [Kau20].

Um ein Objekt erkennen und identifizieren zu können, muss es eine Weile lang auf der Fovea abgebildet werden. Man bezeichnet diesen Vorgang auch als *Fixation*. Brandes u. a. [Bra19] geben als typische Fixationsdauer 200-600 ms an; pro Stunde treten zwischen den Sakkaden etwa 10.000 Fixationsperioden auf.

Auch wenn man den Eindruck hat, einen Punkt völlig stabil und ruhig zu fixieren, führen die Augen dabei unablässig kleine Bewegungen aus:

- Langsame Mikrobewegungen (Drift): Amplitude 2,5 Winkelminuten, Geschwindigkeit 2-8 Winkelminuten/s. Die Verschiebungen durch Drift verhindern Lokaladaption[1].
- Mikrosakkaden: Amplitude 3-50 Winkelminuten, linear abhängige Maximalgeschwindigkeit von 8°/s bei Amplitude von 50 Winkelminuten. Drift und Mikrosakkaden dienen der Fixationsregelung und führen die beiden Augen immer wieder auf das geforderte Fixierobjekt zurück.

[1] Vgl. Troxler-Effekt: Netzhautareale passen sich anhaltenden oder wiederkehrenden Reizen ähnlicher Intensität so an, dass ihre Empfindlichkeit abnimmt.

- Mikrotremor: Amplitude < 1 Winkelminute, Geschwindigkeit bis ca. 10 Winkelminuten/s. Die Frequenz beträgt etwa 30 Hz, sodass Mikrotremor auch als Rauschen interpretiert wird.

2.1.4 Wahrnehmung und Erkennen

Die einzelnen Untersysteme, die im Zusammenspiel die visuelle *Wahrnehmung* ermöglichen, sind heute recht gut verstanden. Wie allerdings letztendlich die Wahrnehmung des Bildes unserer Umgebung zustandekommen, ist bislang unklar [Kau20].

Es existieren zwei Erklärungsansätze dafür, wie aus den komplexen physiologischen Vorgängen beim Sehprozess die Wahrnehmung einer Szene mit Bedeutung entsteht. Der *Strukturalismus* geht davon aus, dass die Wahrnehmung aufgeteilt wird in die Empfindungen der einzelnen Komponenten (Farbe, Helligkeit, Textur, Bewegung etc.), die jeweils mithilfe erlernter Regeln interpretiert werden. Jede komplexe Wahrnehmung wird dann aus den einfachen Empfindungen zusammengebaut.

Die *Gestalttheorie* geht davon aus, dass die Wahrnehmung auf einer Reihe von Prinzipien basiert, die manche Wahrnehmungen bzw. Gruppierungen wahrscheinlicher machen als andere. Bei Forschungen wurden zahlreiche Gestaltprinzipien gefunden, etwa das Prinzip der Prägnanz, das besagt, dass ein Reizmuster so gesehen wird, „…dass die resultierende Struktur so einfach wie möglich ist“ [Gol15].

Neben der Gruppierung von Elementen spielt bei der visuellen Wahrnehmung einer Szene auch die Trennung von Objekten vom Hintergrund eine Rolle (auch: perzeptuelle Segmentierung oder Figur-Grund-Unterscheidung).

Verschiedene Faktoren beeinflussen, ob eine Figur wahrgenommen wird oder nicht. Beispielsweise werden konvexe Seiten von Konturen eher als Figur wahrgenommen oder gewohnte physikalische Regelmäßigkeiten wie die „Licht-von-oben-Heuristik“ lässt Kreisflächen mit aufsteigenden Grauwerten als Erhebungen erscheinen, solche mit absteigenden als Vertiefungen. Auch

subjektive Faktoren bestimmen, was Figur ist und was Grund: Ein Objekt, das Bedeutung enthält, wird eher als Figur erkannt.

Mit dem Strukturalismus lässt sich erklären, wie Neuronen Bilder analysieren, mit der Gestalttheorie visuelles Erleben und visuelle Mehrdeutigkeit. Eine Verbindung beider Schulen ist noch nicht gelungen.

Betrachtet man das *Erkennen*, so kann man sich vorstellen, dass die visuelle Szene mit ihren unbekannten Komponenten bezüglich bekannter Merkmale interpretiert wird [Phi97]. Sieht der menschliche Beobachter das Fahrzeug in der Bildfolge aus Abb. 2.1, so hat er dafür (teilweise unbewusst)

- die visuelle Szene interpretiert,
- ihren Typ festgestellt,
- ein Objekt in der Szene lokalisiert,
- dieses Objekt mit allen Konzepten aller ihm bekannten Objekte verglichen,
- festgestellt, dass es am besten das Konzept „Fahrzeug“ erfüllt,
- und das Objekt unverzüglich als „Fahrzeug“ erkannt.

Der Mensch denkt also in Suchbildern, die aus Konzepten gebildet werden, die mit allen erkennbaren Formen, Oberflächen und Objekten verknüpft sind. Visualisiert man mental ein Fahrzeug, so konstruiert man dessen Suchbild.

In Anlehnung an Philipson [Phi97] kann man Bildfolgenanalyse als eine visuelle Such-Aktivität ansehen: Die visuellen Inhalte der Bildfolge werden analysiert und mit einer Vielzahl an Suchbildern zur Deckung gebracht, sodass die Szene aussagekräftig interpretiert werden kann.

2.1.5 Aufmerksamkeit

Wahrnehmung geschieht nicht nur, weil Stimuli die Netzhaut erreichen und verarbeitet werden. Ein wesentlicher Einflussfaktor ist die Aufmerksamkeit

des Betrachters. Auch bei der Bildfolgenanalyse mit ihrer visuellen Suchaktivität werden manche Objekte der Szene besonders beachtet. Andere werden ignoriert, denn die Verarbeitungskapazität des visuellen Systems ist beschränkt.

Ein wichtiger Mechanismus zur Auswahl von Objekten, „(...) ist das visuelle Abtasten, mit dem wir verschiedene Stellen nacheinander ansehen“ [Gol15]. Die Augen werden dabei bewegt, da Detailsehen auf der Netzhaut nur an der Fovea möglich ist (s.o.). Der Blick ruht eine kurze Weile auf einer Stelle, um sie zu erfassen, und dann wird mithilfe einer Sakkade die Fovea auf die nächste Stelle positioniert. Der erste Schritt der Aufmerksamkeit ist also, dass die Augen auf ein Objekt fokussieren und dadurch seine Verarbeitung und infolge seine Wahrnehmung verbessert wird.

Aufmerksamkeit wird durch verschiedene Aspekte gelenkt. Einer ist die sogenannte *Stimulussalienz*, die dadurch entsteht, dass Merkmale wie Kontrast, Farbe oder Bewegung in einer Szene hervorstechen. Analysen von Bildbetrachtungen ergaben, dass in Bildern zuerst saliente Bereiche fixiert werden. *Top-down-Prozesse* in Form von Zielen oder Erwartungen des Beobachters übernehmen jedoch bereits nach den ersten Fixationen.

Was in welcher Reihenfolge in statischen Szenen visuell abgetastet wird, ist individuell verschieden und wird von Interessen und Vorwissen des Beobachters über das Szenenschema beeinflusst. Ein Objekt wird umso länger fixiert, je ungewöhnlicher es in einer Szene ist. Zudem bestimmt Wissen über die Szenenstatistik – wie wahrscheinlich treten Objekte oder Ereignisse in der Szene wo auf? –, wohin die Aufmerksamkeit gerichtet wird.

Aufmerksamkeit verändert mehrere Aspekte bei der Wahrnehmung. Sie beschleunigt die Reaktionszeit (Reizantwort), denn räumliche Aufmerksamkeit verbessert die Fähigkeit, auf Reize an diesem Ort zu reagieren (245 ms versus 305 ms, [Pos78]). Als Metapher wird gerne der Scheinwerfer verwendet, der eine bestimmte Stelle der Szene exklusiv hell erleuchtet. Aufmerksamkeit beschleunigt auch die Reaktion auf Objekte („Objektidentitätsvorteil“), was mutmaßlich auch für die Wahrnehmung teilverdeckter Objekte hilft.

Aufmerksamkeit beeinflusst zudem die physiologische Antwort in Form erhöhter Gehirnaktivität bestimmter Areale, wenn die Aufmerksamkeit auf ein bestimmtes Objekt oder auf einen bestimmten Ort gerichtet ist. Ist ein Objekt zudem bewegt, so erhöht sich zusätzlich die Aktivität im Areal V5, wo die Objektbewegungsverarbeitung stattfindet.

Man unterscheidet offene Aufmerksamkeit und verdeckte Aufmerksamkeit. Bei ersterer liegt der Blick auf dem beachteten Objekt, bei letzterer liegt er abseits davon.

Fehlende Aufmerksamkeit führt dazu, dass Dinge nicht wahrgenommen werden. Dies kann zum einen daran liegen, dass ein Objekt nicht im Blickfeld liegt. Zum anderen existiert das Phänomen der Unaufmerksamkeitsblindheit, die Situation, dass deutlich sichtbar vor Augen liegende Objekte nicht wahrgenommen werden (vgl. z. B. [Sim99, Ren97, Ren02]).

Man geht heute davon aus, dass eine wichtige Funktion von Aufmerksamkeit ist, Objektmerkmale wie Farbe, Form, räumliche Tiefe, Bewegung und Position, die in unterschiedlichen Gehirnarealen wahrgenommen werden, zu einer integrierten Wahrnehmung zu verbinden (vgl. oben Strukturalismus). Treisman beschriebt diese Bindung mithilfe der Merkmalsintegrationstheorie [Tre86, Tre88, Tre99].

Demnach wird das Abbild eines Objektes zunächst für die Merkmale separat verarbeitet (präattentive Phase, Merkmalsanalyse). Dann werden die Merkmale mithilfe von Aufmerksamkeitsfokussierung kombiniert, sodass ein Objekt als Ganzes wahrgenommen wird (attentive Phase). Beim Ansatz der Merkmalsintegrationstheorie umfasst die Phase der Merkmalsanalyse vorwiegend Bottom-up-Verarbeitung, „weil Wissen in der Regel nicht einbezogen wird. (…) Im Alltag wirken Top-down-Prozesse und Merkmalsanalyse bei der genauen Wahrnehmung vertrauter Objekte zusammen“ [Gol15].

Bei der Aufgabe der Visuellen Suche zeigt sich, dass die Ausprägung der Merkmalssuche nach einem einzelnen variierenden Merkmal (z. B. Farbe) ohne Aufmerksamkeitsfokussierung gelingt. Die Ausprägung der Konjunktionssuche mit mindestens zwei variierenden Merkmalen (z. B. Farbe und Form) erfordert jedoch zum Auffinden des Zielreizes das visuelle Abtasten der Szene

und eine damit verbundene Verlagerung und Fokussierung der Aufmerksamkeit.

Besonders stark wird die Aufmerksamkeit durch Bewegung gebunden. Dies passt zur Gestalttheorie: Ein Objekt wird durch Bewegung sichtbar, weil sich seine Bestandteile gleich bewegen und dadurch zu einer Einheit werden.

2.1.6 Handeln

Wahrnehmung und Handeln greifen ineinander und beeinflussen sich gegenseitig. Im Falle der Aufgabe der Bildfolgenanalyse besteht das Handeln aus mehreren Entscheidungen und Aktionen, an deren Ende die Markierung des Zielobjekts steht (s.o. Abschnitt 2.1).

Man geht heute davon aus, dass die Verarbeitungsprozesse für Wahrnehmung und Handeln in unterschiedlichen Hirnarealen ablaufen, die unterschiedliche Funktionen erfüllen. Die Wahrnehmung eines Objekts und seine Identifizierung übernimmt der sogenannte ventrale Strom (auch „Was?-Strom“), der von der Sehrinde in den Temporallappen fließt. Die Handlung mit dem Objekt – im Beispiel die motorische Aktion zur Zielobjektmarkierung in der Bildfolge – steuert der dorsale Strom (auch „Wo?-Strom“), der von der Sehrinde in den Parietallappen fließt. Dabei liefert der dorsale Strom neben Information über den Ort des Objekts auch Information über das Ausführen der Handlung selbst (das „Wie?“).

Die beiden Ströme sind jedoch nicht vollständig getrennt, sondern tauschen Informationen aus. Außerdem ist der Informationsfluss nicht nur vom visuellen System in Richtung des Temporal- bzw. Parietallappens gerichtet. Vielmehr geben beide auch Feedback-Signale zurück an das visuelle System und sind damit einer der Mechanismen hinter der oben beschriebenen Top-down-Verarbeitung visueller Information.

2.2 Eyetracking und Eyetracker für die MMI

Der DUDEN definiert *Eyetracking* (engl. *eye tracking* von eye = Auge und tracking = Verfolgung) als „(...) die elektronische Messung und Aufzeichnung von Augenbewegungen“[1]. Im Deutschen werden für Eyetracking auch die Begriffe Blickerfassung, Blickregistrierung oder Blick(richtungs)messung verwendet. Das entsprechende Gerät, das in der Lage ist, kontinuierlich die Blickrichtung zu bestimmen, wird *Eyetracker* genannt.

Es existieren verschiedene Geräte und Methoden für Eyetracking. Sie unterscheiden sich bezüglich der Messapparatur, der räumlichen und zeitlichen Auflösung, mit der Blickdaten geliefert werden, im Ausmaß an (Un-)Aufdringlichkeit, dem der Benutzer während der Messung ausgeliefert ist, sowie im Preis.

Die Geräte der Wahl für die MMI sind *nicht-intrusive video-basierte Eyetracker*. Sie beobachten die Blickrichtung direkt über eine Kamera. Die Blickrichtung wird auf Basis der Kamerabilder der Augen mithilfe von zwei Augenmerkmalen bestimmt, der Pupillenmitte und der Cornea-Reflexion.

Die Cornea-Reflexion wird durch Beleuchtung der Augen mit nahem Infrarot erzeugt und erscheint im Kamerabild als weißer Punkt. Aus Pupillenmitte und Cornea-Reflexion lässt sich der Blickrichtungsvektor berechnen. Denn da das Auge (fast) kugelförmig ist und sich um sein Zentrum bewegt, bewirken Augenbewegungen zwar eine Veränderung der Position der Pupillenmitte, die Position der Cornea-Reflexion aber verbleibt praktisch immer am selben Ort.

Video-basierte Eyetracker gibt es in zwei Varianten, statische Tisch-basierte und Kopf-getragene. Kopf-getragene Eyetracker haben den Nachteil, dass das permanente Tragen der Messapparatur auf dem Kopf den Benutzer belastet.

Statische Tisch-basierte Eyetracker lassen sich unterscheiden in turmartige Systeme (engl. *tower-mounted eye tracker*) und Systeme, die berührungslos aus der Distanz messen (engl. *remote eye tracker*).

[1] https://www.duden.de/rechtschreibung/Eyetracking

Turmartige Systeme sind nahe beim Benutzer platziert und schränken die Kopfbewegungsfreiheit über eine Stütze von Kinn und Stirn ein. Daher sind auch sie nicht geeignet für blickbasierte Interaktion. Denn auch diese Bewegungseinschränkung erzeugt unerwünschte Belastung beim Benutzer, da sie ihm unnatürliches Bewegungsverhalten auferlegt.

Geeignet sind hingegen statische video-basierte Eyetracker, die die Blickrichtung berührungslos aus der Distanz messen, im Folgenden kurz mit *Remote-Eyetracker* bezeichnet. Bei Remote-Eyetrackern sind Kamera und Infrarot-Beleuchtung direkt am oder nahe beim Bildschirm platziert.

Abb. 2.3 zeigt den im Rahmen dieser Arbeit überwiegend eingesetzten Tobii X60. Die Messapparatur aus Infrarot-Leuchten und Augenkamera ist so platziert, dass beide auf den Benutzer ausgerichtet sind. Details inklusive der technischen Spezifikationen sind in Abschnitt 4.2 beschrieben.

Remote-Eyetracker haben den Vorteil, dass sie einfach nutzbar sind und dass die Messung aus der Distanz den Benutzer eher vergessen lässt, dass überhaupt ein Eyetracker vorhanden ist. Remote-Eyetracker gibt es in Form des Beispiels aus Abb. 2.3, wo eine Messaparatur aus Kamera und Infrarot-Leuchten auf den Benutzer ausgerichtet sind, der sich an einem eng eingegrenzten Ort befindet, wie etwa der Benutzer am Desktoparbeitsplatz in der Abbildung.

Remote-Eyetracker gibt es jedoch auch als Mehrkamerasysteme, bei denen mehrere einzelne Messeinheiten aus Kamera und Infrarot-Leuchten so verbaut sind, dass sich der Benutzer viel bewegen und den Kopf drehen kann. Anwendungsbeispiele sind Fahrzeuge und Flugsimulatoren.

Auch Remote-Eyetracker wie der Tobii X60 erlauben eine gewisse *Kopfbewegungsfreiheit* innerhalb eines bestimmten räumlichen Volumens, der sogenannten *Head-Movement Box*. Nicht zuletzt aus diesem Grund wird Remote-Eyetrackern nachgesagt, dass sich die Benutzer mit ihnen natürlicher verhalten als bspw. bei turmartigen Systemen (s.o.). Dies bedeutet, dass mit Remote-Eyetrackern aufgezeichnete Blickdaten ökologisch valider wären als etwa die

von turmartigen Systemen. Systematisch untersucht wurde dies jedoch bislang nie [Hol11]. Laut Vertegaal [Ver08] gibt es Eyetracker mit Kopfbewegungsfreiheit seit etwa den Jahren 2004 (System der Firma LC) bzw. 2005 (Tobii 1750, vgl. a. Abschnitt 4.1).

Abbildung 2.3: Remote-Eyetracker Tobii X60, platziert unterhalb des Bildschirms.

In Abb. 2.3 ist zu sehen, dass Augenkamera, Beleuchtung und Stimulusebene (hier der Bildschirm, auf den geblickt wird) in einem festen Aufbau vorliegen. Zudem ist bekannt, wohin der Kopf zeigt. Diese Gegebenheiten ermöglichen, dass die Blickdaten als Koordinaten in einem festgelegten Koordinatensystem geliefert werden können.

Im Vergleich zu turmartigen Eyetrackern mit Kopfstabilisierung liefern Remote-Eyetracker die Blickdaten jedoch mit etwas geringerer Datenqualität bezüglich räumlicher *Genauigkeit* (engl. *accuracy*) und *Präzision* (engl. *precision*). Dies ist darauf zurückzuführen, dass das Modell, das der Blickschätzung zugrunde liegt, während Kopfbewegungen nicht perfekt passt. Manche Remote-Eyetracker haben daher ein Verfahren zum Kopftracking integriert, manche verlangen, dass dem Benutzer dabei eine Marke auf die Stirn geklebt wird[1]. Außerdem wird das Auge typischerweise mit geringerer Auflösung aufgezeichnet.

Die wichtigsten Kenngrößen von Eyetrackern sind die räumliche Genauigkeit, räumliche Präzision, Latenz, zeitliche Präzision und zeitliche Auflösung und werden im Folgenden beschrieben.

Die ***räumliche Genauigkeit*** (engl. ***accuracy***) ist definiert als die Differenz zwischen der wahren Blickposition und der aufgezeichneten Blickposition. Damit diese Differenz möglichst gering ist, muss der Algorithmus zur Blickschätzung die Blickrichtung auf Basis der Positionen der Pupillenmitte bzw. der Cornea-Reflexionen robust berechnen können. Hauptfehlerquellen bei der Berechnung sind das Rauschen der Pupillen- bzw. Cornea-Reflexions-Orte sowie die Nichtübereinstimmung der sphärischen Gestalt der Cornea-Reflexion im Vergleich zu ihrer wahren Gestalt [Hol11]. Gute räumliche Genauigkeit ist eine wichtige Eigenschaft für alle Anwendungen, bei denen relevant ist, wohin der Benutzer exakt blickt.

Theoretisch ist die räumliche Genauigkeit limitiert durch die Größe des Bereichs, den die Fovea umfasst (1,5-2°, s. a. S. 26) [Hol11]. Im Falle einer typischen Aufzeichnungsdistanz von 65-70 cm zwischen Benutzer und Bildschirm, entspricht dieser Bereich etwa einer Größe von 2 cm auf dem Bildschirm. Die Genauigkeit kann aber besser werden, da sie über die präzise Position des Auges während der Kalibrierung bestimmt ist. Falls die Augen während der Kalibrierung so ausgerichtet sind, dass für alle Kalibrierungspunkte gilt, dass sie auf dieselbe Position auf der Fovea abgebildet werden, verbessert

[1] z. B. beim EyeLink 1000 Plus, https://www.sr-research.com/eyelink-1000-plus/

sich die Genauigkeit auf einen Wert < 0,5°, was etwa 0,5 cm auf dem Monitor entspricht. 0,5° ist der Wert, den Eyetracker-Hersteller gewöhnlich angeben.

Dass der Benutzer vor der Nutzung des Eyetrackers einen Kalibriervorgang absolviert, ist erforderlich, weil Augen sich individuell unterscheiden. Beispielsweise variiert der Radius des Augapfels um bis zu 10% bei Erwachsenen und auch die Augapfelform ist unterschiedlich. Solche Unterschiede, aber auch das Tragen von Brillen oder Kontaktlinsen verändern die Geometrie bei der Berechnung der Blickrichtung und führen häufig zu schlechteren Kalibrierergebnissen [Hol11].

Die Kalibrierung erfolgt üblicherweise mithilfe von Kalibrierpunkten, typisch ist eine Anzahl von 5 oder 9. Sie werden auf dem Bildschirm so verteilt präsentiert, dass sie den Bereich, auf dem die Blickmessung stattfinden soll, umfassen; meistens wird auch der Bildschirmmittelpunkt mit einem Kalibrierpunkt belegt. Der Benutzer muss diese Kalibrierpunkte alle nacheinander fixieren.

Dabei zeichnet der Eyetracker für jeden Kalibrierpunkt einige Augenbilder auf, in denen Pupille und Cornea-Reflexionen die für den jeweiligen Kalibrierpunkt charakteristischen Positionen haben. Auf Basis dieser aufgezeichneten Augenbilder und den zugehörigen Kalibrierpunktpositionen kann die Blickaufzeichnungssoftware eine Funktion anpassen, die es erlaubt, auch für alle anderen Monitorpunkte aus gegebenen Positionen von Pupillenmitte und Cornea-Reflexion die Blickposition auf dem Monitor zu schätzen.

Die Genauigkeit während der Nutzung des Eyetrackers kann sich verschlechtern aufgrund individuellen Benutzerverhaltens, aber auch aufgrund von Veränderungen in der Umgebung.

Ein typisches Problem des Benutzerverhaltens ist, dass der Benutzer sich während der Kalibrierung anders verhalten hat als bei der Nutzung. Insbesondere ungeübte Eyetracker-Benutzer versuchen oft, sich bei der Kalibrierung besonders anzustrengen: Sie setzen sich besonders aufrecht hin, halten den Kopf starr nach vorn gerichtet und öffnen die Augen unnatürlich weit. Sobald die Kalibrierung erfolgt ist, entspannen sie sich wieder. Dadurch verändern sich die Position von Kopf und Augen etwas und die Augen werden wieder

nur natürlich weit geöffnet. Infolge kann sich die Genauigkeit verschlechtern [Hol11].

Mehr Tränenflüssigkeit kann dazu führen, dass sich weitere Reflexionen im Augenbild ergeben, die das Blickberechnungsverfahren fälschlicherweise als Cornea-Reflexionen interpretiert, sodass die Berechnung auf Basis einer falschen Position fehlerhaft wird. Holmqvist u. a. [Hol11] berichten in diesen Fällen von Verschlechterungen der Genauigkeit zwischen 0,1-0,3°. Wenn sich aufgrund der Veränderung der Stimulushelligkeit die Pupillenweite verändert, kann die Verschlechterung bis zu 1,5° betragen.

Die ***räumliche Präzision*** (engl. ***precision***) ist definiert als die Fähigkeit des Eyetrackers, eine Blickrichtungsschätzung zuverlässig reproduzieren zu können. Ist diese Kenngröße schlecht, können kleine Blickänderungen nicht erfasst werden, weil sie im Rauschen untergehen. Außerdem wird die Bestimmung von Fixationen und Sakkaden auf Basis der Blickrohdaten fehlerhaft. Die Präzision wird beeinflusst durch verschiedene Arten von Rauschen (systembedingt, okulomotorisch und Umgebungsrauschen) sowie von optischen Artefakten.

Das systembedingte Rauschen wird auch als *räumliche Auflösung* (engl. *spatial precision*) bezeichnet und ist definiert als die bestmögliche Präzision, die für einen Eyetracker erzielbar ist. Der Wert wird mithilfe künstlicher Augen bestimmt, die völlig unbewegt sind. Auf diese Weise ist sichergestellt, dass jede räumliche Abweichung in der Blickrichtungsschätzung durch den Eyetracker hervorgerufen wurde.

Das okulomotorische Rauschen (engl. *jitter*) bezeichnet die Mikrobewegungen des Auges, die während Fixationen auftreten (Tremor, Mikrosakkaden, Drift, s.o. S. 33).

Umgebungsrauschen besteht aufgrund mechanischer (der Benutzer stößt versehentlich an den Tisch, auf dem der Eyetracker steht und versetzt diesen in Bewegung) und elektromagnetischer Störungen.

Optische Artefakte sind falsche, unmögliche Blickbewegungen, die im Vergleich zu den physiologisch möglichen Blickbewegungen zu schnell sind. Sie

entstehen „…aus dem Zusammenspiel der optischen Situation (bspw. Brillen, Kontaktlinsen, zusätzlichen Reflexionen und Schatten und variierendem Umgebungslicht) und dem Algorithmus der Blickschätzung“ [Hol11].

Eine Verbesserung der Präzision kann durch algorithmische Filterung der Blickrohdaten erzielt werden. Der Preis hierfür ist eine Verlängerung der Latenz, mit der der Eyetracker das aktuelle Blickdatum liefert (s.u. Abschnitt 2.4.1).

Die ***Latenz*** eines Eyetrackers ist definiert durch die durchschnittliche Verzögerung zwischen tatsächlicher Augenbewegung und dem Zeitpunkt, zu dem der Eyetracker die Bewegung detektiert hat. Auch sie wird üblicherweise an künstlichen Augen bestimmt. Die Eyetracker-Latenz ist entscheidend bei Echtzeitanwendungen und dann, wenn andere technische Geräte mit dem Eyetracker synchronisiert werden müssen (z. B. ein Brain Computer Interface, das Hirnströme misst).

Holmqvist u. a. [Hol11] nennen eine konstante Latenz von drei Samples als wünschenswert: Während des ersten wird das Augenbild in der Kamera stabilisiert, während des zweiten wird das Augenbild in den Computerbuffer transferiert, während des dritten geschieht die Blickrichtungsschätzung. Drei Samples entsprechen bei einer Abtastfrequenz von 60 Hz $3 * 17$ ms = 54 ms, bei 1000 Hz $3 * 1$ ms = 3 ms. Um dies zu erreichen, wird oft ein separater Prozessor nur für die Blickrichtungsschätzung genutzt.

Wünschenswert ist auch eine hohe ***zeitliche Präzision***, definiert als die Standardabweichung der Eyetracker-Latenzen. In diesem Fall werden die Samples zwar mit Latenz geliefert. Das Intervall zwischen den Samples bleibt jedoch (fast) gleich. Ist die zeitliche Präzision gering, werden die Ergebnisse von Berechnungen von Zeitdauermaßen wie Fixationsdauer oder Sakkadendauer verfälscht.

Die ***zeitliche Auflösung*** von Remote-Eyetrackern liegt bei Abtastfrequenzen zwischen 50 und 2000 Hz. Höhere Abtrastfrequenzen haben den Vorteil, dass sie auch schnellere Augenbewegungen wie Tremor oder Mikrosakkaden detektieren können. Laut Martinez-Conde u. a. [Mar09] nutzen zwei Drittel der Mikrosakkaden-Untersuchungen Eyetracker mit einer Abtastrate von 500 Hz.

Ein weiterer Vorteil höherer Abtastfrequenzen ist, dass sie länger dauernde Augenbewegungen wie Fixationen und Sakkaden präziser detektieren, da sie deren Start und Ende (auch „Onset“ bzw. „Offset“ genannt) besser annähern [Hol11].

Der Preis hierfür ist eine größere Belastung des Benutzers. Denn bei einer höheren Abtastfrequenz ist die Infrarot-Beleuchtungsdauer für jedes Augenkamerabild kürzer, was durch stärkere Infrarot-Beleuchtung ausgeglichen werden muss.

Remote-Eyetracker, die hohe Datenqualität garantieren, sind nach wie vor sehr teuer. Bei niedrigen Abtastfrequenzen wie beim Tobii X60 liegt der Preis bei ca. 20.000 Euro; bei hohen Abtastraten bis zu 1.000 Hz wie beim EyeLink 1000 Plus[1] liegt der Preis bei ca. 30.000 Euro. Während solche Preise im Umfeld von Forschungseinrichtungen akzeptabel sein mögen, verhindern sie, dass Benutzer in großer Zahl einen Eyetracker als Standard-Peripheriegerät zur blickbasierten Informationseingabe nutzen.

Kostengünstige Eyetracker wurden über lange Jahre hinweg zunächst von Forschungseinrichtungen selbst entwickelt, die nicht bereit waren, die hohen Preise zu bezahlen. Diese Geräte waren jedoch nicht käuflich zu erwerben. Vereinzelt brachten auch kommerzielle Hersteller kostengünstige Eyetracker auf den Markt, die preislich bei ca. 500 bis 2.000 Euro lagen, oft aber rasch wieder vom Markt verschwanden.

Inzwischen hat Tobii einen kostengünstigen Eyetracker etabliert, der für blickbasierte Interaktion in Computerspielen gedacht ist und ähnlich einfach wie andere Peripheriegeräte über USB integrierbar ist. Die erste Version war der Tobii EyeX Ende des Jahres 2014[2]. Ende des Jahres 2016 kam der *Tobii Eye Tracker 4C* zu einem Preis von 149 US-Dollar auf den Markt[3], der bezüglich Robustheit deutlich verbessert war.

[1] https://www.sr-research.com/eyelink-1000-plus/

[2] https://www.tobii.com/group/news-media/press-releases/tobii-releases-eyex-plug-in-for-unreal-engine-4/

[3] https://www.tobii.com/group/news-media/press-releases/2016/10/tobii-releases-next-generation-gaming-eye-tracker/

Das Nachfolgegerät aus dem Jahr 2020 ist der *Tobii Eye Tracker 5*, der bei Erscheinen einen Preis von 229 Euro hatte und aktuell 183 Euro kostet[1]. Dieser Preis liegt etwas über dem für hochwertige Gaming-Computermäuse wie beispielsweise der *Razer Viper Ultimate*-Computermaus zu 170 Euro[2].

Im Gegensatz zu den ausführlichen technischen Spezifikationen, die die Hersteller für die teuren Geräte liefern, macht Tobii zu den kostengünstigen Geräten vergleichsweise wenige Angaben. So gibt es beispielsweise keine Angaben zur räumlichen Genauigkeit (engl. *accuracy*) oder zur Latenz.[3][4]

2.3 MMI: Manuelle Selektionsoperation

Die Selektionsoperation ist eine Basisoperation bei der MMI und kommt daher sehr häufig vor. Jede Selektionsoperation hat zwei Eingabekomponenten. Die räumliche Komponente definiert den Ort, an dem die Selektion erfolgen soll; die zeitliche definiert den Zeitpunkt, zu dem die Selektion erfolgen soll. Die räumliche Komponente wird auch als *Zeigeoperation* oder kurz *Zeigen* bezeichnet, die zeitliche als *Selektionsauslösung*.

Bei manueller Selektionsoperation unterscheidet man Eingabegeräte mit *direkter Manipulation* wie den Touchscreen und solche mit *indirekter Manipulation* wie die Computermaus.

2.3.1 Eingabegeräte mit direkter Manipulation

Shneiderman u. a. [Shn18, S. 348] attestieren Eingabegeräten mit direkter Manipulation einfache Erlernbarkeit und einfache Benutzung.

[1] https://gaming.tobii.com/product/eye-tracker-5/

[2] https://www.razer.com/de-de/gaming-mice/razer-viper-ultimate/RZ01-03050100-R3G1

[3] https://help.tobii.com/hc/en-us/articles/213414285-Specifications-for-the-Tobii-Eye-Tracker-4C

[4] https://help.tobii.com/hc/en-us/articles/360008539058-What-s-the-difference-between-Tobii-Eye-Tracker-4C-and-5-

MacKenzie [Mac12] nennt als erste Benutzungsschnittstelle mit Zeigegerät die Kombination aus *Sketchpad* und *Light Pen*, 1962 von Sutherland vorgestellt. Der Light Pen ermöglichte dem Benutzer direkte Manipulation auf dem Bildschirm. Da der Light Pen in die Luft vor den Bildschirm gehalten werden musste, führte die Interaktion jedoch zu rascher Ermüdung des Benutzers.

Direkte Manipulation mit *Touchscreen* wurde Ende der 1960er-/Anfang der 1970er-Jahre für die Anwendung der Luftverkehrskontrolle vorgestellt [Orr68, Gäe80] bzw. für Teilchenbeschleuniger am CERN [Low74]. 1973 wurde das erste Patent für einen optischen Touchscreen erteilt [Ebe73].

Nachteilig bei direkter Manipulation ist, dass Pixel-genaue Selektion sehr schwierig ist. Dies liegt zum einen daran, dass der Benutzer mit seinem Finger den Selektionsort auf dem Bildschirm verdeckt [Shn18]. Zum anderen umfasst der Kontakt einen Bereich des Bildschirms, keinen einzelnen Pixelpunkt (bekannt als „fat finger problem“ [Wig07]. Wie gut Touchscreen-Interaktion gelingt, wird von der Fingerform und -physiologie beeinflusst, sodass Größe und Form des Kontaktbereichs individuell variieren.

Heute werden Touchscreens vor allem für die Interaktion mit Smart Phones genutzt, teilweise auch für Laptops oder für horizontal angebrachte Monitore wie etwa den Digitalen Lagetisch [Str17]. Bei Anwendungen mit vertikalen Desktop-Monitoren sind Touchscreens eher die Ausnahme. Denn der Abstand zwischen Benutzer und Monitor ist zu groß, als dass er bequem mit einer Armlänge ohne zusätzliche Anpassung des Sitzabstands erreichbar wäre. Zudem wird den Arm geradeaus auszustrecken rasch anstrengend und ermüdend.

2.3.2 Eingabegeräte mit indirekter Manipulation

Weniger ermüdend als die direkten Techniken zeigte sich von Beginn an die Interaktion mit der 1963 von Engelbart vorgestellten *Computermaus*, heute Standardausstattung jedes Desktop-Computersystems. Im Gegensatz zu Touchscreen und Light Pen kann sie nahe bei der zur Texteingabe genutzten Computertastatur operiert werden.

Als (x,y)-Positionszeiger ermöglicht die Computermaus indirekte Manipulation auf dem Bildschirm, indem eine Zuordnung zwischen dem Bildschirmraum (engl. *display space*) und dem Raum des Eingabegeräts (engl. *device space*) hergestellt wird. Bereits die erste Computermaus verfügte über einen Knopf, der eine Selektionsauslösung realisierte. So wurde bereits damals für Maus-Interaktionen der Begriff *point-click* bzw. *point-select* üblich.

Ab 1981 war die Computermaus kommerziell verfügbar mit dem Xerox Star System, dem ersten kommerziell verfügbaren Computersystem mit grafischer Benutzungsoberfläche [Mac12]. Dieses System war für die Büroautomatisierung entwickelt worden. Es verfügte über einen Bildschirm mit Rastergrafik, auf dem Fenster, Icons[1], Menüs und ein Zeigegerät (mithilfe des Mauszeigers) dargestellt wurden.

Diese Art der Benutzungsoberfläche wird auch als WIMP-Paradigma bezeichnet (von engl. *W*indows, *I*cons, *M*enus, *P*ointing device). Mit ihr kann die Desktop-Metapher realisiert werden. Anstatt über die Kommandozeile mit dem Computer zu kommunizieren, kann der Benutzer Aktionen im Computer starten, indem er einfach auf ein Icon zeigt oder das Icon selektiert. Die Komplexität des Computersystems wird hinter den Icons vor dem Benutzer verdeckt, was die Benutzung einfach und intuitiv macht.

Weitere Eingabegeräte für indirekte Manipulation sind der Trackball, Joystick, Trackpoint und Touchpad bei Laptops. Shneiderman u. a. [Shn18, S. 348] bemerken zu Geräten mit indirekter Manipulation, dass diese im Gegensatz zu denen mit direkter Manipulation eine deutlich längere Zeit zum Erlernen benötigen.

2.3.3 Selektion statischer Objekte

English u. a. [Eng67] veröffentlichten die mutmaßlich erste Nutzerstudie der MMI [Mac12] zu einem kontrollierten Experiment, in dem die Vorteile der Computermaus gezeigt werden konnten. Die Computermaus wurde dort

[1] DUDEN (duden.de/rechtschreibung/Icon): „grafisches Symbol für Anwendungsprogramme, Dateien u.Ä. auf dem Bildschirm“.

für die Aufgabe der Textauswahl mit anderen Eingabegeräten verglichen, die ebenfalls einerseits die (x,y)-Positionskontrolle eines Cursors auf dem Monitor und andererseits eine Selektionsauslösung realisierten. Dabei erzielte die Computermaus eine nur halb so große Fehlerquote wie die anderen Geräte. Die Computermaus war außerdem unter den schnellsten Geräten und komfortabler nutzbar als der etwas schnellere Light Pen (s. Abschnitt 2.3.1).

Im Fall dieses Experiments waren die zu selektierenden Objekte alle unbewegt. Diese Situation kann mithilfe des Modells von Fitts (oft auch als „Fitts' Law" bezeichnet) beschrieben werden [Fit64]. Der Zeitbedarf T, um auf ein Objekt zu zeigen, hängt von seiner Breite W und von der Distanz (auch Amplitude) A ab, die das Zeigegerät (bei der Computermaus der Mauszeiger) zu Beginn der Zeigeoperation zum Objekt hat:

$$T = a + b * log_2(A/W + 1) \tag{2.3}$$

wobei a und b empirisch bestimmte Konstanten sind.

Je breiter also ein Objekt ist und je näher das Zeigegerät bei Beginn des Zeigeprozesses bei diesem Objekt positioniert ist, desto kürzer ist der Zeitbedarf. Um die Selektion kleiner Objekte effizienter zu machen, wurden verschiedene Methoden entwickelt.

Beim Ansatz *Area Cursor* umfasst der effektive Aktivierungsbereich des Zeigegeräts keinen Punkt, sondern einen Bereich [Kab95]. Aus Sicht des Modells von Fitts wird dort die Breite des Objekts durch die Breite des Area Cursors ersetzt; es leuchtet ein, dass der Zeitbedarf sich dadurch verringert.

Nachteilig beim Area Cursor ist, dass er mit seiner festgelegten Aktivierungsgröße bei kleinen Objekten, die enger beieinander liegen, nicht ohne Weiteres funktioniert; meist wird die Selektion dem Objekt zugeschlagen, dessen Mittelpunkt oder Rand näher bei der Selektionsposition liegt.

Der *Bubble Cursor* versucht diesen Nachteil zu vermeiden. Er baut auf dem Area Cursor auf, passt jedoch die Größe seines Aktivierungsbereichs an in Abhängigkeit von Anzahl und Dichte der Objekte, sodass für jede gegebene Cursorposition stets nur ein Objekt selektiert wird [Gro05].

Ein alternativer Ansatz ist, statt des Aktivierungsbereichs des Cursors den Aktivierungsbereich des Objekts zu vergrößern. Ein Objekt hat dann eine sichtbare Größe und eine größere selektierbare Größe. Solche virtuellen *Objektvergrößerungen* ziehen jedoch Schwierigkeiten nach sich, wenn Objekte dicht platziert sind und sich die Aktivierungsbereiche überlappen, sodass die Selektionsposition nicht mehr eindeutig einem Objekt zugeordnet werden kann [McG02]. McGuffin u. a. [McG05] zeigten, dass eine dynamische Objektvergrößerung die Leistung eines Benutzers sogar dann noch verbessert, wenn sie erst dann geschieht, wenn der Benutzer 90% des Zeigeprozesses bereits erledigt hat. Sie betrachteten verschiedene Kombinationen aus Breite W und Distanz A (mit W zwischen 8 und 64 Pixel und A zwischen 128 und 1024 Pixel).

Anstatt die Breite W zu vergrößern, schlagen andere Autoren vor, die Distanz A zu verkleinern, um schneller selektieren zu können. Dafür werden für ein Objekt Stellvertreterobjekte generiert, die näher beim Cursor platziert werden, sodass die Amplitude des Zeigevorgangs verkleinert wird. Vorschläge sind z. B. der *Vakuum-Filter* [Bez05] oder *Drag-and-Pop* [Bau03] aus Untersuchungen an sehr großen Displays.

Solange die Objekte auf dem Bildschirm sich nicht bewegen, können sie mit der Computermaus vergleichsweise einfach selektiert werden, wobei die genannten Verbesserungen die Selektionsoperation beschleunigen können.

2.3.4 Selektion bewegter Objekte

Bewegen sich Objekte auf dem Bildschirm, wird ihre Selektion erheblich schwieriger. Jagacinski u. a. [Jag80] passten das Modell von Fitts für bewegte Objekte an. Der Zeitbedarf ist jetzt zusätzlich abhängig von der Geschwindigkeit V des Objekts:

$$T = a + b * A + c * (V + 1) * ([1/W] - 1) \tag{2.4}$$

wobei a, b und c empirisch bestimmte Konstanten sind.

Je schneller ein Objekt sich bewegt und je kleiner es ist, desto schwieriger ist es zu selektieren.

Beim Selektionsvorgang folgen die Benutzer typischerweise einer von zwei Strategien. Entweder sie verfolgen das Objekt mit dem Zeigegerät (z. B. Mauszeiger bei der Computermaus oder Finger beim Touchscreen) und lösen die Selektion aus, sobald sie das Zeigegerät über das Objekt bewegt haben. Oder sie schätzen die Bewegung des Objekts zu einem Zeitpunkt in der nahen Zukunft, bewegen das Zeigegerät an diesen Bildschirmort, warten und lösen die Selektion aus, sobald das Objekt sich an diesem Ort befindet.

Die erste Strategie funktioniert prinzipiell für alle Objektbewegungen; je schneller und unvorhersagbarer sich das Objekt bewegt, desto mehr Selektionsversuche sind zu erwarten. Die zweite Strategie funktioniert nur dann, wenn die Bewegung des Objekts zuverlässig vorhersagbar ist.

Für die Erleichterung und Vereinfachung der Bewegtobjektselektion wurden als Ansätze *Bewegungspausierung* und *Objektvergrößerung* vorgeschlagen.

Ilich [Ili09] schlägt vor, die Bewegung zu pausieren. Auf diese Weise wird die Bewegtobjektselektion auf eine Selektion eines unbewegten Objekts zurückgeführt. Bei der Variante *Click-to-Pause* drückt der Benutzer die linke Maustaste und stoppt dadurch alle Bewegung auf dem Bildschirm, er bewegt dann den Mauszeiger (bei weiterhin gedrückter Maustaste) auf das Objekt und bewirkt die Selektionsauslösung durch Loslassen der Maustaste. Bei der Variante *Chase-or-Pause* wird beim Mausklick unmittelbar eine Selektion ausgelöst, falls sich der Mauszeiger über einem Objekt befindet; falls sich unter dem Mauszeiger kein Objekt befindet, gilt der Mechanismus von Click-to-Pause. Ilich zeigt, dass diese beiden Varianten die Selektionszeit im Vergleich zur Standard-Mauseingabe (Ilich nennt diese *Chase-and-Click*) für kleine oder schnelle Objekte verringern. Click-to-Pause ist am schnellsten bei kleinen und schnellen Objekten.

Ragan u. a. [Rag20] betrachten ebenfalls Ansätze zur Bewegungspausierung. Sie berichten, dass das Pausieren der gesamten Szene die Selektionsleistung signifikant verbessert, dabei jedoch das Situationsbewusstsein des Benutzers beeinträchtigt werden kann. Pausieren von Teilen der Szene verbessert die

Selektionsleistung ebenfalls signifikant und beeinträchtigt das Situationsbewusstsein weniger. Die Variante *Cursor Proximity* pausiert alle Objekte (Größe 0,15 cm, 5 Pixel) im Umkreis von einem Radius von 2,2 cm (70 Pixel) um die Cursorposition. Die Variante *Trajectory Pausing* bewirkt, dass Objekte pausiert werden, die sich in dem Bereich befinden, der ab der aktuellen Cursorposition in Bewegungsrichtung des Cursors einen Winkel von 50° umfasst; diese Variante ermöglicht das Pausieren entfernter Objekte. Sowohl für die schnelleren (3,7 cm/s) als auch für die langsameren (1,4 cm/s) Objekte lieferten diese beiden Techniken signifikant und substanziell geringere Selektionsfehlerquoten als die traditionelle Mauseingabe ohne Pausieren.

Hasan u. a. [Has11] untersuchen den Ansatz des Pausierens in Kombination mit einem Stellvertreter. Diese Methode, *Ghost*, erstellt, sobald der Benutzer auf den Bildschirm klickt, für alle bewegten Objekte einen unbewegten Stellvertreter. Während der Benutzer den passenden unbewegten Stellvertreter selektiert, bewegen sich die eigentlichen Objekte weiter. Hasan u. a. untersuchten außerdem den Ansatz einer Objektvergrößerung inspiriert vom Area bzw. Bubble Cursor (s.o. Abschnitt 2.3.3). Diese Methode, *Comet*, realisiert die Objektvergrößerung in Form eines Kometenschweifs, dessen Form in Abhängigkeit von Objektbreite und Objektgeschwindigkeit variiert. Für die Schweiflänge gilt, je schneller das Objekt, desto länger der Schweif:

$$\text{Schweiflänge} = (V/c) + (W/2) \tag{2.5}$$

wobei $c = 1{,}6$ durch initiale Pilottests empirisch bestimmt wurde.

Die Objektgrößen der Experimente betrugen 50, 75, 100 Pixel, die Objektgeschwindigkeiten 500, 650, 800 Pixel/s (1 Pixel entsprach 0,28 mm auf dem genutzten Monitor).

Da im ersten Experiment (Bewegtobjektselektion in 1D) die Benutzer die Selektion vor allem an der Schweifspitze platzierten, wurde für ein zweites Experiment (Bewegtobjektselektion in 2D) der Schweif so gestaltet, dass er sich vom Objekt zur Spitze auf Schweiflänge = Objektdurchmesser ∗ 1,5 verbreiterte. Überlappen sich zwei Schweife, so wird das Objekt selektiert, das näher an der Selektionsposition liegt.

Hasan u. a. betrachteten Comet, Bubble Cursor, Area Cursor und traditionelle Mauseingabe jeweils alleine und in Kombination mit Ghost. Die Selektionsfehlerquoten waren für die Ghost-Varianten durchweg niedriger. Ohne Ghost erzielte bei der Bewegtobjektselektion in 1D der Bubble Cursor die niedrigste Selektionsfehlerquote, in 2D Comet. Die kürzeste Selektionszeit erzielte bei der Bewegtobjektselektion in 1D Comet, in 2D war BubbleGhost am schnellsten, gefolgt von CometGhost, Comet und Bubble.

2.4 MMI: Blickbasierte Informationseingabe

Laut Zhai [Zha03] gilt der menschliche Blick als der beste Vermittler für die menschliche Aufmerksamkeit oder Intention. Misst man also eine Fixation an einem bestimmten Ort, so misst man zugleich auch die Aufmerksamkeit, die auf diesem Ort liegt. Eine Ausnahme ist, wenn der Benutzer ins Leere starrt, ohne visuelle Information aufzunehmen [Hol11].

Als einer der ersten schlug Bolt [Bol82] vor, den Blick neben Sprache und Gesten für multimodale MMI einzusetzen. Er postulierte insbesondere die Nutzung von Blickinformation für Zeigeoperationen als intuitiv. Denn Blickbewegungen erfolgen einerseits oft automatisch, ohne dass der Mensch über das Blickziel nachdenkt („Bottom-up“). Andererseits kann der Mensch seinen Blick auch gezielt willentlich ausrichten und seine Aufmerksamkeit aktiv auf einen Ort lenken („Top-down“, s. Abschnitt 2.1). Wenn nun ein Benutzer mithilfe seiner Augen visuelle Information vom Bildschirm aufnimmt, liefert die Ausrichtung der Augen gleichzeitig umgekehrt einen Hinweis, auf welchem Bildschirm-Ort der aktuelle visuelle Aufmerksamkeitsfokus liegt [Jus76].

Mehrere Autoren formulieren, dass der Blick deutlich mehr Information zum Nutzungskontext des Benutzers enthüllt als Maus oder Computertastatur [Jac90, Zha03]. Jacob [Jac90] weist in diesem Zusammenhang darauf hin, dass die Bandbreite der Informationsübermittlung vom Computer zum Menschen weitaus größer ist als umgekehrt, wenn einzig Computermaus und Tastatur zur Eingabe dienen. Während der Computer-Bildschirm den Systemzustand detailreich darstellt, liefern die manuellen Eingabegeräte nur

eine Information über den Benutzerzustand, wenn dieser die Entscheidung für eine manuelle Eingabe tätigt. Wie der Benutzerzustand darüber hinaus beschaffen ist, ist dem Computersystem unbekannt. Eine Ergänzung der Benutzungsschnittstelle um Informationen zum Blickverhalten kann daher zu mehr Ausgewogenheit beitragen [Jac90].

Blickbasierte Interaktion wurde neben der Nutzung für Zeige- und Selektionsoperationen auch für die Texteingabe (engl. *Eyetyping*), für die Interaktion in Menüs, zum Scrollen oder zur Passworteingabe untersucht (vgl. z. B. [Maj02, Špa05b, Kam08, Kum07c, Kum07a]). Für die vorliegende Arbeit sind sie wenig relevant und werden daher im Weiteren nicht vertieft.

Im Folgenden beschreibt Abschnitt 2.4.1 den Aspekt der Blickdatenfilterung, der für praktisch alle Arten der Blickinteraktion relevant ist. Danach beschreibt Abschnitt 2.4.2 verwandte Arbeiten zur blickbasierten Selektionsoperation, Abschnitt 2.4.3 zur blickbasierten Selektion bewegter Objekte, Abschnitt 2.4.4 zur Interaktion mit Blick und Brain Computer Interfaces (BCI) und Abschnitt 2.4.5 Arbeiten zur blickbasierten Aufgaben- bzw. Tätigkeitsschätzung.

2.4.1 Blickdatenfilterung

Wie oben erwähnt, liefert ein Eyetracker nicht präzise die aktuelle Blickrichtung, sondern ein verrauschtes Signal. Die Datenqualität wird dabei sowohl von Eigenschaften des Eyetrackers als auch von Eigenschaften des Benutzers beeinflusst. Selbst wenn der Eyetracker stets die korrekte Blickrichtung liefern würde und der Benutzer den Kopf stillhalten könnte, würde der menschliche Blick selbst noch ein verrauschtes Signal liefern, weil das Auge bei Fixationen Mikrobewegungen durchführt (vgl. Abschnitt 2.1.3, S. 33).

Die meisten Anwendungen von Eyetracking, darunter auch die in der vorliegenden Arbeit relevante MMI, benötigen keine Blickinformation auf der Detailstufe von Mikrobewegungen. Sie nutzen vielmehr Blickinformation, die davon abstrahiert in Form von Fixationen (ruhender Blick auf Bereichen von

Interesse zur visuellen Informationsaufnahme) und Sakkaden (schnelle Blickbewegungen zwischen Fixationen) [Sal00, Špa12].

Ziel von Blickfilterverfahren ist daher zum einen, die Mikrobewegungen während Fixationen zu glätten. Zum anderen soll auch alle Variation aus dem Blicksignal entfernt werden, die nicht von Augenbewegungen herrührt. Bei Anwendungen, die Blick zur Echtzeit-Informationseingabe in ein System nutzen, muss auch die Filterung in Echtzeit erfolgen. Bei Anwendungen, wo die Blickdatenanalyse nach der Aufzeichnung offline erfolgt, ist keine Echtzeitanforderung für den Filteralgorithmus notwendig.

Abb. 2.4 zeigt Blickdaten aufgezeichnet im Rahmen der vorliegenden Arbeit mit dem Tobii X60 über eine Zeitdauer von 2 Sekunden. Die rote Linie visualisiert die Blickrohdaten der x-Koordinate, die grüne die Blickrohdaten der y-Koordinate. Der Benutzer fixiert nacheinander drei Zielobjekte mit 100 Pixeln Durchmesser. Zu Beginn fixiert er ein Zielobjekt, dessen Mittelpunkt sich auf dem Bildschirm an der (x,y)-Position (1200, 950) befindet. Bei Sample 11 vollzieht er eine Sakkade zum Zielobjekt mit Mittelpunkt (1420, 850), das er als nächstes fixiert. Während dieser Fixation passiert bei Sample 31 ein optisches Artefakt mit einem Ausschlag von einem Sample mit sehr hoher Amplitude. Bei Sample 86 vollzieht der Benutzer eine Sakkade zum Zielobjekt mit Mittelpunkt (950, 450) und fixiert es. Während der drei Fixationen erkennt man das hochfrequente Rauschen mit kleiner Amplitude, das durch die Messunsicherheit des Eyetrackers und die Mikrobewegungen der Augen erzeugt wird.

Die graue und die schwarze Linie visualisieren die mit dem echtzeitfähigen Algorithmus von Kumar u. a. [Kum08] gefilterten Blickdaten, genannt ***Real-Time Saccade Detection and Fixation Smoothing***, im Folgenden kurz *RT-SDFS*. Man kann erkennen, dass dieser Algorithmus die Blickdaten während der Fixationen glättet und auch das optische Artefakt bei Sample 31 eliminiert. Der Preis hierfür ist eine Latenz von einem Sample im Falle von Sakkaden. Man sieht dies bei Sample 11/12 und 86/87: Hier liefert der gefilterte Wert ein Sample zu spät, sodass der Blick bereits auf dem neu fixierten Punkt liegt.

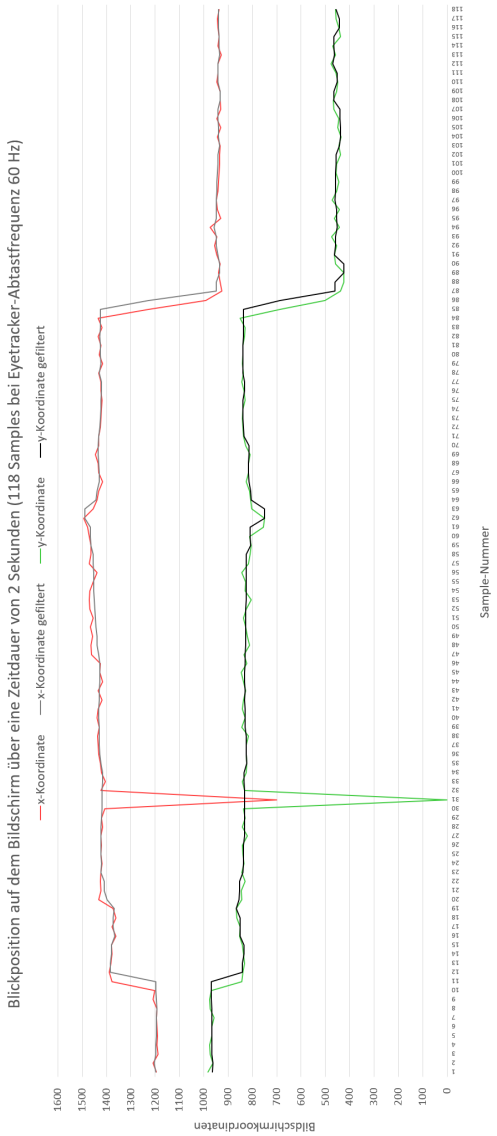


Abbildung 2.4: Blickdaten, roh und gefiltert.

Der *RT-SDFS* ist weit verbreitet bei Anwendungen, die Echtzeit-Filterung für Blickeingabe benötigen. Die Autoren geben selbst an, dass die Verbesserung der Präzision bei ihren eigenen Experimenten die Selektionsfehlerquote stark genug reduzierte, dass Objektselektion auf Basis der Blickposition praktisch nutzbar wurde [Kum07b, Kum08].

Der Algorithmus liefert die gefilterte Blickposition als einen gewichteten Durchschnitt innerhalb eines dynamischen Zeitfensters. Er adressiert die Besonderheiten von Blickbewegungen und unterscheidet dabei drei Situationen für das aktuell gemessene Blickrohdatum $\mathbf{r}_0 = (x_0, y_0)$:

1. $\mathbf{r}_0$ ist der Beginn einer Sakkade.
2. $\mathbf{r}_0$ ist die Fortsetzung einer Fixation.
3. $\mathbf{r}_0$ ist ein Ausreißer innerhalb einer Fixation.

Wenn das aktuelle Blickdatum $\mathbf{r}_0$ der Beginn einer Sakkade ist (Fall 1), ist der euklidische Abstand zum vorangehenden Blickdatum $\mathbf{r}_{-1}$ deutlich größer als in den beiden anderen Fällen 2 und 3. Der Algorithmus nutzt daher einen Schwellenwert S für den euklidischen Abstand zwischen je zwei Blickdaten. Gilt $d_{\mathbf{r}_0,\mathbf{r}_{-1}} = \sqrt{(x_0 - x_{-1}))^2 + (y_0) - y_{-1}))^2} > S$, wird $\mathbf{r}_0$ als Beginn einer Sakkade identifiziert. Diese Vorgehensweise ist vergleichbar mit der des Verfahrens *I-VT* [Sal00], s.u. in diesem Abschnitt.

Das hochfrequente Rauschen mit geringer Amplitude der Mikrobewegungen wird mit zwei Mechanismen behandelt:

1. Der Abstand von $\mathbf{r}_0$ wird nicht zur Position des vorangegangenen Blickdatums $\mathbf{r}_{-1}$ berechnet, sondern zur aktuellen Schätzung der Fixationsposition $\hat{\mathbf{f}}_0 = (\hat{x}_0, \hat{y}_0)$.
2. Wenn der Abstand $d_{\mathbf{r}_0,\hat{\mathbf{f}}_0}$ den Schwellenwert S übersteigt, wird $\mathbf{r}_0$ nicht unmittelbar als Beginn einer Sakkade identifiziert. Stattdessen wird zunächst der Abstand des nachfolgenden Blickdatums $\mathbf{r}_1$ zur geschätzten Fixationsposition $\hat{\mathbf{f}}_0$berechnet. Falls dieser Abstand unterhalb des Schwellenwertes S liegt, gilt $\mathbf{r}_0$ als Ausreißer innerhalb einer ansonsten stabilen Fixation. Diese Vorausschau induziert die

oben beschriebene Latenz von einer Blickschätzung bei der Sakkadendetektion (vgl. Abb. 2.4).

Die aktuelle Fixationsposition $\hat{\mathbf{f}}_0 = (\hat{x}_0, \hat{y}_0)$ wird als gewichtetes Mittel (einseitiger triangulärer Filter) berechnet. Aktuellere Blickpunkte werden stärker gewichtet, um leichte Drift in den Blickdaten schneller zu berücksichtigen. Für n Blickpunkte erfolgt die Berechnung als

$$\hat{\mathbf{f}}_0 = \frac{1\mathbf{r}_{-(n-1)} + 2\mathbf{r}_{-(n-2)} + \cdots + n\mathbf{r}_0}{1 + 2 + \cdots + n}. \tag{2.6}$$

Špakov [Špa12] zeigte in einem Vergleich von sieben verschiedenen Blickfilterverfahren, dass Filterverfahren mit solchen triangulären Kernelfunktionen mit die besten Ergebnisse liefern. Der *RT-SDFS* gehörte zu einem der zwei Verfahren mit der besten Leistung (bzgl. der Kenngrößen Latenz, Ähnlichkeit zum Originalsignal, Glättung).

Ebenfalls echtzeitfähig ist Blickfilterung mithilfe eines ***Kalman-Filters***. Kalman-Filter werden grundsätzlich eingesetzt, um durch rekursive Schätzung den zukünftigen Zustand eines dynamischen Systems aus einer Reihe unvollständiger und verrauschter Messwerte zu bestimmen. Dazu wird der mittlere quadratische Fehler zwischen der Prädiktion des Systemzustands und des aktuellen Messwerts bestimmt. Um eine neue Schätzung für den aktuellen Systemzustand zu erhalten, sind nur der geschätzte Zustand des vorangehenden Zeitschritts sowie die neue Messung notwendig.

Komogortsev u. a. [Kom07] stellen den *Attention Focus Kalman Filter (AFKF)* basierend auf der Arbeit von [Sau91] vor, der Fixationen, Sakkaden und Augenfolgebewegungen detektiert. Die Evaluation erfolgte am Videocomputerspiel World of Warcraft, wobei die Versuchspersonen einen Avatar kontrollierten. Der Mauszeiger wurde dabei auf Basis des *AFKF* blickbasiert gesteuert. Es zeigte sich, dass das Verfahren die Blickrohdaten erfolgreich für die Nutzung zur blickbasierten Systemsteuerung filtern sowie das Blicksignal vorhersagen konnte, wenn das Eyetracker-Signal kurzzeitig ausfiel.

In der vergleichenden Evaluation von Špakov [Špa12] zeigte der *AKFK* jedoch mit die schlechteste Leistung und zwar sowohl bezüglich der Glättung als auch bezüglich der Ähnlichkeit zum Originalsignal.

Auch Kumar u. a. [Kum08] vergleichen ihren *RT-SDFS* (s.o.) mit einem Kalman-Filter (ohne Angabe von dessen Implementierungsdesign) und berichten, dass ihr einseitiger triangulärer Filter während Fixationen besser glättet als der Kalman-Filter. Kumar u. a. weisen darauf hin, dass eine Verbesserung für Kalman-Filter-basierte Blickdatenfilterung möglicherweise mithilfe eines besseren Prozessmodells für Blickbewegungen bzw. mithilfe nicht-linearer Kalman-Filter-Varianten [Aru02] möglich sei. Kumar u. a. weisen jedoch auch darauf hin, dass der *RT-SDFS* Echtzeit-Blickdatenfilterung mit einem im Vergleich zum Kalman-Filter simplen Design bewerkstelligt.

Ein anderer, weit verbreiteter Blickfilteralgorithmus ist der ***I-VT*** (abgekürzt für engl. *Velocity-Threshold Fixation Identification*) von Salvucci u. a. [Sal00]. Er wurde zunächst vorgestellt für die Offline-Detektion von Fixationen und Sakkaden in Blickrohdatenprotokollen, ist jedoch auch echtzeitfähig. Das Ziel ist, Sakkadenpunkte aus dem Blickrohdatenprotokoll zu eliminieren, da während Sakkaden keine Informationsaufnahme stattfindet.

Der *I-VT* separiert Blickdatenpunkte, die zu Fixationen gehören, von solchen, die zu Sakkaden gehören, mithilfe eines Schwellenwertes v_{max}, der die maximal erlaubte Geschwindigkeit innerhalb einer Fixation festlegt.

Im ersten Schritt vergleicht der Algorithmus Blickpunkt-zu-Blickpunkt-Geschwindigkeiten. Liegt die Geschwindigkeit unterhalb des Schwellenwertes v_{max}, wird der Blickpunkt als Fixationspunkt klassifiziert, liegt sie oberhalb, ist er ein Sakkadenpunkt.

Im zweiten Schritt werden unmittelbar aufeinander folgende Fixationspunkte zu Fixationen aggregiert. Die (x,y)-Fixationsposition wird als Zentroid aus allen Fixationspunkt-Koordinaten berechnet. Die Fixationszeit definiert sich durch den Zeitpunkt des ersten Fixationspunkts (auch: Fixations-Onset), die Fixationsdauer als die Differenz zwischen letztem und erstem Fixationspunkt.

Wie der Schwellenwert am besten zu wählen ist, muss je nach Anwendung individuell entschieden werden. Salvucci u. a. [Sal00] geben als groben Richtwert an, dass Fixationen Geschwindigkeiten < 100°/s annehmen, Sakkaden Geschwindigkeiten > 300°/s. Andere Quellen nennen jedoch deutlich überlappende Werte, etwa 30-500°/s für Sakkaden und 15-50°/s für die Mikrosakkaden innerhalb einer Fixation [Hol11], vgl. a. S. 33, sodass die Trennung nicht in allen Fällen ohne Weiteres gelingt. Angesichts dessen, dass der Überlappungsbereich bei 30-50°/s liegt, ist ein Schwellenwert in diesem Bereich naheliegend. Komogortsev u. a. [Kom07] nehmen für Fixationen Geschwindigkeiten < 5°/s und für Sakkaden Geschwindigkeiten > 30°/s an.

Olsen u. a. [Ols12b] bzw. Olsen [Ols12a] schlagen 30°/s vor, weisen jedoch darauf hin, dass bei stärker verrauschten Daten, wie sie bei Blickerfassung ohne Kopfstabilisierung typisch sind, höhere Schwellenwerte vermutlich bessere Ergebnisse liefern. Auch Koh u. a. [Koh09] weisen darauf hin, dass der *I-VT* wenig tolerant auf Rauschen reagiert, das von Eyetracker-Messunsicherheiten oder Mikrobewegungen der Augen herrührt; sie wählen in ihrem System einen Schwellenwert von 75°/s.

Ein weiterer populärer Algorithmus, der häufig zur Offline-Detektion von Fixationen in Blickrohdatenprotokollen genutzt wird, ist der ***I-DT*** (abgekürzt für engl. *Dispersion-Threshold Fixation Identification*) von Salvucci u. a. [Sal00]. Eine Folge von Blickrohdaten wird als Fixation zusammengefasst, wenn sie bestimmte zeitliche und räumliche Grenzen einhalten.

Die zeitliche Bedingung bewertet Fixationsdauern kleiner als ein Schwellenwert t_{min} als unphysiologisch und eliminiert sie als invalide. Die räumliche Bedingung nutzt einen Schwellenwert d_{max} und fordert für eine Fixation, dass die Summe der maximalen horizontalen und der maximalen vertikalen Distanz zwischen allen Einzelblickpunkten kleiner ist:

$$(x_{max} - x_{min}) + (y_{max} - y_{min}) < d_{max} \tag{2.7}$$

Andere Dispersions-basierte Verfahren nutzen alternative Metriken zur Berechnung des Schwellenwertes d_{max}, bei denen der Berechnungsaufwand jedoch höher ist [Shi08].

Der *I-DT* arbeitet mit einem dynamischen Fenster, das aufeinanderfolgende Blickrohdaten aufnimmt und die zeitliche und räumliche Bedingung überprüft. Wird nach Hinzufügen eines Blickpunktes der räumliche Schwellenwert überschritten, beschreiben die Blickpunkte keine Fixation. In diesem Fall bewegt sich das Fenster um einen Punkt weiter. Wird der räumliche Schwellenwert nicht überschritten, beschreiben die Blickpunkte eine Fixation und das Fenster wird um diesen Blickpunkt erweitert; dies wird so lange fortgesetzt, bis der räumliche Schwellenwert überschritten wird.

Für eine detektierte Fixation wird dann die (x,y)-Fixationsposition als Zentroid aus den Koordinaten aller im Fenster enthaltenen Blickpunkte berechnet; der Zeitpunkt des ersten Blickpunktes gilt als Fixationsbeginn (engl. *Fixation Onset*). Die Fixationsdauer berechnet sich aus der Differenz zwischen erstem und letztem Blickpunkt des Fensters.

Salvucci u. a. [Sal00] schlagen vor, den räumlichen Schwellenwert d_{max} so zu wählen, dass er 1° Sehwinkel umfasst, oder ihn auf Basis einer explorativen Analyse der vorliegenden Blickrohdaten angemessen zu schätzen. Als zeitlichen Schwellenwert t_{min} empfehlen sie 100 bis 200 ms (in Anlehnung an [Wid84]).

2.4.2 Blickbasierte Selektionsoperation

Shneiderman u. a. [Shn18] bringen Eyetracker als Zeigegeräte unter der Rubrik „Novel devices and strategies (for special purposes)“, in die sie auch etliche andere Geräte eingruppieren, z. B. verschiedene Sensoren (Gyroskop, Beschleunigungssensor, Tiefenkamera), Datenhandschuhe, Haptisches Feedback und Fußtasten.

Eyetracking zum Zeigen auf dem Bildschirm zu nutzen, gilt wie eingangs von Abschnitt 2.4 beschrieben als inuitiv und daher naheliegend für eine Nutzung für die MMI. Der Benutzer nutzt seinen Blick dabei doppelt: einerseits zur visuellen Informationsaufnahme, der typischen Aufgabe der Augen, und andererseits zur Informationseingabe, die typischerweise manuell erfolgt.

Um zu verhindern, dass jede Blick-Zeigeposition auf dem Monitor eine Systemeingabe bewirkt, muss ein Mechanismus implementiert werden, der Blickpunkte auf dem Bildschirm, die zur Informationseingabe genutzt werden sollen, eindeutig und zuverlässig identifiziert.

Gelingt dies nicht, resultiert der sogenannte *Midas Touch-Effekt*[1] [Jac90]: Im schlimmsten Fall werden alle Blickpunkte auf dem Monitor als Eingabekommando interpretiert, sodass unablässig ungewollte Systemeingaben stattfinden.

So wie bei der Mauseingabe der Mausklick bestimmt, wann an der aktuellen Mauszeigerposition eine Selektion stattfindet, benötigt auch die Blickeingabe einen Mechanismus zur Selektionsauslösung, der bestimmt, wann an der aktuellen Blickposition eine Selektion stattfinden soll. Wenn man diesen Mechanismus geeignet wählt, kann der Midas Touch-Effekt verhindert werden.

Die Vorschläge der Fachliteratur unterscheiden *unimodale* und *multimodale* Realisierungen blickbasierter Interaktion. Beide nutzen den Blick zum Zeigen. Die unimodalen nutzen das Auge/den Blick jedoch zusätzlich für die Selektionsauslösung, während die multimodalen hierfür eine andere Modalität nutzen.

2.4.2.1 Unimodale Realisierungen

In der Fachliteratur vorgeschlagene *unimodale* Realisierungen sind:

- Verweildauer (engl. *dwell time*): Eine Selektion wird dann ausgelöst, wenn die Blickrichtung eine bestimmte Zeit lang auf demselben Ort/Objekt verweilt;
- Lidschlag;

[1] Dieser Effekt ist nach dem sagenhaften König Midas aus der griechischen Mythologie benannt. Dieser wünschte sich aus Gier vom Gott Dionysos, dass alles, was er, Midas, berühren würde, zu Gold werde. Als er bemerkte, dass ihm dadurch der Tod durch Verhungern und Verdursten drohte, bat er um Rücknahme der Gabe. Vgl. Ovid, Verwandlungen 11 (85-193) [Pub96].

- Zeigen/Verweildauer auf eine(r) separaten Softwaretaste (GUI-Schaltfläche).

Die ***Verweildauer*** wurde für die Selektion statischer Objekte häufig untersucht [War86, Jac91, Jac95, Sib00, Ver08] und gilt als schnelle Interaktionstechnik [Jac91]. Sie ist eine naheliegende Realisierung, da die Blickrichtung den Aufmerksamkeitsfokus eines Menschen enthüllt und eine Verweildauer ab 100 ms auf eine Fixation, d.h. auf visuelle Informationsaufnahme, hindeutet.

Die verwendeten Verweildauern unterscheiden sich stark. Vertegaal [Ver08] nutzte 100 ms, Sibert u. a. [Sib00] nutzten 150 ms, Ware u. a. [War86] 400 ms. Jacob [Jac95] nennt 200 bis 600 ms als typische Verweildauer, Majaranta u. a. [Maj02] nennen 600 bis 1000 ms, Majaranta u. a. [Maj09] 300 ms für geübte Benutzer bei ausreichender Qualität des Eyetrackings.

Die großen Unterschiede der genutzten Verweildauern weisen auf die unterliegende Problematik hin, nämlich eine auch bei Dauernutzung robuste und somit praktikable Verweildauer zu definieren. Eine kurze Verweildauer schließt die Selektionsoperation schnell ab, löst jedoch eher falsch positive Selektionen aus. Nutzt man demgegenüber eine sehr lange Verweildauer, etwa 2 oder 3 Sekunden, so wird die Anzahl falsch positiver Selektionen reduziert. Die Interaktionstechnik wird dann jedoch sehr zeitaufwändig.

Bei der Nutzung von ***Lidschlag*** ist die Schwierigkeit ebenfalls, eine praktikable Lidschlagdauer zu finden. Denn hier müssen Lidschläge, die unwillkürlich zum Schutz des Auges geschehen, von den willentlich gesteuerten Lidschlägen algorithmisch unterschieden werden [Jac91]. Wählt man die Lidschlagdauer kurz, sind viele falsch positive Selektionen zu erwarten, denn auch die unwillkürlichen Lidschläge werden mit einer Dauer von ca. 300 ms sehr rasch durchgeführt (s. S.28). Zur Abgrenzung müsste also eine recht lange willentliche Lidschlagdauer für die Selektionsauslösung genutzt werden. Auch diese Interaktionstechnik wäre recht zeitaufwändig.

Zudem ergibt sich bei einem längeren Lidschluss zur Selektionsauslösung das Problem, dass der Benutzer währenddessen die visuelle Informationsaufnahme unterbricht und nicht unmittelbar visuell überprüfen kann, ob die Selektion wie gewünscht erfolgt ist. Lidschlag wird daher traditionell eher für motorisch eingeschränkte Benutzer genutzt [Ras99]. Eine neuere Arbeit wählt den Ansatz, zur Selektionsauslösung nur ein Auge zu schließen und untersucht dies bei einer Drag & Drop-Aufgabe an einer Objektgröße von 5° [Ram21].

Lidschlag wurde mehrmals zur Selektionsauslösung in Computerspielen betrachtet, für Schach [Špa05a] bzw. für Puzzles und Memorys [Wil08, Dja15]; Benutzergruppe waren motorisch eingeschränkte Personen. Špakov [Špa05a] berichtet für ein Schachbrett mit Feldern von 64 Pixeln Seitenlänge (mit Tobii 1750 entspricht dies bei 60 cm Augenabstand vom Monitor 1,6°), dass die Benutzer Blick+Verweildauer (mit 1,8 s Verweildauer) gegenüber Blick+Lidschlag (Lidschlagdauer 350-1.000 ms) bevorzugten. Pro Zug mussten zwei Selektionen durchgeführt werden: die erste selektierte die Figur, die zweite das Feld, auf das die Figur bewegt werden sollte.

Die Verwendung einer ***separaten Softwaretaste*** bedeutet zusätzliche Blickbewegungen, was ebenfalls in einer zeitaufwändigen Interaktionstechnik resultiert [War86] und zudem verlangt, die visuelle Aufmerksamkeit vom Selektionsort weg zu verlagern.

Istance u. a. [Ist09] erprobten Verweildauer auf Softwaretasten zur Steuerung des 3D-Computerspiels „World of Warcraft“ und berichten, dass Verweildauer die Mauseingabe nicht mit gleicher Leistung ersetzen konnte. Castellina u. a. [Cas08] betrachten ein 3D-Computerspiel, in dem Aktionen eines Avatars ebenfalls mit Verweildauer auf Softwaretasten ausgelöst wurden. In beiden Untersuchungen erfolgten die Selektionen an unbewegten Zielobjekten, Nutzergruppe waren motorisch eingeschränkte Benutzer.

Einige Autoren votieren generell gegen unimodale Blickinteraktion. Denn während der Part des Zeigens keine bewusste Anstrengung erfordert [Sta95] – zumindest, wenn der Benutzer sich an die Blickinteraktion gewöhnt hat und gelernt hat, zu vertrauen, dass sie funktioniert – ist dies beim Auslösepart anders. Sowohl bei der Verweildauer als auch beim Lidschlag ist es so, dass

das Wissen darum, dass sie eine bestimmte Zeitspanne dauern müssen, in unnatürlicher Blicknutzung resultiert. Unnatürliches Verhalten wiederum erzeugt zusätzliche Belastung, was nicht wünschenswert ist [Zha99].

2.4.2.2 Multimodale Realisierungen

In der Fachliteratur vorgeschlagene *multimodale* Realisierungen für die Selektionsauslösung sind:

- Hardware-Taste: Der Benutzer drückt eine Taste;
- Sprache: Der Benutzer spricht ein Kommandowort;
- Mimik: Gesichtsmuskelaktivität wie Stirnrunzeln, Lächeln;
- Computermaus: Der Blick wird für das grobe Zeigen verwendet, die Computermaus für das feine Zeigen und die Selektionsauslösung.

Die ***Hardware-Taste*** kommt in verschiedenen Ausprägungen vor:

- Manueller Tastendruck auf einer Computertastatur, Computermaus oder einem Gamepad;
- Tastendruck mit dem Fuß auf einer Fußtaste.

Manueller Tastendruck auf einer Computertastatur, im Folgenden kurz ***Blick+Taste***, ist für Desktopsysteme sehr naheliegend, da diese typischerweise über eine Computertastatur verfügen. Blick+Taste ist also ohne weiteren technischen und mit geringem softwareseitigem Aufwand möglich.

Blick+Taste wurde sehr häufig für die Selektion statischer Objekte als Alternative zur Mauseingabe untersucht [War86, Jac91, Zha07, Kum07b, Ver08]. Verwendete Tasten waren ENTER oder SPACE.

Ware u. a. [War86] untersuchten Blick+Taste[1] für Objektgrößen von 2,0° x 1,65° sowie für Objekte mit den Seitenlängen 3,0°, 2,25°, 1,5°, 0.75° und 0.45°.

[1] Die Taste wurde als „hardware button“ bezeichnet und nicht weiter spezifiziert, außer dass der Benutzer ihn an einen angenehmen Ort („convenient location“) platzieren durfte.

Die Selektionsfehlerquoten lagen für 2,25° und 3,0° bei ca. 5%, für 2,0° x 1,65° bei 8,5%, für 1,5° bei ca. 10%. Für die Größe 0,75° betrug sie etwa 25%, für 0,45° etwa 45%. Die Selektionszeiten lagen für die Größen 1,5° und größer zwischen 600 und 750 ms, für 0,75° bei ca. 1 Sekunde und für 0,45° bei ca. 1,4 Sekunden.

Die Ergebnisse zeigen, dass für Zielobjektgrößen bis 1,5° Blick+Taste eine sehr schnelle Selektionstechnik ist. Bei Zielobjektgrößen kleiner 1,5° ist sie weder effektiv noch effizient, denn hier steigen sowohl die Selektionsfehlerquote als auch die Selektionszeit stark an. Das Experiment wurde mit Kopfstabilisierung durchgeführt.

Ware u. a. führen die Fehlerquote zumindest teilweise auf mangelnde Synchronisation zwischen Blickausrichtung und Tastendruck zurück. Sie mutmaßen, dass Training helfen könnte, die Synchronisation einzuüben und danach noch schneller und fehlerärmer selektieren zu können. Auch Kumar u. a. [Kum08] adressieren die Problematik einer unzureichenden Hand-Auge-Koordination und führen zwei Varianten von schlechter Synchronisation an. Beim zu frühen Tastendruck („early press“) ist der Blick noch nicht sorgfältig auf das Zielobjekt ausgerichtet, beim zu späten Tastendruck schweift der Blick bereits auf einen anderen Ort ab. Bader u. a. [Bad11] (vgl. a. [Bad14]) untersuchten in diesem Zusammenhang den Einfluss des mentalen Modells des Benutzers bei Objektmanipulation auf einem Desktop-Computer (Tobii 1750). Die Probanden mussten mit multimodaler Interaktion – mit Computertaste (gedrückt/nicht gedrückt) zusammen mit einem Stifttablett (Stiftbewegung auf dem Tablett horizontal nach links bzw. rechts) – geometrische Objekte auf dem Desktop-Computer von einer Startposition an eine Zielposition verschieben. Dabei wurde das Blickverhalten aufgezeichnet. Die Autoren fanden im Zusammenhang mit der aktuellen Position des manipulierten Objekts sowohl reaktive objektbezogene als auch proaktive antizipatorische Blickpositionen.

Zhang u. a. [Zha07] untersuchten Blick+Taste (SPACE) für Zielobjekte mit Durchmesser 1,88° (75 Pixel) und 2,5° (100 Pixel) an einem Tipp-Test mit mehreren Richtungen (engl. *Multi-directional Fitts' law task* oder *2D Fitts discrete task*), ähnlich wie von der Norm DIN CEN ISO/TS 9241-411 vorgeschlagen

(vgl. [DIN14], Anhang B.6.2.2). Die Distanzen zwischen Startpunkt und Zielpunkt betrugen 6,88° oder 8,75° (entsprechend 275 oder 350 Pixeln bei 60 cm Augenabstand vom Bildschirm).

Für jede Einzelselektionsaufgabe musste ein Punktepaar selektiert werden (auch als Startobjekt und Zielobjekt bezeichnet). Die Selektionszeit wird berechnet als Zeitdifferenz zwischen den beiden Selektionszeitpunkten, die Selektionsfehlerquote anhand der prozentual erfolgreichen Selektionen des Zielobjekts. Blick+Taste erzielte hier eine mittlere Selektionszeit von ca. 600 ms, die Mauseingabe benötigte 830 ms. Die Selektionsfehlerquote betrug im Mittel mit Blick+Taste 17%, mit Mauseingabe 1%.

Zhang u. a. legten den Probanden auch den Fragebogen zur Einzelbewertung von Interaktionstechniken der DIN CEN ISO/TS 9241-411 vor (vgl. [DIN14], Anhang C.2) und ergänzten ihn (mutmaßlich als Erste) um das Merkmal Augenermüdung. Auf einer 7-Punkte-Skala (1: schlechteste Bewertung, 7: beste Bewertung) wurde die Augenermüdung für die Eyetrackernutzung mit einem Mittelwert ± 1 Standardabweichung von 3,6 ± 1,6 am schlechtesten unter allen Merkmalen bewertet. Das Experiment wurde mit Kopfstabilisierung durchgeführt.

Vertegaal [Ver08] untersuchte ***Blick+linke Maustaste*** für Zielobjekte mit Seitenlänge 4° (140 Pixel), 3° (100 Pixel) und 2° (70 Pixel) an einem Tipp-Test mit einer Richtung (vgl. [DIN14], Anhang B.6.2.1). Die Distanzen zwischen Startpunkt und Zielpunkt betrugen 6°, 12° oder 24° (entsprechend 200, 400 oder 800 Pixeln bei 60 cm Abstand der Augen zum Bildschirm). Auch hier wird die Selektionszeit berechnet als Zeitdifferenz zwischen den beiden Selektionszeitpunkten und die Selektionsfehlerquote anhand der prozentual erfolgreichen Selektionen des Zielobjekts.

Die Ergebnisse zeigten hier eine mittlere Selektionsfehlerquote ±1 Standardabweichung mit Blick+linke Maustaste von 11,7 ± 3,5 % und eine mittlere Selektionszeit von 570 ± 40 ms. Die Ergebnisse der ebenfalls evaluierten Mauseingabe lagen bei 4,6 ± 1,3 % bzw. bei 660 ± 30 ms. Das Experiment wurde ohne Kopfstabilisierung mit dem Gerät Rev II Eyegaze-Eyetracker von LC Technologies durchgeführt.

Monden u. a. [Mon05] untersuchte Blick+linke Maustaste sowie *Blick+linke Maustaste mit Selektion des nächstliegenden Objekts* (im Paper *SemiAuto* genannt) für die Selektion von 9 statischen Zielobjekten mit Seitenlänge 1 cm in einer Anordnung von 3 x 3. Der Abstand zwischen den Zielobjekten betrug 0, 1, 3 oder 5 cm. Der Abstand der Augen vom Bildschirm betrug 50 cm. Das Experiment nutzte den EMR-NC-Eyetracker ohne Kopfstabilisierung.

Es zeigte sich, dass reine Blick+linke Maustaste ab einem Zwischenobjektabstand von 3 cm mit 700 ms die kürzeste Selektionszeit und eine geringe Selektionsfehlerquote von unter 6% erzielte. Die Technik SemiAuto erzielte für alle Zwischenobjektabstände eine Selektionsfehlerquote von unter 5%. Die Selektionszeit betrug zwischen ca. 900 ms (Abstand 0 cm) und 750 ms (Abstand 5 cm).

Bednarik u. a. [Bed09] betrachten als Computerspiel ein 8-teiliges Puzzle (Puzzleteile mit 5,1° Seitenlänge); Selektion eines Puzzleteils bewirkte, dass es sich auf den freien Platz bewegte. Im Vergleich zu reiner Mauseingabe und Verweildauer (1.000 ms) zeigte sich Blick+linke Maustaste besser für die Problemlösungsstrategie und mit weniger Fehlern, mehr Immersion sowie besser bewerteter User Experience.

Djamasbi u. a. [Dja15] betrachten als Computerspiele Memory und Puzzle. Blick+linke Maustaste wurde von den Versuchspersonen gegenüber Blick+Verweildauer bzw. Blick+Lidschlag bevorzugt und besser bewertet bzgl. wahrgenommener Kontrolle über das System, Natürlichkeit der Interaktion sowie Frustration.

Eine ***Fußtaste*** zur Selektionsauslösung erfordert zusätzliche Hardware. Der Aufwand hierfür ist jedoch überschaubar, denn Fußtasten sind heute günstig zu bekommen und einfach über USB anschließbar. Die Nutzung einer Fußtaste wurde im Vergleich zur Handtaste eher selten betrachtet. Die Übersicht von Velloso u. a. [Vel15b] über die Nutzung von Fußeingabe in der MMI führt nur einen einzigen Beitrag für die Nutzung von Blick+Fußtaste auf: Göbel

u. a. [Göb13] untersuchten Pan und Zoom an einer Desktopanwendung. Afkari u. a. [Afk14, Afk18] untersuchen Blick+Fußtaste im medizinischen Umfeld der Mikrochirurgie, Rajanna u. a. [Raj16] für Interaktionsaufgaben mit Selektionen in Windows 10.

Zuletzt wurde eine Arbeit veröffentlicht, die die Selektion von Objekten auf Desktopmonitoren ohne Kopfstabilisierung betrachtet [Zha21]. Für Objektgrößen W von 80, 120 und 160 Pixeln und Distanzen A von 500, 750 und 1000 Pixeln (24"-Bildschirm mit Auflösung 1920 x 1200 Pixel, Augenabstand vom Monitor 70 cm) berichten die Autoren eine mittlere Selektionszeit von 1265 ms, 1001 ms bzw. 932 ms mit zunehmender Objektgröße und von 1022 ms, 1046 ms bzw. 1131 ms mit zunehmender Distanz. Die Selektionsfehlerquoten liegen für $W = 160\ Pixel$ unter 3%, für $W = 120\ Pixel$ zwischen 5 und 8%, für $W = 80\ Pixel$ zwischen 11 und 16%.

Sprache zur Selektionsauslösung ist eine interessante Alternative, wenn der Benutzer überhaupt keine manuelle Aktion durchführen soll. Nachteilig ist, dass Hard- und Software zur Spracheingabe erforderlich sind, deren korrektes Funktionieren eine weitere Fehlerquelle darstellen. Die Selektionszeiten liegen jedoch deutlich über denen mit Blick+Verweildauer. Wilcox u. a. [Wil08] betrachteten Blick+Sprache für die Nutzergruppe motorisch eingeschränkter Benutzer für ein 3D-Puzzle-Computerspiel.

Miniotas u. a. [Min06] betrachten die Selektion kleiner Bildschirmelemente (5 x 5-Anordnung von Quadraten mit 20 bis 40 Pixeln Seitenlänge, Zwischenräume 10 bzw. 20 Pixel). Sie berichten für Blick+Sprache im Vergleich zu Blick+Verweildauer substanziell bessere Selektionsfehlerquoten, die Selektionszeiten sind jedoch erheblich länger. Van der Kamp u. a. [Van11] schlagen Blick+Sprache für eine Zeichenanwendung für motorisch beeinträchtigte Benutzer vor.

Parisay u. a. [Par21] betrachten ein akustisches Geräusch anstelle von Sprache, der Hintergrund sind auch hier Anwender mit motorischen Einschränkungen. Sie nutzten den kostengünstigen Tobii 4C-Eyetracker und ein Headset-Mikrofon (Logitec H370) für die akustischen Eingaben. Für eine Selektionsaufgabe mit Objekten einer Größe von ca. 3° x 2° (Breite x

Höhe) berichten sie gute Nutzbarkeit bei Präsenz von Umgebungsgeräuschen mit kürzerer Selektionszeit und geringerer kogntiver Belastung (erhoben mithilfe des NASA-TLX) als Sprache. Die Versuchspersonen erzielten jedoch mit den ebenfalls evaluierten Interaktionstechniken Mauseingabe und Blick+Verweildauer (500 ms) erheblich kürzere Selektionszeiten und niedrigere Selektionsfehlerquoten.

Gesichtsmuskelaktivität wie Stirnrunzeln benötigt eine separate Messaparatur zur Erfassung und ist auf die Dauer ermüdend [Par01, Sur04]. Tuisku u. a. [Tui16] untersuchten Point-Click-Operationen für Stirnrunzeln, Hochziehen der Augenbrauen und Lächeln für Zielobjekte mit Größe 1,79°, 2,15° sowie 2,86°.

Die erzielten Selektionsfehlerquoten waren am besten für Lächeln und lagen im Mittel ± 1 Standardabweichung zwischen 7,5 ± 1,3 % (bei 2,86° Größe) und 25,3 ± 3,3 % (bei 1,79° Größe). Für die beiden anderen Selektionsauslösemethoden lagen sie noch höher. Die Selektionszeit war etwa 100 bis 200 ms länger als die ebenfalls getestete Verweildauer-Realisierung mit 400 ms. Blick+Gesichtsmuskelaktivität ist also keine konkurrenzfähige Alternative zu Blick+Verweildauer oder Blick+ Taste und wird daher eher für die Benutzergruppe motorisch eingeschränkter Menschen genutzt.

Die Interaktionstechnik ***MAGIC pointing*** kombiniert ebenfalls einen Eyetracker und manuelle Eingabe. Der Blick wird aber nur zum groben Zeigen verwendet und ein TrackPoint als manuelles Eingabegerät zum feinen Zeigen und zur Selektionsauslösung [Zha99]; eine Computermaus wäre auch zur Nutzung möglich gewesen.

Zhai u. a. [Zha99] schlagen diese Kombination aus verschiedenen Gründen vor. Zum einen ist die räumliche Genauigkeit von Eyetracking grundsätzlich limitiert (s.o. Abschnitt 2.2), ein manuelles, indirektes Eingabegerät wie TrackPoint oder Computermaus kann hingegen Pixel-genau selektieren.

Zum anderen sei es unnatürlich, einen Wahrnehmungskanal wie den Sehsinn mit einer motorischen Steuerungsaufgabe zu überladen. Denn dies stehe grundsätzlich im Widerspruch zum natürlichen mentalen Modell des Benutzers bei einer Selektionsoperation, bei dem die Augen Information suchen und

aufnehmen und die Hand die Ausgabe produziert und die externen Objekte manipuliert.

Die Autoren nehmen zudem an, dass der Benutzer seinen Blick weniger bewusst und somit entspannter einsetzen kann, wenn der Blick nur zum groben Zeigen dient und nicht die finale Selektionsposition definiert.

MAGIC pointing nutzt die Blickposition, um einen Großteil der manuellen Cursorbewegung obsolet zu machen, indem der Cursor an die vom Eyetracker gelieferte Blickposition platziert wird. Ausgehend von der Annahme, dass der Benutzer vor einer Selektion ein Zielobjekt anblickt, wird Eyetracking dazu genutzt, die Position des Cursors dynamisch anzupassen, sodass er sich immer in der Nähe des Zielobjekts befindet.

Die Autoren schlagen zwei MAGIC pointing-Varianten vor, *MAGIC pointing konservativ* und *MAGIC pointing liberal*, im Folgenden kurz *MAGIC-kons* bzw. *MAGIC-lib*.

MAGIC-lib versetzt den Cursor zu jedem Objekt, das der Benutzer anblickt. Eine blickbasierte Versetzung findet nur dann statt, wenn die neue Blickposition eine ausreichend große Distanz zur aktuellen Cursorposition hat (z. B. 120 Pixel). Dies soll verhindern, dass der Cursor jede Mikrobewegung des Auges mitmacht und den Benutzer dadurch irritiert (vgl. Abschnitt 2.4.2.3). Will der Benutzer selektieren, so übernimmt er die Cursorsteuerung manuell. Will der Benutzer nicht selektieren, ignoriert er den Cursor und sucht das nächste Objekt.

Die Autoren beschreiben MAGIC-lib als proaktiv, da der Cursor in der Nähe jedes potenziellen Zielobjekts „wartet“. Falls der Benutzer selektieren will, ist dies nützlich. Falls der Benutzer ein Ziel nur anschauen will, mag ihm der Cursor überaktiv erscheinen; nach und nach könnte sich der Benutzer aber angewöhnen, das Cursorverhalten zu ignorieren.

MAGIC-kons versetzt den Cursor erst dann an die aktuelle Blickposition, wenn der Benutzer das manuelle Eingabegerät aktiviert, d.h. bewegt hat. Der Benutzer lenkt den Cursor dann manuell auf das Zielobjekt und selektiert es. Diese Variante entspricht eher der Arbeitsweise mit Cursorn auf GUIs: Der

Cursor ruht an irgendeiner Stelle auf dem Bildschirm und wird nur versetzt, wenn der Benutzer auch selektieren will.

MAGIC-kons ist niemals überaktiv und springt dadurch nicht an Orte, wo der Benutzer nicht selektieren will. Die Autoren mutmaßen jedoch, dass hier mehr Hand-Auge-Koordination eingeübt werden muss als bei MAGIC-lib. Denn jetzt besteht Unsicherheit über den exakten Ort, an dem der Cursor auftauchen wird. Dies könnte den Benutzer zu einer umständlichen Strategie zwingen: (1) Bewegung des manuellen Eingabegeräts, um den Cursor erscheinen zu lassen (2) Warten auf das Erscheinen des Cursors (3) Manuelles Bewegen des Cursors auf das Zielobjekt.

Als potenzielle Vorteile von MAGIC pointing nennen die Autoren

1. Reduzierung manueller Belastung und Ermüdung;
2. Praktikablerer Akkuratheitslevel durch das manuelle, feine Zeigen;
3. Natürlicheres mentales Modell, da sich der Benutzer der Rolle seines Blicks als Eingabegerät nicht bewusst sein muss und Zeigen in letzter Instanz eine manuelle Aufgabe bleibt;
4. Verkürzung der Selektionszeit (als logische Folge aus dem Modell von Fitts, da die Selektionszeit von der Distanz zum Zielobjekt abhängt, vgl. Formel (2.3) S. 50)[1];
5. Verbesserte subjektiv empfundene Geschwindigkeit und Bedienkomfort[2].

Als potenzielle Nachteile von MAGIC pointing gelten neben den Eigenheiten verfügbarer Eyetracker (Latenz, Fehlfunktionen, Unbequemlichkeit) die potenzielle Überaktivität des Cursors bei MAGIC-lib und die Unsicherheit über den exakten Erscheinungsort des Cursors bei MAGIC-kons. Außerdem ist die

[1] Es ist umstritten, ob Blickinteraktion dem Modell von Fitts folgt, da die blickbasierte Selektionszeit nur geringfügig mit der Distanz wächst, siehe z. B. [Sib00].

[2] „Da die manuelle Zeigeamplitude kleiner ist, mag der Benutzer das MAGIC-System subjektiv als schneller und angenehmer empfinden als rein manuelle Steuerung, auch dann, wenn es gleich schnell oder langsamer funktioniert." Eigene Übersetzung aus [Zha99].

Verkürzung der Selektionszeit nicht gesichert: Denn „…mit reinem manuellem Zeigen kann der Benutzer, wenn er die aktuelle Cursorposition kennt, seine motorischen Aktionen möglicherweise parallel zur visuellen Suche durchführen. Die motorische Aktion mag also starten, sobald der Blick auf dem Ziel ruht. Mit MAGIC pointing kann die motorische Aktion erst bei Erscheinen des Cursors starten, was möglicherweise die durch die Reduzierung der Bewegungsamplitude gewonnene Zeit zunichtemacht“ (eigene Übersetzung aus [Zha99]).

Im Experiment untersucht wurden Zielobjekte mit Durchmesser von 0,53° und 1,61° an einem Tipptest (Selektion von Punktepaaren) in drei Richtungen horizontal, vertikal und diagonal. Die Distanzen zwischen Startpunkt und Zielpunkt betrugen 5,37°, 13,37° oder 21,24°. Die Selektionszeit berechnet sich als Zeitdifferenz zwischen den beiden Selektionszeitpunkten und die Selektionsfehlerquote anhand der prozentual erfolgreichen Selektionen des Zielobjekts.

Als Eyetracker wurde ein eigenes System (30 Hz, Kosten 2.000 US $) realisiert. Das Blicksignal wurde mit dem Algorithmus von Jacob [Jac93] gefiltert. Kopfstabilisierung wurde für die Brillenträger genutzt (wie viele der 9 Probanden das betraf, wird nicht berichtet).

Die Ergebnisse zeigten eine mittlere Selektionsfehlerquote mit MAGIC-lib von 7,5%, MAGIC-kons von 7% und ohne Blickunterstützung von 8,2%. Die mittlere Selektionszeit betrug 1,33 s für MAGIC-lib, 1,52 s für MAGIC kons. und 1,4 s ohne Blickunterstützung. Für die Objektgröße von 0,53° betrug sie 1,5 s (MAGIC-lib), 1,7 s (MAGIC-kons) und 1,65 s ohne Blickunterstützung. Für die Objektgröße von 1,61° betrug sie 1,15 s (MAGIC-lib), 1,35 s (MAGIC-kons) und 1,25 s ohne Blickunterstützung.

Die subjektive Befragung bestätigte die Vor- bzw. Nachteile der beiden MAGIC pointing-Varianten. Einige Probanden äußerten, dass die konservative Variante mehr Anstrengung für die Koordination des Timings der Auge-Hand-Kooperation verlangte. Andere fanden sie weniger ablenkend und störend als die liberale. Allerdings machte die Unsicherheit, wo der Cursor erscheinen würde, die Gewöhnung schwieriger.

Insgesamt wurde die liberale Variante aufgrund ihrer schnelleren Reaktionsfähigkeit bevorzugt. Die Autoren bemerken jedoch, dass bei einer realistischeren Anwendungsaufgabe möglicherweise der proaktive Cursor mehr stören könnte, sodass dort eher die konservative Variante bevorzugt werden könnte. Außerdem bemerken sie die Notwendigkeit für weitere Validierung der Ergebnisse sowie für die Gestaltung weiterer MAGIC pointing-Varianten.

2.4.2.3 Visuelles Feedback der Blickposition

Der Benutzer ist von der Benutzung der Computermaus und der Computertastatur daran gewöhnt, während des Interaktionsvorgangs Feedback in verschiedenen Formen zu bekommen. Er fühlt die Tasten, hört ihr Klicken und sieht die Veränderungen auf dem Bildschirm, wenn der Cursor seine Position verändert oder Buchstaben erscheinen.

Die aktuelle Blickposition permanent zu visualisieren, gilt für Benutzer ohne motorische Einschränkungen jedoch als nicht hilfreich. Die permanente Visualisierung belastet die visuelle Wahrnehmung zusätzlich und lenkt ab. Dies gilt besonders dann, wenn die Kalibrierung nicht akkurat ist. Der Benutzer ist irritiert und beginnt möglicherweise, mit den Augen hinter dem visualisierten Blickpunkt nachzujagen, anstatt die Interaktionsaufgabe zu erledigen [Jac95].

Das frühe LC Eyegaze System, gedacht für Benutzer mit motorischen Einschränkungen, nutzte einen kleinen roten Punkt, der die Blickposition anzeigte. Diese Benutzergruppe war aber nur schwer in der Lage, ihre Fixationen zu kontrollieren. Der rote Punkt half daher, die Fixation am gewünschten Ort zu stabilisieren: Beim Eyetyping beispielsweise „zogen“ die Benutzer quasi mit dem Blick den roten Punkt auf die gewünschte Softwaretaste [Cle94].

Statt permanenter visueller Rückmeldung der Blickposition ist es besser, Bildschirmelemente visuell hervorzuheben, sobald der Blick auf sie gerichtet ist. Möglichkeiten sind eine Veränderung der Farbe, Veränderungen (z. B. Vergrößerung) an visuellen Teilelementen wie dem Mittelpunkt oder dem Rand oder die Anzeige eines Rahmens um das Element herum (vgl. z. B. Mausinteraktion bei Windows 10, wenn der Mauszeiger auf einem Desktop-Icon platziert ist).

Bei Blickinteraktion mittels Verweildauer wird Feedback auch eingesetzt, um dem Benutzer zu signalisieren, wann die Verweildauer abgelaufen ist [Hut89, Lan00]; die Autoren schlagen vor, das visuelle Feedback konfigurierbar anzubieten.

Bei der Selektionsauslösung über Tasten kann visuelles Feedback als Bestätigung des Tastendrucks realisiert werden, etwa, indem das selektierte Objekt markiert wird.

2.4.3 Blickbasierte Selektion bewegter Objekte

Wie bei der Mauseingabe so existieren auch für die blickbasierte Selektion erheblich mehr verwandte Arbeiten zur Selektion unbewegter Objekte als zur Selektion bewegter Objekte. Blickbasierte Interaktion mit Anwendungen mit dynamischen Szenen wurde vor allem für die Anwendungsdomäne der Computerspiele betrachtet, die traditionell offen ist für neuartige Eingabegeräte. Übersichten für blickbasierte Interaktion bei Computerspielen liefern Isokoski u. a. [Iso09], Velloso u. a. [Vel16] und Ramirez Gomez u. a. [Ram19].

Die Entwickung von Blickinteraktion für Computerspiele nahm Fahrt auf, nachdem Tobii ab 2014 einen kostengünstigen Gaming-Eyetracker auf den Markt brachte (s.o. S.46), für den inzwischen mit einer Vielzahl nutzbarer Spiele geworben wird[1]. Velloso u. a. [Vel16] nennen als Meilenstein das Jahr 2015, als mit Assessin's Creed Rogue das erste AAA-Game[2] mit Blickinteraktion warb. Ramirez Gomez u. a. [Ram19] nennen die Anzahl von mehr als 140 kommerziellen Computerspielen, bei denen die Nutzung von Eyetracking möglich sei.

Velloso u. a. [Vel16] nennen fünf Kategorien der Blicknutzung in Computerspielen: Navigation, Zielen und Schießen, Auswahl und Kommandos, implizite Interaktion sowie visuelle Effekte. Implizite Interaktion umfasst dabei die

[1] gaming.tobii.com

[2] Einstufung der Computerspielindustrie: AAA sind Computerspiele mit dem höchsten Budget (Entwicklung, Werbung etc.), vgl. https://de.wikipedia.org/wiki/AAA_(Computerspiele)

soziale Interaktion mit einem Avatar [Vid15], die Steuerung der Kameraausrichtung [Smi06, Nac10] sowie Rendering der Szene [Hil08]. Am häufigsten findet sich blickbasierte Interaktion für die Kategorien „Navigation“ und „Zielen und Schießen“.

Die blickbasierte Selektion bewegter Objekte fällt in die Kategorie Zielen und Schießen.

Ramirez Gomez u. a. [Ram19] implementierten drei kleine Computerspiele (Unity Game Engine 2D, 27“-Monitor mit Auflösung 1.920 x 1.080, Tobii EyeX) mit dem Ziel, den Gestaltungsspielraum blickbasierter Computerspiele auszuloten. Eines davon („Witch Hunt“) umfasst neben der Bewegung des Avatars (über Pfeiltasten) auch die Selektion bewegter Objekte mit *Blick+Taste (SPACE)*. Die Autoren machen keine Angaben zu Objektgrößen und -geschwindigkeiten; aus ihrer Figure 1 lässt sich grob ableiten, dass die Objekte ca. 100 Pixel Durchmesser haben, was bei 40 cm Augenabstand vom Monitor 4,47° entspricht (Annahme: 27“ hat bei 16:9 Ratio eine Breite von 60 cm[1], 1 Pixel ist dann 1.920/600 mm = 0,3125 mm groß, 100 Pixel entsprechen dann 31,25 mm; Berechnung mit Formel (2.1), S.26).

Leyba u. a. [Ley04] realisierten ein einfaches 3D-Computerspiel, bei dem 25 magentafarbene Bälle auf schwarzem Hintergrund abzuschießen waren. Angaben zu Objektgrößen und -geschwindigkeiten wurde keine gemacht (aus ihrer Abbildung lässt sich aus der Cursorgröße von 10 Pixeln auf einen Balldurchmesser von ca. 50 Pixeln schließen, was mit Augenabstand zum Monitor (Tobii 1750) von 60 cm ca. 1,25° entspricht. Mauseingabe war signifikant besser (Selektionstrefferquote und Selektionszeit) als Blick+linker Mausklick. Die Autoren führen dies teilweise auf schlechte Eyetracker-Kalibrierungen zurück.

Jönsson [Jön05] betrachtet zwei First-Person-Shooter-Computerspiele (FPS), „Halflife“ und „Sacrifice“. Interaktionsaufgaben sind die Navigation in der Szene (Ausrichtung des Sichtfeldes) sowie das Abschießen von Zielobjekten (keine Angabe zu Objektgrößen und -geschwindigkeiten; aus den Abbildungen

[1] https://de.wikipedia.org/wiki/Bildschirmdiagonale

geschätzt, sind die Figuren mehrere Grad Sehwinkel groß). Jönsson vergleicht drei Interaktionsdesigns: (1) Blick zum Zeigen (Zielen), Maus für Sichtfeldausrichtung und Schussauslösung (linker Mausklick) (2) Blick zum Zielen und zur Sichtfeldausrichtung, Maus zur Schussauslösung (3) Mauseingabe für alle Interaktionskomponenten. Es wurde keine quantitative Analyse durchgeführt, aber die Versuchspersonen gaben für die beiden blickbasierten Interaktionsdesigns größere Begeisterung sowie höhere subjektiv empfundene Interaktionsgeschwindigkeit an.

Smith u. a. [Smi06] betrachteten Blick+linke Maustaste im Vergleich zur Mauseingabe beim Computerspiel „Missile Command“, wo wenige Pixel große Flugkörper abgeschossen werden mussten. Dazu mussten nicht die bewegten Flugkörper direkt getroffen werden, sondern es musste durch Selektion einer Stelle in ihrer Flugbahn eine Bombe platziert werden, sodass die Explosion den Flugkörper zerstören würde. Mit reiner Mauseingabe erzielten die Versuchspersonen eine um 25% höhere Trefferzahl.

Isokoski u. a. [Iso06] implementierten ein FPS-artiges Computerspiel, präsentiert auf dem Monitor eines Tobii 1750. In einer Naturumgebung erscheinen Ziele, die abzuschießen sind. Wann immer ein Zielobjekt getroffen wurde, erschien ein neues an einem zufälligen Ort in der Szene. Es wurden keine Angaben zu Objektgröße und Geschwindigkeiten gemacht (aus ihrer Figure 1 geschätzt, sind die Zielobjekte mehrere Grad Sehwinkel im Durchmesser).

Die Spielerbewegung in der Szene erfolgte mit den Pfeiltasten einer Computertastatur, die Ausrichtung des Sichtfeldes erfolgte mit Mauseingabe. Das Abschießen erfolgt entweder mit Mauseingabe (linker Mausklick) oder mit Blick+rechtem Mausklick (Blickposition als kleiner roter Punkt visualisiert). Die dritte Alternative war Interaktion mit einem Gamepad-Kontroller mit den typischen Belegungen (Joysticks für Spielerbewegung bzw. Sichtfeldausrichtung, Zeigefinger-Schusstasten). Die Trefferzahl war mit Maus+Tastatur am besten; Maus+Tastatur+Blick erzielte dasselbe Ergebnis wie das Gamepad.

Isokoski u. a. [Iso07] nutzten dasselbe FPS-artige Computerspiel wie Isokoski u. a. [Iso06]. Verglichen wurden jetzt die Interaktionsdesigns (1) Gamepad der Xbox 360 (Sichtfeldausrichtung, Spielerbewegung, Zielen+Schießen)

(2) Gamepad (Sichtfeldausrichtung, Spielerbewegung, Schießen) + Blick (Zielen) (3) Gamepad (Spielerbewegung + Schießen) + Blick (Sichtfeldausrichtung + Zielen). Die Trefferzahl war ähnlich für alle drei Interaktionsdesigns. Die blickbasierten Varianten benötigten aber mehr Schussversuche. Angaben zu den Selektionszeiten wurden keine gemacht.

Isokoski u. a. [Iso09] vergleichen Blickeingabe und Mauseingabe für das klassische 2D-Fun-Shooting-Spiel „Chicken Shoot“[1]. Die Aufgabe ist, Hühner zu treffen, die kreuz und quer durch die Szene fliegen.

Mit Mauseingabe wird das Fadenkreuz zum Zielen mit dem Mauszeiger platziert; ein Schuss wird mit linkem Mausklick ausgelöst; nach 10 Schüssen muss die Waffe wieder geladen werden (rechter Mausklick). Die 2D-Navigation in der Szene nach links bzw. rechts erfolgt mit den Pfeiltasten der Computertastatur.

Mit Blickeingabe wird das Fadenkreuz zum Zielen mit der Blickposition gesetzt; der Schuss erfolgt über eine Softwaretaste (Verweildauer 500 ms), wobei mit einer Rate von 2 Hz alle 10 Schüsse abgefeuert werden (Aufladen automatisch nach 10 Schüssen). Die 2D-Navigation in der Szene erfolgt über Softwaretasten links und rechts der Szene mit Verweildauer 500 ms.

Anfänglich erzielten die Versuchspersonen eine höhere Trefferzahl mit Maus+Tastatur; nach 4 bis 5 Versuchen erzielten die meisten Versuchspersonen höhere Trefferzahlen mit Blickeingabe. Zudem berichteten die Versuchspersonen eine höhere Immersion mit Blickeingabe.

Velloso u. a. [Vel15a] implementierten drei Computerspiele (Unity 3D) für ihre „Arcade+“-Plattform, mit der multimodale Computerspiele im öffentlichen Raum ermöglicht und evaluiert werden sollen. Das Spiel „StarGazing“ erfordert blickbasierte Bewegtobjektselektion. Der Spieler befindet sich auf der Brücke eines Raumschiffs und muss verhindern, dass nahende Asteroiden sein Schiff treffen. Die Umgebung ist dunkel, außer dort, wohin sein Blick gerichtet ist. Der Spieler muss also kontinuierlich den Horizont visuell abtasten und, wenn er einen Asteroiden erblickt, diesen zerstören. Dazu zeigt er mit dem

[1] https://www.chickenshoot.com/ger/index.html

Blick auf den Asteroiden und drückt zur Schussauslösung einen Hardware-Knopf des Arcade+-Systems. Angaben zu Größen und Geschwindigkeiten der Asteroiden wurden keine gemacht; aus Figure 3 in [Vel15a] lässt sich grob abschätzen, dass die Objektgrößen vmtl. wenige Grad Sehwinkel umfassen.

2.4.3.1 Selektion auf Basis von Folgebewegungen – Pursuits

Vidal u. a. [Vid13] führten eine neue Art der blickbasierten Interaktion speziell für die Interaktion mit bewegten Objekten ein, die sogenannten *Pursuits*. Die Bezeichnung ist angelehnt an die englische Bezeichnung für Folgebewegungen der Augen, engl. *Smooth Pursuits* (vgl. S. 31).

Pursuits als Interaktionstechnik stützen sich darauf, dass bei sich (langsam) bewegenden Objekten die Trajektorie, die das Objekt vollzieht, von den Augenfolgebewegungen nachvollzogen wird. Die Idee ist nun, beide Trajektorien zu korrelieren (Pearsons Korrelationskoeffizienz, s.u.), um zu detektieren, welches Objekt der Beobachter betrachtet.

Im Gegensatz zu direkter Blickinteraktion ist die Pursuits-Technik unabhängig von der Blickposition und benötigt daher keine Kalibrierung des Eyetrackers. Da die Korrelationsformel inhärent die Daten normalisiert, müssen sich die Blickkoordinaten nicht im selben Bereich wie die Objektkoordinaten bewegen, sodass Pursuits mit jedem Eyetracker bzw. mit jeder Monitorgröße nutzbar sind. Zudem wird die herausfordernde Selektion bewegter Objekte in natürlicher Weise adressiert. Pursuits sind in gewisser Weise ähnlich zur Selektion über Verweildauer, vermeiden aber die Unnatürlichkeit einer willentlich verlängerten Fixation.

Als Eingabe benötigt *Pursuits* die synchronisierten Zeitreihen der horizontalen und vertikalen Blickkoordinaten und der Koordinaten aller Objekte auf dem Monitor. Für jedes Objekt wird der Pearsonsche Korrelationskoeffizient berechnet (Formel für x-Koordinate, Berechnung gleichermaßen für y):

$$corr_x = \frac{E[(Blick_x - \overline{Blick_x})(Obj_x - \overline{Obj_x})]}{\sigma_{Blick_x}\sigma_{Obj_x}} \tag{2.8}$$

Bei der Echtzeit-Interaktion werden für jeden neuen Datenpunkt $corr_x$ und $corr_y$ für alle Paare aus Blick und Objekt $(Obj(i), Blick), i = 1, \ldots, n$ berechnet. n ist dabei die Anzahl Objekte, die sich während des Zeitfensters auf dem Monitor befunden haben, das die letzten w ms umfasst. Wenn für ein Objekt sowohl $corr_x$ als auch $corr_y$ über einem Schwellenwert S_{corr} liegen, wird das Objekt als eines identifiziert, das vom Blick verfolgt wird und es wird selektiert. Der Algorithmus nutzt also zwei Parameter: die Länge des Zeitfensters w, über dem die Korrelationskoeffizienten bestimmt werden, sowie den Schwellenwert S_{corr} des Korrelationskoeffizienten.

Die Autoren nennen folgende Eigenschaften ihrer Pursuits-Technik.

Erstens ist die Objektselektion nicht mehr von der Objektgröße abhängig, sondern nur von der Objektgeschwindigkeit; auf diese Weise können sehr kleine Objekte selektiert werden, die mit positionsbasierten Selektionstechniken schwierig zu selektieren sind (vgl. a. Formel (2.4), S. 51).

Zweitens ist Pursuits nur für bewegte Objekte anwendbar; denn bei einem unbewegten Objekt können die erforderlichen Standardabweichungen der x- bzw. y-Koordinaten nicht berechnet werden.

Drittens benötigt Pursuits deutlich unterschiedliche Korrelationskoeffizienten, um zwischen Objekttrajektorien unterscheiden zu können; bei zwei linearen Trajektorien muss sich bspw. die Winkeldifferenz ihrer Vektoren unterscheiden.

Viertens müssen sich die Koordinaten von Objekt und Blick in gleicher Weise entwickeln; der Monitor sollte also orthogonal zur Sichtlinie ausgerichtet sein.

Für die Evaluation der Schlüsseleigenschaften von Pursuits bezogen auf die Anzahl, Geschwindigkeit und Trajektorie von Objekten führten Vidal u. a. eine Nutzerstudie durch. Die Versuchspersonen wurden jeweils instruiert, genau ein Objekt mit den Augen zu verfolgen. Andere Objekte waren ebenfalls präsent, wurden jedoch in der Hintergrundfarbe angezeigt, um Ablenkungen zu verhindern.

Variiert wurden die Variablen Anzahl Zielobjekte (2 bis 20), Objektgeschwindigkeiten (100 bis 850 Pixel/s), Trajektorientyp (linear, kreisförmig), S_{corr} (0,2

bis 0,9), w (100 ms, 500 ms). Die Blickdaten wurden mit dem Tobii X300 erfasst (Samplingrate 300 Hz), der unterhalb eines 40“-Monitors (Auflösung 1.920 x 1.080) platziert war.

Die Ergbnisse zeigten, dass mit steigender Anzahl Objekte die Detektionsquote sank. Der Trajektorientyp war ohne großen Einfluss. Bezüglich der Fenstergröße w zeigten sich die Ergebnisse deutlich besser mit 500 ms als mit 100 ms. Allerdings stieg die Anzahl verpasster Objekte ab einer Geschwindigkeit von 650 Pixel/s mit 500 ms deutlich an. Die Autoren empfehlen daher für Systeme, die hohe Reaktivität erfordern, $w = 100\ ms$.

Die Unabhängigkeit von der Objektgröße zeigten die Autoren anhand eines Computerspiels, bei dem ein Frosch Fliegen fangen soll, die um ihn kreisen. Verfolgt der Benutzer eine Folge mit den Augen, fängt der Frosch diese Fliege. Das Spiel startet mit einer Fliege mit einer Geschwindigkeit von 650 Pixel/s; jede verspeiste Fliege wird durch eine mit höherer Geschwindigkeit und geringerer Größe ersetzt. Als Parameter wurden $w = 300\ ms$ und $S_{corr} = 0{,}3$ gewählt.

Die erzielten Selektionszeiten für das Frosch-Spiel (und andere getestete Anwendungen, vgl. [Vid13]) lagen in der Größenordnung von Sekunden. Sie war für das Frosch-Spiel sechsmal so lang wie das Pursuit-Fenster $w = 300\ ms$. Die Selektionsquote lag bei ca. 89%.

Als Nachteile nennen Vidal u. a. [Vid13], dass Pursuits bei längerer Nutzung anstrengend ist, dass langsam bewegte Objekte schwer zu detektieren sind sowie dass Kopfbewegungen die Blicktrajektorie verändern können und Pursuits möglicherweise nicht mehr gut funktioniert.

Velloso u. a. [Vel16] erwähnen in ihrer Übersicht zu Blickeingabe in Computerspielen auch Pursuits. Sie bestätigen einerseits die Ansicht, dass Pursuits für die blickbasierte Bewegtobjektselektion natürlicher erscheinen mögen als die Verweildauer. Sie weisen jedoch auch auf die Nachteile der langen Selektionszeiten sowie ungewollter Aktivierung hin.

Velloso u. a. [Vel17] geben eine Übersicht über Arbeiten, die Bewegungskorrelation (engl. *motion correlation*) zur Systemeingabe nutzen. Sie beschreiben

Anwendungen für unterschiedliche Monitorgrößen (Smart Watch bis große Anzeigen im öffentlichen Raum wie z. B. Schaufenster) mit unterschiedlichen Trajektorien (geometrische Figuren: Gerade, Ellipse, Kreis, Quadrat, Dreieck).

Die Autoren bestätigen die Aussage der langen Selektionszeiten, die darauf zurückzuführen sind, dass das Pursuits-Fenster w ggf. sehr lange (mehrere Sekunden) sein muss, damit sich ähnliche Trajektorien über den Korrelationskoeffizienten unterscheiden lassen. Zudem weisen sie darauf hin, dass Pursuits umso besser funktioniert, je weniger Objekte gleichzeitig sichtbar sind; wenngleich Pursuits auch für mehr als 10 Objekte funktioniert, ergeht die Empfehlung, die Anzahl auf ein Minimum zu beschränken.

Isachenko u. a. [Isa18] betrachten die Frage, ob Blickeingabe schneller sein kann als Mauseingabe, wenn Zielobjekte und Nicht-Zielobjekte sich visuell kaum unterscheiden. Sie untersuchen dies an einer Versuchsaufgabe mit 20 Bällen (Durchmesser 2,8°), die sich linear in unterschiedlichen Richtungen mit 344 Pixel/s (12°/s) auf einem Monitor (18,5“ mit Auflösung 1.440 x 900 Pixel) bewegen; die Geschwindigkeit wurde absichtlich passend zur typischen Geschwindigkeit für Augenfolgebewegungen gewählt. Die Bälle waren durchnumeriert, sodass über die Ansage der Nummer das nächste Zielobjekt bestimmt wurde.

Sie vergleichen drei Selektionstechniken: (1) traditionelle Mauseingabe (2) Mausfolgebewegung und (3) Augenfolgebewegung. Eine Selektion war erfolgreich

- mit traditioneller Mauseingabe, wenn der Mausklick mit einer Distanz < 2,4° vom Ballmittelpunkt erfolgte;
- mit Mausfolgebewegung, wenn der Median der Distanzen des *Mauszeigers* vom Zielobjekt über ein Zeitfenster von 867 ms < 1,7° war und dies zugleich der kleinste Median aller Objekte war;
- mit Augenfolgebewegung (ohne Blickpunktvisualisierung), wenn der Median der Distanzen der *Blickposition* vom Zielobjekt über ein Zeitfenster von 867 ms < 1,7° war und dies zugleich der kleinste Median aller Objekte war.

Die Blickdaten wurden mit einem Tobii 4C-Eyetracker erfasst. 16 Versuchspersonen erzielten die kürzeste Selektionszeit mit der Augenfolgebewegung mit im Mittel ± 1 Standardabweichung 1,13 ± 0,16 s; dies war im Mittel doppelt so schnell wie die traditionelle Mauseingabe und dazu mit geringerer Standardabweichung (Trad. Maus 2,29 ± 0,41 s; Mausfolgebewegung 1,60 ± 0,39 s). Selektionen wurden nur in Einzelfällen verpasst; auch hier war die Augenfolgebewegung geringfügig besser als die Maus-Techniken. Alle Versuchspersonen nannten zudem die Augenfolgebewegung als komfortabelste Eingabetechnik.

Zhao u. a. [Zha20] verglichen an einem ähnlichen Szenario wie Isachenko u. a. [Isa18] die Bewegtobjektselektion für vier Interaktionstechniken, alle basierend auf Augenfolgebewegungen (AF): (1) AF + linke Maustaste (2) AF + Sprachkommando Ballnummer (3) AF + Sprachkommando „ty“[1] (4) AF + Sprachkommando selbstgewähltes Wort. Hintergrund ist die Nutzergruppe motorisch eingeschränkter Benutzer. Ziel der Untersuchung war herauszufinden, wie leistungsfähig ein Sprachkommando im Vergleich zu einer manuellen Selektionsauslösung ist.

Die Versuchsaufgabe enthielt 10 Bälle (Durchmesser 100 Pixel (2,6°); Geschwindigkeit 225 Pixel/s (6°/s)), die ihre Bewegung in natürlicher Weise änderten, wenn sie zusammenprallten oder den Rand des Monitors trafen. Ab dem Zeitpunkt des Kommandos zur Selektionsauslösung (Mausklick bzw. Sprache) wird der Median der Distanzen der Blickposition zum Ballzentrum für alle Bälle bestimmt. Das betrachtete Fenster betrug 867 ms für die Versuchspersonen 1 bis 10 und 400 ms für die Versuchspersonen 11 bis 23. Eine Selektion war erfolgreich, wenn der Median < 60 Pixel betrug und am kleinsten unter allen Bällen war.

Die Blickdaten wurden mit einem Tobii 4C-Eyetracker erfasst. AF + linke Maustaste wurde subjektiv bevorzugt und erzielte mit ca. 92% die beste Trefferquote sowie mit im Mittel ± 1 Standardabweichung 1269 ± 265 ms die kürzeste Selektionszeit (Sprachkommandos mit 60 bis 80% bzw. 1581 ± 305 ms bis 1721 ± 348 ms.).

[1] Russisch für „du“. Die Versuchspersonen waren alle russische Muttersprachler.

2.4.4 Interaktion mit Blick und Brain-Computer-Interfaces (BCI)

Reize, die die menschlichen Sinnesorgane empfangen, werden großteils im Gehirn weiter verarbeitet. Die Verarbeitung geschieht durch ein komplexes Netz von Neuronen. Sobald ein internes oder externes Ereignis auftritt, erzeugt das Gehirn dazu ein ereigniskorreliertes Potenzial (EKP). Externe Ereignisse sind bspw. die Stimuli, die die Sinnesorgane empfangen, interne Ereignisse sind bspw. Entscheidungen, die der Mensch trifft.

Nach der Darbietung eines Reizes tritt also eine Abfolge unterschiedlicher EKP auf. Nach internationaler Vereinbarung werden EKP gemäß ihrer Polarität (im Vergleich zum Ruhepotenzial einer Nervenzelle) mit N (negativ) bzw. P (positiv) bezeichnet. Hinzu kommt entweder die laufende Nummer des negativen bzw. positiven Potenzials oder die Angabe der mittleren Latenz, mit der das EKP nach dem Reizzeitpunkt auftritt. P300 bezeichnet demnach ein EKP mit positivem Potenzial, das 300 ms nach dem Reizzeitpunkt auftritt.

Bei visuellen Reizen werden die EKP-Komponenten bis ca. 100 ms nach Reizzeitpunkt von externen Ereignissen erzeugt. Danach werden diese durch intern generierte Komponenten überlagert (P1 nach 80-120 ms, N1 nach 160-200 ms), wobei P1 und N1 Prozesse räumlicher Aufmerksamkeit repräsentieren [Buc14]. Die P300 (P3b) „(…) reflektiert Gehirnaktivität, die erforderlich ist, um eine Repräsentation im Kurzzeitgedächtnis aufrechtzuerhalten (…)“; die Größe der Amplitude der P300 ist proportional zur „(…) Intensität der Aufmerksamkeit, die einer Aufgabe gewidmet wird.“ Ihre maximale Amplitude erreicht die P300 300-500 ms nach Reizzeitpunkt.

Die Messanordnung zur Ableitung der P300 wird typischerweise mit dem sogenannten „Oddball“-Paradigma abgeleitet. Dabei werden der Versuchsperson zwei Reize dargeboten, bspw. visuell präsentierte Buchstaben X und O. Einer der Buchstaben dient als Standardreiz und wird häufiger dargeboten (in ca. 90% der Fälle [Buc14]), der andere dient als Zielreiz und wird seltener dargeboten (in den verbleibenden 10% der Fälle). Die Versuchsperson hat dann die Aufgabe, „(…) auf den seltenen Reiz mit Tastendruck zu reagieren oder die

Anzahl der in einem Stimulusblock präsentierten seltenen (Ziel-)Reize zu zählen“ [Buc14]. Im Elektroenzephalogramm (kurz EEG) tritt dann bei Präsentation des Zielreizes die P300 auf, bei Präsentation des Standardreizes nicht (vgl. z. B. Buchner u. a. [Buc14, S. 96, Abb. 7.6].

Bechara u. a. [Bec00] zeigten, dass Menschen unbewusst auf auffällige oder relevante Ereignisse bereits reagieren, bevor sie bewusst reagieren oder auch in Fällen, wo keine bewusste Reaktion erfolgt. Diese unbewusste, sehr schnelle Reaktion kann aus EEG-Signalen extrahiert werden, auch wenn danach keine bewusste Reaktion mehr folgt.

Heutzutage ist es möglich, EKP nicht-invasiv zu erfassen. Dafür werden Elektroden am Kopf angebracht, um dort die elektrische Aktivität zu messen, auch Elektroenzephalographie genannt. Ein Gerät hierfür ist beispielsweise der actiCHamp Recorder [Bra21].

Die Hauptnutzung von EEG liegt im diagnostischen Bereich zur Erforschung der Wirkweise des Gehirns in den Neurowissenschaften sowie zur Feststellung medizinischer Befunde. Inzwischen existieren jedoch auch Brain-Computer-Interfaces (BCI), die das Ziel haben, die gemessenen Hirnströme zur Informationseingabe in ein Computersystem zu nutzen [Bro21].

EEG-basierte BCI nutzen zur Detektion von Ereignissen, die die Aufmerksamkeit des Benutzers erregen, oft das sogenannte P300 Speller-Paradigma [Kru08]. Es wird genutzt, damit Personen, die vollständig unfähig sind, motorisch zu interagieren oder zu kommunizieren, buchstabieren können. Dabei werden alle Buchstaben in einer Matrix präsentiert und nacheinander visuell hervorgehoben. Dies löst ein charakteristisches EKP – den P300 – etwa 300 ms nach der Hervorhebung des Ziel-Buchstabens aus. Dieses EKP kann im EEG automatisch detektiert werden, sodass der gewünschte Buchstabe abgeleitet werden kann.

Eine andere Anwendung ist die Assistenz bei der Bildanalyse [She08, Ger06]. Anstatt dass menschliche Bildauswerter manuelle Bilder annotieren, werden ihnen kontinuierlich Bilder präsentiert, wobei ein EEG während des Betrachtens aufgezeichnet wird. Diese EEG-Aufzeichung wird danach automatisch auf neuronale Marker hin analysiert, die sich auf bestimmte Bilder beziehen.

Yong u. a. [Yon11] betrachten die Kombination von EEG und Eyetracking, um Texteingabe mit einem BCI-Speller an beliebigen Stellen auf Basis der Blickposition zu ermöglichen.

Huang u. a. [Hua13] integrieren EEG-Information, um die Steuerung eines Blick-basierten Cursors kontrollierbarer zur machen und berichten deutlich verbesserte Präzision verglichen mit rein blickbasierter Steuerung.

Zander u. a. [Zan10] realisieren einen visuell-neuronalen Objektselektionsprozess, der ganz ohne manuelles Eingreifen funktioniert. Auch hier lokalisiert Eyetracking den räumlichen Ort der Aufmerksamkeit des Benutzers, das EEG-Signal wird anstelle blickbasierter Verweildauer (vgl. S. 63) für die Bestimmung der zeitlichen Komponente der Selektion genutzt. Zander u. a. nutzen diesen Ansatz für eine Such- und Selektionsaufgabe auf einer einfachen GUI.

Um eine Selektion auszulösen, muss der Benutzer aktiv ein mentales Kommando initiieren, wofür eine Handbewegung imaginiert werden musste (Vorstellung, ein Handtuch mit beiden Händen auszuwringen, indem man die Hände gegenläufig verdreht). Die Autoren begründen die Nutzung eines expliziten, aktiv generierten mentalen Kommandos damit, dass ein solches höhere zeitliche Genauigkeit bietet als eine implizite Selektionsauslösung auf Basis der P300. Allerdings ist die Reaktionszeit eines aktiven BCI deutlich länger als implizite Methoden.

Aufgrund des Zusammenhangs der P300 mit selten auftretenden Zielreizen sowie aufgrund der Verfügbarkeit nicht-intrusiver EEG-Messung ist es denkbar, EEG als Mechanismus für die Selektionsauslösung zu nutzen und analog zur Blick+Taste, Blick+Verweildauer etc. (s.o.) eine Selektionstechnik Blick+EEG zu realisieren.

2.4.5 Blickbasierte Aufgaben- bzw. Tätigkeitsklassifikation

Die Aufgabe oder Tätigkeit, die ein menschlicher Beobachter durchführt, aus seinen Blickbewegungen abzuleiten, wurde von etlichen Autoren für unterschiedliche Beobachtungsaufgaben untersucht. Ein großer Teil der Arbeiten stammt aus der kognitiven Psychologie und hat das Ziel herauszufinden, wie die Mechanismen menschlicher Aufmerksamkeit auf neuronaler und auf Verhaltensebene funktionieren.

Neuere Beiträge stammen auch aus anderen Anwendungsdomänen wie der Mensch-Maschine-Interaktion oder Robotik. Henderson u. a. [Hen13] stellen als praktischen Nutzen solcher Tätigkeits- oder Aufgabenklassifikation den Einsatz als passive Eingabequelle für die MMI oder andere Anwendungen, die auf den kognitiven Zustand des Benutzers reagieren, heraus[1].

Die überwiegende Anzahl der Beiträge untersucht das Blickverhalten bei der Betrachtung von Einzelbildern, wenige adressieren die Betrachtung von Bildfolgen oder andere dynamische Szenarien.

Buswell [Bus35] gilt als der erste, der versuchte, aus dem Blickverhalten eines menschlichen Beobachters auf seine Tätigkeit oder Aufgabe zu schließen. Buswell ließ Versuchspersonen zuerst ein Bild von einem Turm betrachten, und zwar frei ohne weitere Instruktion. Als Zweites sollten die Versuchspersonen eine Person in einem Fenster des Turmes finden. Das Ergebnis war, dass sich die Fixationsverteilung stark unterschied: die Instruktion bewirkte, dass sowohl die Anzahl als auch die Dauer der Fixationen deutlich höher bzw. länger waren. In einer anderen Aufgabe sollte eine Szene zunächst wieder frei und darauf nach dem Lesen einer Szenenbeschreibung betrachtet

[1] „From a practical standpoint, successful classification opens the door to using eye movements as a passive input source for human-computer interfaces and other applications that react to the user's cognitive state." [Hen13]. Vgl. a. Boisvert u. a. [Boi16] „One evident benefit from an applied perspective is the capability to infer task or intentions from fixations for applications in human machine interaction and human centric computing."

werden. Nach dem Lesen erhöhte sich die Anzahl Fixationen um 75%. Buswell mutmaßte, dass die Instruktionen die Veränderungen der Fixationscharakteristika bewirken, indem sie das Interesse des Betrachters für bestimmte Bildregionen wecken. Die Unterschiede leuchten ein, da Fixationen anzeigen, wohin der Betrachter seine Aufmerksamkeit richtet [Jus76].

Yarbus [Yar67] ließ einen Betrachter das Gemälde „Unerwartete Besucher“[1] unter sieben verschiedenen Fragestellungen/Instruktionen zum Bildinhalt betrachten (z. B. Wie alt sind die Personen? Wie ist der materielle Status der Familie?) und fand unterschiedliche Fixationsorte je nach Fragestellung. Je nach Fragestellung sind zur Beantwortung der Frage unterschiedliche Bildbereiche relevant und auch die erforderliche Detailtiefe variiert. Demzufolge variieren Anzahl und Länge der auftretenden Fixationen. Die Ergebnisse zeigten, dass das menschliche visuelle System aktiv vorgeht, was auf den großen Einfluss der Top-down-Komponente visueller Informationsverarbeitung hinweist (s.o. Abschnitt 2.1, S. 22, oder Abschnitt 2.1.5).

DeAngelus u. a. [DeA09] wiederholten das Experiment von Yarbus innerhalb eines größeren Experiments mit 57 digitalen Bildern von Gemälden, Fotografien und Zeichnungen mit 17 Versuchspersonen. Das Blickverhalten war auch hier instruktionsabhängig unterschiedlich.

Castelhano u. a. [Cas09] ließen Fotografien komplexer natürlicher Innen- und Außenraumszenen betrachten, jeweils einmal unter der Instruktion, bestimmte Objekte im Foto zu suchen, und einmal, um sich an Szeneninhalte zu erinnern. Die Ergebnisse zeigten je nach Instruktion Unterschiede auf der Ebene aggregierter Blickmaße wie der durchschnittlichen Fixationsdauer und der durchschnittlichen Sakkadenamplitude.

Hild u. a. [Hil12] ließen Videoclips einer PTZ-Überwachungskamera mit fest eingestelltem Szenenausschnitt betrachten. Die Instruktionen definierten vier Tätigkeiten: Erkunden, Suchen, Beobachten und Verfolgen. Die Ergebnisse zeigten Unterschiede bezüglich einzelner aggregierter Blickparameter wie der

[1] s. https://de.wikipedia.org/wiki/Datei:Ilya_Repin_Unexpected_visitors.jpg

durchschnittlichen Fixationsdauer oder der durchschnittlichen Sakkadenamplitude. Die Wertebereiche für die einzelnen Tätigkeiten überlappten jedoch erheblich.

Neuere Arbeiten berichten nicht nur aggregierte Blickmaße, sondern wenden maschinelle Lernverfahren an, um aus dem aufgezeichneten Blickverhalten auf die Beobachteraufgabe zu schließen. Zur Klassifikation werden zwei Arten von Merkmalen genutzt [Bor14]:

- Maße, die die Fixationsdynamik beschreiben, z. B. die Fixationsdauer. Sie spiegeln die Top-down-Komponente der visuellen Aufmerksamkeit.
- Maße, die die Bildeigenschaften beschreiben, z. B. Salienzkarten. Sie spiegeln die Bottom-up-Komponente visueller Fixation wider.

Einige Autoren nutzen eine Kombination aus Merkmalen der Fixationsdynamik und Merkmalen der Bildeigenschaften [Boi16, Bor14, Dor12], einen Überblick geben Boisvert u. a. [Boi16] sowie Borji [Bor21]. Andere nutzen ausschließlich Merkmale der Fixationsdynamik [Hen13, Kan14, Kar15, Gre12].

Henderson u. a. [Hen13] weisen dabei auf einen entscheidenden Vorteil hin, ausschließlich allgemeine Merkmale der Blickbewegungen zu nutzen, nämlich dass es insbesondere in dynamischen Umgebungen herausfordernd sein kann, die Blickpositionen den sich ständig ändernden visuellen Objekten zuzuordnen. Dies wäre erforderlich, wenn man Bildmerkmale nutzen will. Da die vorliegende Arbeit zur blickbasierten Tätigkeitsklassifikation bei der Bildfolgenanalyse eine erste Untersuchung beiträgt, wurden nur Maße für die Blickdynamik (Fixations- und Sakkadenmerkmale) zur Klassifikation herangezogen (vgl. Abschnitt 6.4). Im Folgenden sind daher nur verwandte Arbeiten aufgeführt, die ausschließlich Merkmale der Blickdynamik genutzt haben.

Greene u. a. [Gre12] ließen 16 Versuchspersonen 20 Schwarz-Weiß-Fotografien unter vier Instruktionen betrachten: (a) Bildinhalte erinnern (b) Dekade der Bildaufnahme bestimmen (c) Bestimmen, wie gut sich die Personen im Bild kennen (d) Wohlstand der Personen im Bild bestimmen. Die Klassifikation der Aufgabe/Instruktion erfolgte mit dem Klassifikationsverfahren

LDA (Lineare Diskriminanzanalyse) mit einem Merkmalsvektor mit den vier Blickmerkmalen (1) mittlere Anzahl Fixationen, (2) mittlere Fixationsdauer, (3) mittlere Sakkadenamplitude sowie (4) Prozent des Bildes, die von Fixationen überdeckt wurden. Die Klassifikation gelang mit 25,9% korrekt, (95% KI = 21-31%, p=0,70), also nicht besser als die Zufallswahrscheinlichkeit, die bei vier Instruktionen 25% beträgt.

Borji u. a. [Bor14] berichten zwei Experimente mit Einzelbildbetrachtungen. Im ersten analysieren sie die Daten von Greene u. a. erneut und erzielen mit dem RUSBoost-Algorithmus [Sei09] (Merkmalsvektor wie bei Greene u. a.) eine Korrektklassifikationsrate für die Klassifikation der Instruktion/Aufgabe für Einzelbilder, die signifikant über der Zufallswahrscheinlichkeit von 25% lag. Im zweiten Experiment wurden 15 Gemälde betrachtet, darunter das Bild von Repin aus Yarbus' Untersuchung, sodass seine sieben Instruktionen übernommen werden konnten. Der RUSBoost-Algorithmus (Merkmal: geglättete Fixationskarte) erzielt eine Korrektklassifikationsrate der Instruktion/Aufgabe signifikant über (fast doppelt so hoch) der Zufallswahrscheinlichkeit. Die Ergebnisse bestätigen damit, dass Fixationen Informationen liefern bezüglich des mentalen Zustands sowie der Aufgabe des Betrachters.

Kanan u. a. [Kan14] analysieren ebenfalls die Daten der 16 Versuchspersonen aus dem Experiment von Greene u. a. [Gre12]. Klassifikation mit einer Radial-Basis Function Support Vector Machine mit einem Merkmalsvektor aus mittlerer Fixationsdauer und Anzahl Fixationen verbesserte das Ergebnis von Greene u. a. kaum (26,3% korrekt, 95% KI = 21,4-31,1%, p=0,61). Ein Klassifikationsverfahren, das als Merkmale die Fixationen mit ihren (x,y)-Koordinaten und Fixationsdauern nutzt (vgl. den Ansatz von Borji u. a. [Bor14] mit der geglätteten Fixationskarte, diese Merkmale erhalten die zeitliche Dynamik der Blickpositionen), erzielt ein Korrektklassifikationsergebnis signifikant über der Zufallswahrscheinlichkeit mit 33,1% (95% KI = 27,9-38,3%).

Henderson u. a. [Hen13] klassifizierten die vier Aufgaben Visuelle Suche in einer Szene, Erinnern einer Szene, sowie Lesen und Pseudo-Lesen signifikant über Zufallswahrscheinlichkkeit mit einem linearen Klassifikator (Naiver Bayesscher Klassifikator). Der Merkmalsvektor umfasste 8 Fixations- und Sakkaden-bezogene Merkmale.

Kardan u. a. [Kar15] klassifizieren die drei Aufgaben Visuelle Suche, Szenenerinnerung und Entscheidung der ästhetischen Präferenz - signifikant über Zufallswahrscheinlichkeit mit einem linearen Klassifikator (Lineare Diskriminanzanalyse). Der Merkmalsvektor umfasste 7 Fixations- und Sakkaden-bezogene Merkmale.

2.5 Leistungsbewertung in der MMI: Normen, Metriken, Fragebögen

Über die Jahre wurden die Erkenntnisse über die menschliche Leistungsfähigkeit bzw. über die Methoden und Metriken, mit denen man sie bestimmt, in nationalen und internationalen Normen festgelegt. Relevant für die vorliegende Arbeit sind die Normen der Gruppe DIN EN ISO 9241, die Richtlinien für die Mensch-Computer-Interaktion beschreiben. Zentral für die Bewertung der Qualität eines Systems ist der Begriff der Gebrauchstauglichkeit.

Gebrauchstauglichkeit ist definiert als das „Ausmaß, in dem ein System, ein Produkt oder eine Dienstleistung durch bestimmte Benutzer in einem bestimmten Nutzungskontext genutzt werden kann, um bestimmte Ziele effektiv, effizient und zufriedenstellend zu erreichen." [DIN18b]. Gebrauchstauglichkeit setzt sich also aus drei Merkmalen zusammen: Effektivität, Effizienz und Zufriedenstellung der Benutzer.

Effektivität ist definiert als „Genauigkeit und Vollständigkeit, mit denen Benutzer bestimmte Ziele erreichen". Genauigkeit ist dabei „…der Grad der Übereinstimmung eines tatsächlichen mit einem angestrebten Ergebnis". Vollständigkeit ist definiert als „…das Ausmaß, in dem die Benutzer des Systems, des Produkts oder der Dienstleistung in der Lage sind, alle angestrebten Ergebnisse zu erreichen." Dabei ist es möglich, dass die Ergebnisse vollständig sind, ohne dass alle Ergebnisse vollkommen genau sind.

Ein wichtiges Maß für die Effektivität ist der Grad der Zielerreichung (in Prozent) oder die Fehlerquote (vgl. die Selektionsfehlerquoten in den Evaluationen blickbasierter Interaktionstechniken in Abschnitt 2.4.2).

Effizienz ist definiert als „Eingesetzte Ressourcen im Verhältnis zu den erreichten Ergebnissen“. Als typische Ressourcen werden angeführt die verwendete Zeit, der aufgewendete menschliche Aufwand sowie aufgewendete finanzielle Ressourcen und Materialien, mit denen ein Ziel erreicht wird.

Ein wichtiges Maß für Effizienz ist die Zeit für die Erledigung einer Aufgabe (vgl. die Selektionszeiten in den Evaluationen blickbasierter Interaktionstechniken in Abschnitt 2.4.2). Der aufgewendete menschliche Aufwand lässt sich zum Beispiel mit Hilfe des NASA-TLX-Fragebogens erfassen, der die subjektiv empfundene Belastung erfasst [DIN18a, Har88, Har06].

Zufriedenstellung ist definiert als „Ausmaß der Übereinstimmung der physischen, kognitiven und emotionalen Reaktionen des Benutzers, die aus der Benutzung eines Systems, eines Produkts oder einer Dienstleistung resultieren, mit den Benutzererfordernissen und Benutzererwartungen“. Die Vorgängernorm definierte Zufriedenstellung als „Freiheit von Beeinträchtigungen und positive Einstellungen gegenüber der Nutzung des Produkts“ [DIN18a].

Zufriedenstellung kann gemessen werden mit einem Fragebogen zur Einzelbewertung von Interaktionstechniken (vgl. die Evaluation von Zhang u. a. [Zha07] in Abschnitt 2.4.2), wie er von der Norm DIN CEN ISO/TS 9241-411:2014 (bzw. der Vorgängernorm 9241-9:2000) im Anhang C als Tabelle C.1 vorgeschlagen wird [DIN14, DIN00].

Neben Fragebögen gibt die Norm DIN CEN ISO/TS 9241-411:2014 auch Hinweise und Vorschläge, wie Tests der Gebrauchstauglichkeit durchführbar sind. Vorschläge, die in der Fachliteratur genutzt werden, sind die Tipptests, sowohl der Tipptest mit einer Richtung (vgl. Vertegaal [Ver08] in Abschnitt 2.4.2) als auch der Tipptest mit mehreren Richtungen (vgl. Zhang u. a. [Zha07] in Abschnitt 2.4.2).

Als menschzentrierte, zu erfüllende Qualitäten gelten neben der Gebrauchstauglichkeit auch die Barrierefreiheit, die User Experience sowie die Vermeidung nutzungsbedingter Schäden [DIN18b]. Diese werden in der vorliegenden Arbeit nicht vertieft.

Shneiderman u. a. [Shn18, S. 348] nennen als Erfolgskriterien, mit denen Zeigegeräte zu bewerten sind

- Geschwindigkeit und Akkuratheit
- Wirksamkeit (engl. *efficacy*) für die vorgesehene Aufgabe
- Anlernzeit
- Kosten und Zuverlässigkeit
- Größe und Gewicht

Verglichen mit den Kriterien und Metriken aus den Normen gehören Akkuratheit und Wirksamkeit zur Effektivität; Geschwindigkeit gehört zur Effizienz. Eigenschaften wie Anlernzeit und Kosten, aber auch Größe und Gewicht gelten in den Normen ebenfalls als Aspekte von Effizienz. Zuverlässigkeit betrifft die technische Zuverlässigkeit im Betrieb und liegt damit außerhalb der Gebrauchstauglichkeit. Sie ist jedoch ebenfalls ein wichtiger Faktor, den eine alternative Selektionstechnik erfüllen muss.

3 Konzept für blickbasierte Interaktion in Bildfolgen mit dynamischen Szenen

Wie in Kapitel 1 dargelegt, besteht die Problemstellung der vorliegenden Arbeit darin, Mechanismen zur Leistungssteigerung und zur Entlastung eines Benutzers bei der Bildfolgenanalyse zu erforschen. Der Lösungsansatz ist, die Benutzungsschnittstelle um Eyetracking bzw. die Erfassung des Blicks des Benutzers zu ergänzen. Abschnitt 1.3 beschreibt die im Lösungsansatz adressierten Themen und die Vorgehensweise für die drei Beiträge, die die vorliegende Arbeit leistet.

In diesem Kapitel werden die konzeptionellen Arbeiten beschrieben, für jeden der drei Beiträge separat.

Abschnitt 3.1 beschreibt das Konzept für Beitrag 1 (Blickbasierte Interaktion für Selektionsoperationen an bewegten Objekten, vgl. Abschnitt 1.3.1). Ziel ist, eine im Vergleich zur Mauseingabe *leistungsfähigere und belastungsärmere Selektionstechnik* zu identifizieren. Die Nutzerstudien, anhand derer das Konzept überprüft wird, sind in Kapitel 5 und Kapitel 6, Abschnitt 6.1, beschrieben.

Abschnitt 3.2 beschreibt das Konzept für Beitrag 2 (Blickbasierte Interaktion bei automatischen Bildanalyseverfahren, vgl. Abschnitt 1.3.2). Die Nutzerstudien, anhand derer das Konzept überprüft wird, sind in Abschnitt 6.2 beschrieben.

Abschnitt 3.3 beschreibt das Konzept für Beitrag 3 (Blickbasierte Klassifikation der Benutzertätigkeit bei der Bildfolgenanalyse, vgl. Abschnitt 1.3.3). Die

Datenerhebung, anhand der das Konzept überprüft wird, ist in Abschnitt 6.4 beschrieben.

3.1 Konzept für die Identifikation geeigneter Blickinteraktionstechniken für die Bewegtobjektselektion

In Abschnitt 1.3.1 wurden die Eigenschaften formuliert, die die neue Interaktionstechnik zu Bewegtobjektselektion im Vergleich zur traditionellen Mausinteraktion haben soll:

(A) *Die Aufmerksamkeit des Benutzers soll kontinuierlich auf der Bildfolge und den zu analysierenden Objekten liegen können.*
Die Erwartung ist, dass mit ununterbrochener Aufmerksamkeit auf der Bildfolge weniger Zielobjekte unentdeckt bleiben. Außerdem bedeutet ungeteilte Aufmerksamkeit auf dem Bildmaterial eine kognitive Entlastung.

(B) *Die Selektion soll schneller durchführbar sein.*
Die Erwartung ist, dass der Benutzer mit einer Interaktionstechnik, die eine kürzere Selektionszeit benötigt, auch kurz sichtbare Zielobjekte selektieren kann. Außerdem kann der Benutzer seine Aufmerksamkeit nach erfolgter Selektion rascher auf andere Bildinhalte richten.

(C) *Die Selektion soll mit geringerer Belastung - sowohl kognitiv als auch manuell - durchführbar sein.*
Die Erwartung ist, dass geringere Belastung zu weniger verpassten Selektionen führt und sich die Benutzer komfortabler und zufriedener fühlen.

Vergleicht man diese Ziele mit den Vorgaben, die die passenden Normen aus der Normengruppe 9241 [DIN18b, DIN14] zur Erreichung guter Gebrauchstauglichkeit für eine Interaktionstechnik formulieren, so hat die Erfüllung von

Forderung (A) das Potenzial, bessere Effektivität aufgrund geringerer Selektionsfehlerquoten sowie bessere Effizienz aufgrund geringerer kognitiver Belastung zu bewirken. Die Erfüllung von Forderung (B) hat das Potenzial, bessere Effizienz aufgrund kürzerer Selektionszeit zu bewirken. Die Erfüllung von Forderung (C) hat das Potenzial, bessere Effizienz sowie höhere subjektive Zufriedenstellung aufgrund geringerer kognitiver und manueller Belastung zu bewirken.

Um aufzuzeigen, wie blickbasierte Interaktion im Vergleich zur Mauseingabe die Forderungen (A)-(C) erfüllen kann, zerlegen wir den Gesamtprozess einer Selektionsoperation in drei Schritte:

1. Visuelle *Wahrnehmung* des Zielobjekts.
2. Auswahl des Zielobjekts zur Selektion durch *Zeigen* auf das Zielobjekt (ggf. mit visueller Kontrolle der Zeigerposition).
3. *Auslösen* der Selektion.

Die Schritte erfolgen zeitlich aufeinander, überlappen aber auch. Die visuelle Wahrnehmung des Zielobjekts erfolgt über den gesamten Zeitraum eines Selektionsprozesses. Das Zeigen auf das Zielobjekt (räumliche Komponente der Selektionsoperation) und das Auslösen der Selektion (zeitliche Komponente der Selektionsoperation) kommen nacheinander hinzu.

Mit traditioneller Mausinteraktion bewerkstelligt der Benutzer (1) mit den Augen, (2) mit Koordination von Hand und Auge, (3) mit der Hand. Optimierungen im Sinne der Forderung (A)-(C) sind bei allen drei Schritten möglich. Beitrag 1 betrachtet Optimierungen bei den Schritten (2) und (3).

Wird die Blickrichtung zum Zeigen auf dem Bildschirm genutzt, so fällt Schritt (2) mit Schritt (1) zusammen. Wenn, wie in der Fachliteratur postuliert, das Zeigen mit dem Blick ohne permanente visuelle Rückmeldung der Blickposition auf dem Bildschirm realisiert wird (vgl. Abschnitt 2.4.2.3), ist Forderung (A) während des Zeigevorgangs erfüllt. Denn der Benutzer kann seine Aufmerksamkeit jetzt kontinuierlich auf das Zielobjekt richten, da keine Cursorpositionierung mit dem Auge überwacht werden muss. Das

Zusammenfallen von Schritt (1) und (2) lässt eine Reduzierung der Selektionszeit erwarten und erfüllt damit Forderung (B). Da der Zeigevorgang ausschließlich mit dem Auge bzw. ohne Hand-Auge-Koordination erfolgt, ist die Komplexität der Interaktion deutlich reduziert, wodurch Forderung (C) nach sowohl kognitiver als auch manueller Entlastung erfüllt ist.

Um diese Vorteile gegenüber der Mausinteraktion nicht zu konterkarieren, muss die Realisierung von Schritt (3) sorgfältig gewählt werden. Der folgende Abschnitt stellt blickbasierte Selektionstechniken aus Abschnitt 2.4.2 zusammen und überprüft sie auf Erfüllung der Forderungen (A)-(C).

3.1.1 Mechanismus zur Selektionsauslösung

3.1.1.1 Unimodale Realisierungen

Naheliegend erscheint auf den ersten Blick die Realisierung mit *Verweildauer*, bei der das Auslösen durch eine (verlängerte) Fixation geschieht (vgl. S.63). Denn der Blick ruht mit Beginn des blickbasierten Zeigens auf dem Zielobjekt und kann für das Auslösen ohne Aufmerksamkeitswechsel dorthin gerichtet bleiben. (A) und (C) wären also erfüllt. Als unimodale Interaktionstechnik hat die Verweildauer jedoch den Nachteil, dass sie das Midas Touch-Problem weniger zuverlässig löst als multimodale Interaktionstechniken (Abschnitt 2.4.2). Falsch positive Selektionen würden mehrfache Selektionsversuche nach sich ziehen und so die Interaktionszeit insgesamt erhöhen, was Forderung (B) konterkarieren würde. Verweildauer ist daher nicht geeignet.

Lidschlag ist ungeeignet, da auch dieser das Midas Touch-Problem nicht zuverlässig löst bzw. es nur dann löst, wenn ein unnatürlich langer Lidschluss genutzt wird. Dies konterkariert (B), aber auch (A), da langer Lidschluss die visuelle Aufmerksamkeit unterbrechen kann.

Softwaretaste (Schaltfläche auf der GUI) mit Verweildauer unterbricht ebenfalls die Aufmerksamkeit auf dem Zielobjekt und erzwingt eine zusätzliche Augenbewegung. (A) und (C) sind nicht erfüllt.

3.1.1.2 Multimodale Realisierungen

Von den multimodalen blickbasierten Interaktionstechniken ist die Realisierung mit *Hardware-Taste* günstig, wenn diese gedrückt werden kann, ohne den Blick auf die Taste zu richten. In diesem Fall ist (A) erfüllt.

Dies ist der Fall für eine *Fußtaste*. Die Interaktionstechnik **Blick+Fußtaste** (im Folgenden auch abgekürzt mit „BFT“) kommt komplett ohne manuelle Intervention aus. Dadurch hat sie den Vorteil, die Hand und auch den restlichen Bewegungsapparat der oberen Extremität (Arm, Schulter, Nacken, Finger) zu entlasten, sodass (C) für den Anteil der manuellen Entlastung erfüllt ist. Fußinteraktion ist jedoch untypisch für die Interaktion mit Desktopsystemen. Benutzer sind daher ungeübter, eine Fußtaste zu drücken als eine Handtaste, wenngleich vielen Benutzern Fußinteraktion vom Autofahren her geläufig ist. Es ist daher zu überprüfen, ob (B) bei Nutzung einer Fußtaste erfüllt ist. Blick+Fußtaste gehört daher zu den evaluierten Interaktionstechniken (vgl. Abschnitt 5.1.3).

Für eine *Handtaste* ist „blindes“ Drücken möglich, wenn sie in besonderer Weise exponiert ist, sodass sie alleine über die Haptik gegriffen werden kann. Für Computertastaturen gilt dies für die SPACE-Taste als die größte Taste sowie für die ENTER-Taste des NumPad, die exponiert in der rechten unteren Ecke der Tastatur liegt. Denkbar wären weitere exponierte Tasten wie die CTRL-Taste, die in der linken unteren Ecke liegt. Alle drei liegen am unteren Rand der Tastatur, sodass ein versehentliches Drücken anderer Tasten durch den Handballen ausgeschlossen ist. (A) ist hier also erfüllt.

In der vorliegenden Arbeit fiel die Wahl auf die ENTER-Taste des NumPad. Der Grund dafür ist die mit dieser Taste verknüpfte Semantik und Funktion aus der Alltagsnutzung, die eine leichtere mentale Modellbildung ermöglicht. Denn während die SPACE-Taste zur Strukturierung von Texteingaben dient, ist ENTER, zu Deutsch „Eingabe“, unmittelbar verknüpft mit dem Abschluss einer Systemeingabe – eine Selektionsauslösung ist der Abschluss eines Selektionsprozesses.

Es ist zu erwarten, dass auch (B) und (C) erfüllt sind, da die manuelle Zeigeoperation wegfällt. Im Folgenden wird diese Interaktionstechnik als **Blick+Taste**, kurz „BT“, bezeichnet und evaluiert (vgl. Abschnitte 5.1.1, 5.1.3, 5.1.4, 6.1 und 6.2).

Blindes Drücken ist auch für Tasten der Computermaus möglich. Die Realisierung von Blick+linke Maustaste wurde jedoch nicht weiter betrachtet. Denn gegeben, dass der Benutzer die Computertastatur für gelegentliche Texteingabe nutzt, muss er zum Erreichen der Maus eine größere Armbewegung (verbunden mit mehr manueller Anstrengung) machen als zum Erreichen der NumPad-ENTER-Taste. Gegeben die Situation bei ABUL, wo der Benutzer die kleinen Bildschirmelemente manuell selektiert, spräche für die linke Maustaste, dass der Zeigefinger dort bereits zu deren Selektion dient. Solche Doppelnutzung könnte einerseits die Selektionszeit kurzhalten (B). Andererseits wurde in Abschnitt 1.2.2 auf die Problematik medizinischer Probleme hingewiesen, wenn der Benutzer die Hand über längere Zeit starr auf der Computermaus hält. In der vorliegenden Arbeit wurde Blick+linke Maustaste nicht evaluiert.

Auch bei Realisierung der Auslösung über ein *Sprachkommando* kann die Aufmerksamkeit ungeteilt auf dem Bildschirm bleiben (A) und manuelle Entlastung ist dadurch gesichert, dass keinerlei manuelle Aktion erforderlich ist (C). Verwandte Arbeiten zeigten jedoch, dass Blick+Sprache im Vergleich zu Mauseingabe (s. S.71) oder zu Blick+linke Maustaste (s. S.84) erheblich längere Selektionszeiten aufweist. (B) ist also nicht erfüllt. Zudem ist neben dem Eyetracker zusätzliche technische Ausrüstung sowie spezielle Software für Spracherkennung erforderlich. Korrekte Spracherkennung kann erschwert sein, wenn in der Umgebung des Benutzers Hintergrundgeräusche oder Gespräche zu hören sind. Blick+Sprache wurde daher als ungeeignet bewertet und nicht evaluiert.

Ebenfalls nicht betrachtet wurde Blick+*Mimik*. Der Grund ist das Erfordernis spezieller Software zur Erkennung von Mimik. Außerdem ist eine mimische Geste, die mit der Auslösung einer Aktion intuitiv mental verknüpft wäre,

schwierig zu finden. Am ehesten nutzbar wären Gesten, die mit positiver Bestätigung verknüpft sind, wie Kopfnicken oder Lächeln. Kopfnicken hat jedoch den Nachteil, dass die Augen mit dem Kopf zusammen nach unten, weg vom Bildschirminhalt gerichtet werden, (A) wäre nicht erfüllt. Lächeln könnte als Auslöse-Mechanismus sicher gelernt werden; es bliebe vermutlich trotzdem wenig intuitiv und damit für den Benutzer kognitiv und zudem motorisch belastend, (C) wäre nicht erfüllt.

Während alle bisher genannten Realisierungen eine explizite, bewusste Benutzer-Aktion erfordern, bietet eine Selektionsauslösung mit *EEG* die Möglichkeit einer impliziten, passiven Selektionstechnik, im Folgenden mit **Blick+EEG** bezeichnet. Zur Selektionsauslösung bietet sich das ereigniskorrelierte Potenzial der P300 an (s. Abschnitt 2.4.4), die ein starkes Signal liefert. Mit der P300 ist bei Blick+EEG nicht nur keine manuelle Aktion, sondern überhaupt keine bewusste motorische Aktion erforderlich, (C) ist also erfüllt. Es ist zu erwarten, dass auch (B) erfüllt ist, da der Zeitbedarf für eine motorische Aktion wegfällt. Es ist zu erwarten, dass (A) erfüllt ist, da Blick+EEG keinen Anlass zur Abwendung der Aufmerksamkeit von der Bildfolge gibt.

Ohne Zweifel ist jedoch das Tragen einer EEG-Kappe eine Zusatz-Belastung für den Benutzer, und die Kosten bzw. Größe und Gewicht sind im Vergleich zu einer Hand- oder Fußtaste enorm. Zudem ist die Erfassung der P300 algorithmisch herausfordernd. Nach Stand der Forschung muss vor einer Realisierung von Blick+EEG (P300) als echtzeitfähiger Selektionstechnik in eine GUI daher zuerst ihre Machbarkeit erforscht werden. Blick+EEG wird daher ebenfalls evaluiert, und zwar in Form von Datenerhebungen bei Selektionsaufgaben, wobei die Ermittlung von Kenngrößen für ihre Effektivität und Effizienz mithilfe von Offline-Analyse der Blick- und EEG-Daten erfolgt (vgl. Abschnitte 5.2.1 und 5.2.2).

3.1.2 Mechanismen zur Verbesserung der Zeigegenauigkeit

Da der menschliche Blick ein verrauschtes Signal ist, ist die Zeigegenauigkeit mit dem Blick nicht Pixel-genau, sondern hat eine räumliche Genauigkeit in der Größenordnung des fovealen Bereichs von ca. 1,5-2° (vgl. Abschnitt 2.2, Seite 26). Hinzu kommt noch die Messunsicherheit des Eyetrackers. Verwandte Arbeiten zur Selektion statischer Objekte bestätigen die effektive Selektierbarkeit einer Objektgröße ab 2°, wenn die Blickeingabe mit Kopfstabilisierung durchgeführt wird (z. B. [War86, Zha07], vgl. Abschnitt 2.4.2).

Kopfstabilisierung ist jedoch für Eyetracking in der MMI nicht akzeptabel, wenn es um Nutzung in der Praxis geht. Dies gilt auch für den in der vorliegenden Arbeit betrachteten Anwendungsfall der Bildfolgenanalyse, vor dessen Hintergrund alle Untersuchungen durchgeführt werden.

Ohne Kopfstabilisierung nimmt die räumliche Zeigegenauigkeit ab, sodass zahlreiche Arbeiten Objektgrößen zwischen 2° und 5° betrachten (z. B. [Ver08, Mon05, Sib00, Sta95, Bed09]). Stampe u. a. [Sta95, Maj02] berichten aus Vortests mit einem kopfgetragenen Eyetracker, dass sich die Selektionsfehlerquote für Objektgrößen bis 4° deutlich verbesserte, für größere Größen jedoch nur noch minimal. Die wenigen verwandten Arbeiten zu blickbasierter Bewegtobjektselektion (überwiegend aus dem Anwendungsgebiet der Computerspiele) machen kaum Angaben zu genutzten Objektgrößen; aus Abbildungen lässt sich jedoch auf ähnliche Größen von mehreren Grad Sehwinkel schließen (ca. 4,47° [Ram19], 4° [Vel15a], s.a. Abschnitt 2.4.3).

Erfolgt der Zeigeanteil einer Selektionstechnik also auf Basis des Blicksignals, ist es zweckmäßig die Blickrohdaten mit einem **Blickfilterverfahren** zu glätten. Es gibt jedoch nur wenige echtzeitfähige Blickfilterverfahren und alle haben den Nachteil, dass sie die aktuelle Blickposition zumindest mit einer kurzen Latenz liefern. In der vorliegenden Arbeit wurde bei der Evaluation der Interaktionstechniken daher zunächst gar nicht gefiltert. Als sich bei einer Untersuchung der vorliegenden Arbeit nach einer Offline-Nachverarbeitung der Blickrohdaten mit dem Algorithmus **RT-SDFS** von Kumar u. a. [Kum08]

(Abschnitt 2.4.1) bessere Genauigkeit ergab, wurde bei unseren Experimenten danach stets mit diesem Algorithmus gefiltert.

Die speziell für die Bewegtobjektselektion vorgeschlagene Interaktionstechnik *Pursuits* umgeht das Problem unzureichender Zeigegenauigkeit, indem die Trajektorien von Objektbewegung und Blickbewegung über einen Zeitraum korreliert werden. Als mit der Verweildauer verwandte Interaktionstechnik (vgl. S.80 bzw. [Vid13]) erfüllt sie Forderung (A) und (C), nicht jedoch (B) aufgrund vergleichweise langer Selektionszeiten sowie ungewollter Auslösung [Vel16]. Zudem erfordert *Pursuits* Kenntnis der Objektpositionen. Eine Nutzung zur Markierung von Objekten in einem Überwachungsvideo bei der Verkehrs- und Umweltüberwachung ist also nicht ohne weitere Maßnahmen wie automatischer Objektdetektion möglich. Aus diesen Gründen wurde Pursuits in der vorliegenden Arbeit nicht evaluiert.

Eine andere Möglichkeit, bessere Selektionsgenauigkeit zu erzielen, ist, den Blick nur zum groben Zeigen zu nutzen, und für das feine Zeigen die Computermaus zu nutzen. Diesen Ansatz verfolgt die Interaktionstechnik **MAGIC pointing** (vgl. Abschnitt 2.4.2, S. 71). Da der Mauszeiger unmittelbar vor der Selektion in Blickpositionsnähe positioniert ist – wenn nicht im Bereich der Fovea (2°), so ist der Zeiger doch im parafovealen Bereich (5°) zu erwarten – ist (A) im Vergleich zur Mauseingabe eher erfüllt, denn die Mauszeigersuche auf dem Monitor entfällt. Was bleibt, ist, dass die Aufmerksamkeit zwischen zwei visuellen Objekten – Zielobjekt und Mauszeiger – aufgeteilt werden muss. (B) und (C) bzgl. manueller Entlastung sind erfüllt, da der manuelle Zeigevorgang eine kleinere Amplitude überbrücken muss. Für die Selektion statischer Zielobjekte gelang mit der Variante MAGIC pointing-liberal tatsächlich eine schnellere Selektion als mit Mauseingabe.

Andererseits ist MAGIC pointing eine komplexe Interaktionstechnik, weil die Kontrolle über den Mauszeiger zwischen Hand und Blick wechselt. Die Forderung (C) nach kognitiver Entlastung ist also möglicherweise nicht gegeben. Höhere kognitive Belastung könnte daher auch die kürzere Selektionszeit (B) beeinträchtigen. Der Vorteil, dass MAGIC pointing wie die Mauseingabe grundsätzlich pixelgenaue Selektion ermöglicht, ist jedoch groß genug, dass MAGIC poiting ebenfalls evaluiert wird (vgl. Abschnitte 5.1.1 und 5.1.2).

3.1.2.1 Fazit

Die folgenden vier blickbasierten Selektionstechniken erscheinen aufgrund theoretischer Überlegungen als geeignete Kandidaten für eine Alternative zur Computermaus:

- Blick+Fußtaste
- Blick+Taste (NumPad ENTER)
- Blick+EEG
- MAGIC pointing

Ihre Eignung wird in der vorliegenden Arbeit in mehreren Nutzerstudien experimentell überprüft. Unabhängige Variable der Experimente ist dabei stets die Interaktionstechnik (mit den getesteten Ausprägungen als Anzahl Stufen dieses Faktors). Als abhängige Variable werden Merkmale der Gebrauchstauglichkeit (Effektivität, Effizienz, Zufriedenstellung) betrachtet. Effektivität wird objektiv gemessen als Selektionsfehlerquote oder Selektionstrefferquote sowie als Selektionsgenauigkeit, Effizienz wird objektiv gemessen als Selektionszeit; Zufriedenstellung wird subjektiv gemessen mithilfe des standardisierten Fragebogens aus der Norm DIN CEN ISO/TS 9241-411 [DIN14].

Mit Ausnahme der Untersuchungen zu Blick+EEG wird zusätzlich zu den blickbasierten Interaktionstechniken stets auch die Mauseingabe als Stand der Technik für Selektionsoperationen bei Desktop-Computersystemen bzw. der Videobildauswertung getestet.

Angaben der getesteten Objektgrößen sowie die Ergebnisse für Selektionsgenauigkeit erfolgen in Grad °, weil dies angemessen ist für Blickinteraktion, dazu in Pixel, weil dies angemessen ist für Mausinteraktion, und in mm oder cm, weil der Mensch gewohnt ist, in diesen Einheiten zu denken. Angaben der getesteten Objektgeschwindigkeiten erfolgen entsprechend als °/s, Pixel/s und mm/s bzw. cm/s.

Um eine Balance zwischen interner Validität und externer Validität zu erzielen, werden einerseits gut kontrollierte Experimente durchgeführt und andererseits solche mit mehr Realitätsbezug zur Bildfolgenanalyse bzw. Videobildauswertung, wie sie in der Praxis vorkommt.

Jede der betrachteten Interaktionstechniken wird in mindestens einer kontrollierten Studie an abstrakten visuellen Stimuli untersucht, bei denen die Objekteigenschaften Größe und Geschwindigkeit kontrolliert werden. Versuchsparadigmata sind leichte Abwandlungen der in der DIN CEN ISO/TS 9241-411 vorgeschlagenen Tipptests mit Selektionen von Punktepaaren als Einzelaufgaben (engl. *trials*), wie sie von anderen Autoren vorgeschlagen wurden (vgl. die Nutzerstudien Abschnitte 5.1.2 bis 5.1.4 und 5.2.1). In unserem Fall besteht jedes Punktepaar aus einem stationären Startobjekt und einem bewegten Zielobjekt. Die Abwandlungen stellen sicher, dass die Position des bewegten Zielobjekts erst nach erfolgreicher Selektion des Startobjekts enthüllt wird. Auf diese Weise umfasst jede Einzelaufgabe alle drei Schritte des oben beschriebenen Selektionsprozesses (s. S. 97).

Ergänzt werden die Tipptest-Studien zum einen durch Studien mit abstrakten visuellen Stimuli, die nicht dem Trial-Paradigma folgen, sondern Versuchsaufgaben so gestalten, dass sie einfache, typische Aufgaben der Bildfolgenanalyse bzw. Videobildauswertung simulieren (Abschnitte 5.1.1, 5.1.4 und 5.2.2). Zum anderen werden zwei Studien an Full Motion Video-Datenmaterial durchgeführt (Abschnitt 6.1).

Die Experimente umfassen mehrere Querschnittstudien und eine Längsschnittstudie. Wie in der MMI-Forschung häufig der Fall, sind die Versuchspersonen in der Mehrzahl Nicht-Experten für Bildfolgenanalyse. Ergänzt werden diese durch ein Experiment mit Videoauswerteexperten.

Im Sinne interner Validität wurden die Untersuchungsbedingungen so weit möglich standardisiert, um Versuchsleiterartefakte (vgl. [Bor09], dort S. 82ff) minimal zu halten. Dazu wurden folgende Maßnahmen ergriffen:

- Die Instruktionen zu den Testaufgaben wurden schriftlich festgehalten und vom Versuchsleiter abgelesen, um sicherzustellen,

dass die Versuchspersonen dieselbe Instruktion erhalten; Verständnisprobleme wurden individuell ausgeräumt.

- Alle Experimente umfassten eine Trainingsphase für jede der getesteten Interaktionstechniken, in der der Versuchsleiter anwesend war, um Fragen individuell zu beantworten. Die Trainingsphase war für Interaktionstechniken mit zu erwartender längerer Anlernzeit umfangreicher und von längerer Dauer.
- Die Durchführung der Testaufgaben erfolgte sofern möglich ohne Anwesenheit des Versuchsleiters; d.h. die Versuchspersonen führten gesteuert von der jeweiligen Versuchssoftware die Testaufgaben eigenständig durch. Versuchssoftware sowie Funktionalität der Eingabegeräte, insbesondere des Eyetrackers wurden ausgiebig getestet, um Datenverluste während der Datenerhebung zu vermeiden.
- Versuchsort war bei den Experimenten mit Nicht-Experten das Blickmesslabor des Fraunhofer IOSB, wo die Umgebungsbedingungen (Lichtquellen, Temperatur, Geräuschpegel) gut kontrollierbar waren. Das Experiment mit Videoauswerteexperten fand in verschiedenen Räumlichkeiten an unterschiedlichen Bundeswehr-Standorten statt, ebenfalls mit kontrollierten Umgebungsbedingungen.

3.2 Konzept für blickbasierte Interaktion bei automatischen Bildanalyseverfahren

Bevor der Videoexperte im Rahmen eines Videoauswerteprozesses ein Zielobjekt selektiert, um es zu markieren und die Information dadurch für weitere Verantwortliche nutzbar zu machen, muss er andere Tätigkeiten durchführen. Er muss Zielobjekte als solche wahrnehmen und dafür die Szene visuell durchmustern. Dabei muss er die visuelle Information verarbeiten und mit der Zielobjektspezifikation abgleichen, um ein detektiertes Objekt als „Zielobjekt“

oder „Kein Zielobjekt“ zu klassifizieren und entsprechend die Entscheidung zur Markierung zu treffen (vgl. Abb. 2.1).

Der Vorgang der Wahrnehmung und Informationsverarbeitung kann aufgrund der Gegebenheiten der Szene schwierig sein (Abschnitt 1.2.1). Um zu einer Entscheidung bezüglich der Klassifikation als Zielobjekt zu kommen, muss der menschliche Beobachter das Objekt im Video eine Weile mit den Augen verfolgen und seine Merkmale eingehend inspizieren, um sicher zu sein, dass das Objekt als Zielobjekt zu markieren ist. Oder die Situation ist so, dass ein Objekt zwar äußerlich alle visuellen Merkmale der Zielobjektspezifikation erfüllt, aber erst durch sein Verhalten zum Zielobjekt wird, etwa ein Fahrzeug, wenn es an einem bestimmten Ort parkt. Auch in diesem Fall muss der menschliche Beobachter das Objekt über einen Zeitraum verfolgen, was in einem unruhigen Videostrom (Abb. 5.3) anstrengend und herausfordernd sein kann.

Bei beiden Tätigkeiten – *Detektion eines Objekts* in der Szene sowie *Objektverfolgung* – kann der Mensch Fehler machen. Das menschliche Auge ist zwar versiert, bewegte Objekte zu detektieren. Bei schwierigen Situationen mit kleinen, zahlreichen und über die gesamte Szene verteilten Objekten kann es jedoch passieren, dass ein Objekt übersehen wird. Ein **automatisches Verfahren zur Bewegungsdetektion** kann hier assistieren, indem es jede Bewegung in der Szene visuell hervorhebt.

Das menschliche Auge ist auch versiert darin, ein sich langsam bewegendes Objekt vor einem Hintergrund zu verfolgen – mit der Gleitenden Folgebewegung (vgl. S. 31) hat die Evolution hierfür einen eigenen Mechanismus hervorgebracht. Ein Objekt über einen längeren Zeitraum zu beobachten, kann jedoch anstrengend und herausfordernd sein. Wenn beispielsweise die Szene unruhig ist, weil der Sensor durch Böen stark bewegt wird, kann das Objekt aus der Szene geraten und wieder in sie eintreten. Der menschliche Beobachter muss das Objekt dann immer wieder neu in der Szene detektieren und als Zielobjekt erkennen. Objektverfolgung über einen längeren Zeitraum kann zudem monoton sein, sodass die Eintönigkeit der Tätigkeit den Benutzer ermüdet und unaufmerksam werden lässt. In beiden Fällen kann ein **automatisches Verfahren für Einzelobjekttracking** assistieren.

Heutige Algorithmen für automatische Bewegungsdetektion und Einzelobjekttracking liefern ihre Verfahrensergebnisse robust. Allerdings sind sie in manchen Situationen nicht zu hundert Prozent korrekt. Automatische Bewegungsdetektion kann falsch positive Bewegung detektieren oder Bewegung nicht detektieren; automatisches Objekttracking kann das Objekt verlieren und abbrechen.

Aus diesem Grund ist stets ein menschlicher Auswerter im Auswerteprozess erforderlich, der die Verfahrensergebnisse überprüft und angemessen berücksichtigt. Im Falle der automatischen Bewegungsdetektion überprüft der menschliche Auswerter nicht nur Bildbereiche, für die das Verfahren Bewegung visualisiert, sondern auch für andere, an denen Bewegung zu erwarten ist. Im Falle des automatischen Objekttrackers wird der menschliche Auswerter bei Trackingabbruch gegebenenfalls das Verfahren neu auf ein Objekt anwenden, wenn es weiter getrackt werden soll.

Wenn ein menschlicher Benutzer eines der genannten automatischen Bildanalyseverfahren nutzt, liegt ein Fall von *Informationsfusion* vor. Denn das Ziel ist, durch die Fusion der Information, die das automatische Verfahren liefert, und der Information, die der Benutzer liefert, insgesamt ein besseres Analyseergebnis zu erzielen. Dasarathy [Das01, S. 1] definiert: „Information Fusion encompasses theory, techniques and tools conceived and employed for exploiting the synergy in the information acquired from multiple sources (sensor, databases, information gathered by human, etc.) such that the resulting decision or action is in some sense better (qualitatively or quantitatively, in terms of accuracy, robustness, etc.) than would be possible, if any of these sources were used individually without such synergy exploitation.“

Die Nutzung automatischer Verfahren dient dem menschlichen Benutzer zur Aufmerksamkeitssteuerung und reduziert seine perzeptive und kognitive Belastung. Gleichzeitig produzieren die Verfahrensergebnisse jedoch zusätzlichen visuellen Input, der vom Benutzer wahrgenommen werden muss, was die visuelle Perzeption wiederum zusätzlich belastet. Außerdem entsteht durch die Nutzung der automatischen Verfahren zusätzlicher Interaktionsaufwand.

Um die Wirkung automatischer Verfahrensergebnisse bzw. die Interaktion mit automatischen Bildanalyseverfahren zu erforschen, werden in der vorliegenden Arbeit Nutzerstudien anhand von Full Motion Video-Datenmaterial durchgeführt. Wie die Untersuchungen aus Beitrag 1 (vgl. Abschnitt 3.1 bzw. Kapitel 5 und Abschnitt 6.1) beinhalten sie die Interaktionsaufgabe der Bewegtobjektselektion.

Wie in Abschnitt 6.1 werden die leistungsfähigste blickbasierte Interaktionstechnik aus Kapitel 5 und die Mauseingabe evaluiert und verglichen.

Bei der Untersuchung zu automatischer *Bewegungsdetektion* (vgl. Abschnitt 6.2.1) findet keine Interaktion mit dem Verfahren selbst statt. Vielmehr wird der Aspekt der Informationsfusion betrachtet für die Aufgabe der Bewegtobjektmarkierung bei Assistenz durch automatische Verfahrensergebnisse. Die Versuchsaufgabe ist hier, spezifizierte Zielobjekte zu finden, und zwar mit und ohne Verfügbarkeit automatischer Bewegungsdetektion.

Mit Verfügbarkeit bedeutet, dass das automatische Verfahren eingeschaltet ist und kontinuierlich die detektierten Bewegungen als visuelles Overlay anzeigt. Das automatische Verfahren assistiert dem menschlichen Beobachter dadurch bei der Detektion der bewegten Objekte, indem es permanent bewegte Objekte als potenzielle Zielobjekte vorschlägt. Der menschliche Beobachter bestätigt dann im Falle zutreffender Spezifikation die potenziellen Zielobjekte als tatsächliche Zielobjekte, indem er sie durch eine Selektion markiert.

Ziel ist herauszufinden, welche Leistung und Belastung die Versuchspersonen mit bzw. ohne Nutzung der automatischen Verfahrensergebnisse erzielen bzw. welchen Einfluss die Informationsfusion dabei hat.

Bei der Untersuchung zu automatischem *Einzelobjekttracking* wird die Interaktion betrachtet, die der Benutzer durchführen muss, um das Verfahren auf ein bestimmtes Objekt im Video anzuwenden. Dies wird auch als Initialisierung oder Aufsetzen eines automatischen Einzelobjekttrackers bezeichnet.

Der Benutzer muss dabei durch Bewegtobjektselektion dem System die Position des Zielobjekts genügend präzise mitteilen, sodass das automatische

Einzelobjekttracking-Verfahren in der Lage ist, das Zielobjekt robust zu tracken. Falls die Präzision nicht ausreicht, gelingt das Initialisieren entweder nicht oder das Verfahren trackt das Objekt nur kurz und bricht dann ab. In beiden Fällen muss dann erneut aufgesetzt werden.

Dies wird für ein Einzelobjekttracking-Verfahren in einer Pilotstudie (Abschnitt 6.2.2) und in einer Studie mit Videoanalyseexperten untersucht (Abschnitt 6.2.3).

3.3 Konzept für die blickbasierte Klassifikation der Benutzertätigkeit bei der Bildfolgenanalyse

Wie in Abschnitt 2.4.5 beschrieben, konnten verschiedene Autoren auf Basis von aggregierten Maßen der Blickdynamik (Fixations- und Sakkadendynamik) auf die Aufgabe schließen, die die Versuchspersonen bei der Betrachtung von Einzelbildern durchführten.

Im Rahmen der vorliegenden Arbeit wird untersucht, ob dies auch für Aufgaben bzw. Tätigkeiten[1] gelingt, die bei der Bildfolgenanalyse vorkommen.

Betrachtet werden die vier Tätigkeiten, die Hild u. a. [Hil12] vorgeschlagen haben:

1. Suchen
2. Erkunden
3. Beobachten
4. Verfolgen

[1] Wir verwenden die Begriffe „Aufgabe“ und „Tätigkeit“ hier nebeneinander, da sie zwei Seiten desselben Vorgangs beschreiben: „Aufgabe“ ist im Sinne von Aufgabenstellung zu verstehen, und „Tätigkeit“ beschreibt die Handlung, die der Mensch durchführt, um die Aufgabe zu erfüllen. Letztlich wird ein Videoanalyseexperte im Rahmen der Erfüllung einer Aufgabenstellung (eines Auftrags) mehrere Tätigkeiten im Wechsel durchführen.

Diese Menge von Tätigkeiten ergab sich zum einen auf Basis von Literaturrecherche. Theiler u. a. [The03] nennen *Suchen* als grundsätzliche Aufgabe der Bildauswertung („The job of the professional image analyst is to find things in imagery"). Sie unterscheiden dabei zwischen Suchaufgaben, bei denen der Auftrag genau spezifiziert ist, d.h. es ist klar, welches Objekt zu suchen ist, und solchen, bei denen der Auftrag vage bleibt und allgemeiner nach „ungewöhnlichen" Dingen und Ereignissen gesucht werden soll.

Bowyer [Bow04] nennen *Erkunden* und *Überwachen* als Aufgaben eines Videobildauswerters.

Während Suchen und Erkunden die Durchmusterung der (ggf. gesamten) Szene nach einem Zielobjekt beinhalten, kann Überwachen entweder auf einen Ort in der Szene (ggf. auch auf die gesamte Szene) bezogen sein oder aber nur auf ein bestimmtes Zielobjekt. Diese beiden Situationen wurden bei der Konzeption der betrachteten Tätigkeiten berücksichtigt, indem jede als separate Tätigkeit definiert wurde: Das Überwachen einer Szene wurde als *Beobachten* aufgenommen, das Überwachen eines Zielobjekts als *Verfolgen*.

Die Unterschiedlichkeit der vier Tätigkeiten lässt sich untermauern mithilfe des von Theiler u. a. [The03] genannten Unterscheidungsmerkmals der Konkretheit der Auftragsbeschreibung als eher vage oder eher konkret (s.o.). Wendet man das Begriffspaar vage-konkret auf die Zielobjekteigenschaften *Was* ist das Objekt und *Wo* ist es *Wann* an (vgl. a. die Teilaufgaben Sach-, Orts- und Zeitbestimmung bei der Szenenanalyse bei Geisler [Gei06]), so ergibt sich folgendes:

Suchen ist *konkret* bezüglich *Was* und *vage* bezüglich *Wo* und *Wann*. Passend dazu beschreibt der Duden [Wer10] *suchen* als „(…) etwas (…) Verstecktes zu finden". Ein beispielhafter Auftrag könnte lauten, ein genau spezifiziertes Fahrzeug (Marke, Modell, Farbe) zu suchen.

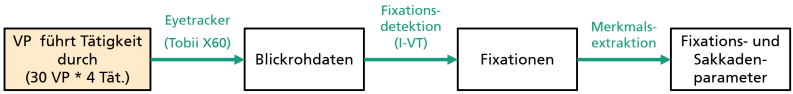
Erkunden ist *vage* sowohl bezüglich *Was* als auch bezüglich *Wo* und *Wann*. Passend dazu beschreibt der Duden [Wer10] *erkunden* als „Sich genaue Kenntnis von etwas (bisher Unbekanntem) verschaffen". Ein beispielhafter Auftrag könnte lauten, ein bestimmtes Gebiet nach Auffälligkeiten zu erkunden, ohne

spezifizierte Angabe der Art der Auffälligkeit. Hier würde der Videoanalyseexperte basierend auf seiner Erfahrung ggf. mehrere Objekte als (potenziell) relevant bewerten.

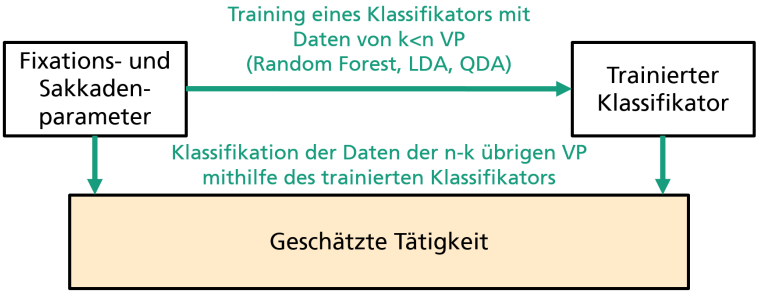
Beobachten ist *konkret* bezüglich *Was*, *konkret* oder *vage* bezüglich *Wo* sowie *vage* bezüglich *Wann*. Anders als bei *Suchen* wird das *Was* hier nicht als spezifiziertes Einzelobjekt konkret, sondern in Form eines zu überprüfenden Regelkatalogs. Passend dazu beschreibt der Duden [Wer10] *beobachten* als „(…) zu einem bestimmten Zweck kontrollierend auf jmd., etwas achten; (…)". Im Vergleich zu *Suchen*, das man als aktiven Vorgang der Suche nach einem genau spezifizierten Zielobjekt in der Szene betrachten kann, hat *Beobachten* eher passiven Charakter. Ein beispielhafter Auftrag könnte lauten, ein bestimmtes Gebiet bezüglich spezifizierter Regelverstöße zu überwachen.

Verfolgen ist definiert als *konkret* bezüglich *Was* und *Wo* und *Wann*. Passend dazu beschreibt der Duden [Wer10] *verfolgen* als „(…) den Verlauf (von etwas) genau betrachten". Ein beispielhafter Auftrag könnte lauten, ein genau spezifiziertes und bereits gefundenes Zielobjekt (Fahrzeug mit Marke, Modell, Farbe, KFZ-Kennzeichen) über eine bestimmte Strecke oder Zeitdauer hinweg zu überwachen.

Die Vorgehensweise bei der Klassifikation der vier Tätigkeiten ist wie folgt (Abb. 3.1). Basis sind die Blickrohdaten von Versuchspersonen, die in einer Datenerhebung während der Durchführung der vier Tätigkeiten aufgezeichnet werden (Abb. 3.1a). Dazu wird für jede Tätigkeit eine entsprechende Versuchsaufgabe gestaltet. Die aufgezeichneten Blickrohdaten werden dann mithilfe des I-VT-Algorithmus (vgl. S. 60) in eine Menge von Fixationen aggregiert. Auf Basis der Fixationen werden dann aggregierte Fixations- und Sakkadenparameter berechnet, die als Merkmale für die Klassifikation dienen (Abb. 3.1b).



(a) Merkmalsextraktion.



(b) Klassifikation.

Abbildung 3.1: Vorgehensweise bei der Klassifikation der Benutzertätigkeiten.

Verwendete Fixations- und Sakkadenparameter sind in Anlehnung an die Fachliteratur [DeA09, Jar10, Gre12, Bor14, Kan14, Kar15, Boi16, Le 17]:

1. Fixationen pro Sekunde (arithmetisches Mittel)
2. Fixationsdauer (arithmetisches Mittel, Varianz)
3. Fixationsdurchmesser (artithmetisches Mittel)
4. Sakkadenamplitude (arithmetisches Mittel, Varianz). Jede einzelne Sakkadenamplitude wird über den euklidischen Abstand zwischen zwei aufeinanderfolgenden Fixationen berechnet.
5. Sakkadengeschwindigkeit (arithmetisches Mittel, Varianz)

6. Fixationswinkel (arithmetisches Mittel, Varianz). Die Winkel zwischen den Fixationen werden anhand von Dreiecken berechnet, wobei je drei Fixationen zur Winkelberechnung herangezogen werden und aus diesen ein Dreieck gebildet wird. Mithilfe des Kosinussatzes lässt sich der Winkel berechnen, der kleiner oder gleich 180° sein muss.

Der Hintergrund der Hinzunahme des Merkmals Fixationswinkel (bestimmt für je drei aufeinanderfolgende Fixationen; sonst nur verwendet in [Jar10]) ist die Annahme, damit die Tätigkeit *Verfolgen* von den anderen abgrenzen zu können. Denn bei Verfolgen ist bei den gegebenen Bewegungsgeschwindigkeiten der Objekte in der Videosequenz zu erwarten, dass ein menschlicher Beobachter Gleitende Folgebewegungen (Smooth Pursuits) (vgl. S. 31) mit zahlreichen einholenden („catch-up“) und antizipatorischen Sakkaden entlang der Bewegungstrajektorie des verfolgten Objekts vollführt. Daher sind für Verfolgen überwiegend Winkel zu erwarten, die nur geringe Bewegungsänderung des Blickpfads anzeigen.

Verwendete Klassifikationsverfahren sind in Anlehnung an die Fachliteratur [Gre12, Kar15, Boi16] Random Forests (RF), die Lineare Diskriminanzanalyse (LDA) sowie die Quadratische Diskriminanzanalyse (QDA).

Klassifikation mit Random Forests basiert auf einer Mehrheitsentscheidung einer Anzahl distinkter Entscheidungsbäume (siehe z. B. [Bre01]). Jeder Entscheidungsbaum besteht aus einem Wurzelknoten, von dem mehrere Kind-Knoten abzweigen. Am Ende jedes Astes befindet sich ein Blattknoten, der die prädizierte Tätigkeit anzeigt. Bei jedem Knoten wird für ein Merkmal mithilfe eines Schwellenwertes entschieden, ob dem linken oder rechten Kind-Knoten zu folgen ist. Schlussendlich bestimmt die Mehrheit der prädizierten Klassen aller Entscheidungsbäume die prädizierte Klasse.

Bei der LDA werden Merkmalskombinationen gelernt, um Klassen zu trennen (siehe z. B. [Alp14]). Die Trainingsmenge wird verwendet, um gewichtete Kombinationen von Merkmalen für jede Klasse zu bestimmen, wobei die Kovarianz innerhalb der Klassen minimiert und die Kovarianz zwischen den Klassen maximiert wird. Für die Klassifikation wird jedes Datensample der

Testmenge derjenigen Klasse zugeordnet, bei der die gewichtete Differenz zum Class-Mean-Vector minimal ist.

Bei der QDA sind die Merkmalskombinationen keine linearen Funktionen, sondern komplexere Funktion wie bspw. Hyperbeln im zweidimensionalen Raum (siehe z. B. [Alp14]). In Fällen, wo die Kovarianzmatrizen von Klassen nicht gleich sind, könnte das Ergebnis mit QDA besser werden im Vergleich zu LDA.

4 Systembeschreibung

Im Rahmen der vorliegenden Arbeit wurden mehrere Versuchssysteme realisiert, an denen die Untersuchungen durchgeführt wurden. Hardware-seitig verfügen alle über die typischen Komponenten von Desktop-PC-Systemen: Rechner, Monitor, Computermaus und Computertastatur. Hinzu kommen ein Eyetracker und in Einzelfällen eine USB-Fußtaste bzw. die EEG-Kappe der Kollegen der Uni Bremen.

Als Versuchssoftware wurden zum einen JAVA-Applikationen realisiert, zum anderen konnte das ABUL-Videoauswertesystem genutzt werden.

Als wichtiges Teilziel des Lösungsansatzes, Blickbewegungen für die Interaktion mit Bildfolgen zu nutzen, wurde die Reduzierung der Benutzerbelastung formuliert (vgl. Abschnitt 1.3).

Aus diesem Grund kommen als Eyetracker für den Systemaufbau nur videobasierte, nicht-intrusive Remote-Eyetracker in Frage (vgl. Abschnitt 2.2). Zudem müssen die Geräte erlauben, dass das Eyetracking ohne Kopfstabilisierung wie eine Stirn-Kinn-Stütze mit ausreichender Qualität möglich ist. Im Rahmen dieser Arbeit wurden zwei solche Eyetracker genutzt.

4.1 Versuchssystem I mit Tobii 1750

Für die erste Untersuchung (Kapitel 5, Abschnitt 5.1.1) wurde der zum damaligen Zeitpunkt am Fraunhofer IOSB vorhandene Tobii 1750 verwendet (Abb. 4.1). Die folgende Beschreibung seiner Eigenschaften ist aus der Produktbeschreibung von Tobii Technologies übernommen [AB05].

Abbildung 4.1: Tobii 1750-Eyetracker.

Beim Tobii 1750 ist die Hardware für die Blickmessung integriert in einen 17-Zoll-TFT-Monitor mit einer Auflösung von 1280 x 1024 Pixeln. Die Kamera zur Blickerfassung ist mittig unterhalb der aktiven Anzeige des Monitors in den Monitorrahmen integriert. Die Infrarot-LEDs sind seitlich links und rechts der Kamera sowie über der aktiven Anzeige des Monitors in den Monitorrahmen integriert. Während des Eyetrackings erzeugt der Tobii 1750 Cornea-Reflexionsmuster auf den Augen des Benutzers. Die Kamera erfasst diese und andere visuelle Information des Benutzers und liefert sie mit einer *Abtastrate* von 50 Hz an den Tobii Eye Tracker Server.

Der Tobii Eyetracker Server ist die Software, die das Tobii 1750-Gerät steuert und die Berechnung der Blickposition auf dem Monitor durchführt. Er läuft auf einem Standard-PC, der über ein USB-Kabel mit dem Tobii 1750 verbunden ist. Bildverarbeitungsalgorithmen identifizieren in den Kamerabildern relevante Blickmerkmale wie das Pupillenzentrum und die Cornea-Reflexionen.

Daraus wird mit mathematischen Modellen die dreidimensionale Position jedes Augapfels im Raum berechnet. Darauf basierend wird die Blickposition auf dem Monitor berechnet. Die *Latenz*[1] des Tobii 1750 wird mit 25 - 35 ms angegeben.

Die *räumliche Genauigkeit* (engl. *accuracy*) der Blickmessung bezeichnet die „(...) typische Differenz zwischen der gemessenen und der wirklichen Blickrichtung auf unterschiedlichen Teilen des Monitors für eine Person, die im Zentrum der Eyetracking-Box positioniert ist."[2]. Diese Messunsicherheit, mit der die Blickposition geliefert wird, wird vom Hersteller mit 0,5° angegeben [AB05].[3]

Da die Abweichung von 0,5° in alle Richtungen vorkommen kann, ist die gelieferte Blickposition in einer Unsicherheitsregion von 1° zu erwarten. Sitzt der Benutzer 60 cm vom Monitor entfernt, dann entspricht 1° 10,4 mm auf dem Monitor. Umgerechnet in Pixel sind dies bei der gegebenen Monitorauflösung von 1280 x 1024 Pixeln 36 Pixel.

Der Wert variiert in Abhängigkeit verschiedener Bedingungen wie Beleuchtung, Qualität der individuellen Kalibrierung oder individueller Augenmerkmale. In der praktischen Nutzung, wo die Benutzer den Kopf nicht angestrengt stillhalten, sondern in natürlicher Art und Weise zumindest minimal bewegen, sind daher etwas schlechtere Werte als 0,5° zu erwarten.

[1] Für die Definition der hier und nachfolgend genannten technischen Eyetracker-Merkmale vgl. a. Abschnitt 2.2

[2] Eigene Übersetzung, engl. Original: „The typical difference between the Measured Gaze Direction and the Actual Gaze Direction at different parts of the screen for a person positioned in the center of the eye tracking box" [AB08]

[3] Die Vorgehensweise, die zum Ergebnis von 0,5° führte, ist in der Produktbeschreibung auf Seite 11 als Fußnote beschrieben. Demnach führten die Versuchspersonen zunächst eine Kalibrierung durch und betrachteten danach 64 Punkte, die gleichmäßig auf dem Monitor verteilt waren. Dabei hielten die Versuchspersonen ihren Kopf still, nutzten jedoch keinerlei Kopfstabilisierungsmechanismus wie eine Kinnstütze oder eine Beißschiene. Für jeden der 64 Punkte wurde der gemittelte Blickpunkt aus dem Durchschnitt von fünf Blickdatenpunkten berechnet. Die *accuracy* wurde dann definiert als der Durchschnitt der Abweichung der 64 Punkte von den intendierten Blickpunkten. Die durchschnittliche *accuracy* der Ergebnisse von 10 typischen Benutzern betrug 0,5°.

Die *räumliche Auflösung* beträgt 0,25°, sie bezeichnet die Variation der gemessenen Blickposition zwischen je zwei Kamerabildern. Die *Drift* beträgt < 1°. Unter Drift wird verstanden, dass die Kalibrierung des Eyetrackers mit der Zeit schlechter wird, was vorkommen kann, wenn sich die Umgebungsbedingungen ändern. Beispiele sind veränderte Beleuchtungsbedingungen (Umgebungsbeleuchtung oder Helligkeit des Monitorinhalts) oder trockene Augen.

Das *Sichtfeld der Kamera* beträgt 21 x 16 x 20 cm, wenn der Benutzer einen Abstand von 60 cm zur Kamera hat. Da es ausreicht, wenn nur ein Auge im Sichtfeld der Kamera liegt, ergibt sich eine *Kopfbewegungsfreiheit* von 30 cm horizontal, 16 cm vertikal und 20 cm in der Tiefe. Laut Hersteller ist dies ausreichend, um die Kopfpositionen, die Benutzer von Desktop-Systemen typischerweise einnehmen, zu kompensieren. Denn durch das binokulare Tracking genügt es, wenn ein Auge im Sichtfeld liegt. Kopfbewegungen können im gesamten Sichtfeld der Kamera kompensiert werden mit einem Fehler kleiner 1°.

4.2 Versuchssystem II mit Tobii X60

Alle anderen Untersuchungen nutzten den Tobii X60, der vom Fraunhofer IOSB als Nachfolger für den Tobii 1750 beschafft wurde. Die folgende Beschreibung seiner Eigenschaften ist aus der Produktbeschreibung von Tobii Technologies übernommen [AB10].

Der Tobii X60 ist anders als der Tobii 1750 nicht in einen Monitor integriert, sondern als separate, freistehende Einheit realisiert. Der X60 wird über ein LAN-Kabel mit einem Standard-PC verbunden und nutzt den dort angeschlossenen Monitor.

Um die Blickposition auf dem Monitor korrekt bestimmen zu können, müssen die geometrischen Beziehungen zwischen X60 und Monitor sowie die Monitorgröße dem Eyetracker bekannt sein. Dies geschieht mithilfe eines Konfigurationstools der Tobii-Software, die folgende Angaben benötigt: Monitorgröße (Breite und Höhe in mm), Monitorwinkel (gekippt nach hinten oder vorn in °), Eyetrackerwinkel (in °, gemessen von der Ebene, auf der der Eyetracker

steht, als 0°), Abstand des Eyetrackers vom Monitor sowie die Höhendifferenz zwischen Monitor und Eyetracker.

Über die Kamera und die Infrarot-Dioden sowie ihre Anordnung macht die Produktbeschreibung keine Angaben. In der praktischen Nutzung kann man LEDs an den Stellen erkennen, die in Abb. 4.2 rot eingezeichnet sind. Ob die Kamera in der Mitte platziert ist, also zwischen den beiden LED-Ringen, ist unklar. Möglicherweise werden auch zwei Kameras genutzt, die in den Ringen platziert sind. Anhaltspunkt dafür ist, dass die Produktbeschreibung an einer Stelle [AB10, S.4] von „image sensors“ in der Mehrzahl spricht, die die visuelle Information über den Benutzer erfassen.

Die *Abtastrate* beim X60 beträgt 60 Hz. Die Berechnung der Blickposition auf dem Monitor erfolgt auf dem Eyetracker selbst. Die Berechnung nutzt dieselben Schritte wie der Tobii 1750: Die Infrarot-LEDs generieren ein Reflexionsmuster auf der Cornea der Augen; die Kamera(s) erfasst(/en) dieses Muster zusammen mit anderen visuellen Informationen über den Benutzer; Bildverarbeitungsalgorithmen identifizieren relevante Merkmale (Augen, Cornea-Reflexionen); daraus wird die Position jedes Augapfels in 3D bestimmt und schließlich daraus die Blickposition auf dem Monitor berechnet. Die *Latenz* des Tobii X60 wird mit maximal 33 ms angegeben.

Abbildung 4.2: Tobii X60-Eyetracker.

Die Tracking-Qualität wird optimiert und für eine größere Bevölkerungsgruppe robustifiziert, indem während der Benutzer-Kalibrierung sowohl die Bright- als auch die Dark-Pupil-Methode [Hol11, S.25, S.127] angewendet werden und danach automatisch die besser funktionierende weiter genutzt wird.

Die *Akkuratheit* der Blickmessung wird vom Hersteller mit „typischerweise 0,5°“ angegeben [AB10]. Die *räumliche Auflösung* beträgt 0,2°, die *Drift* 0,1°.

Die *Kopfbewegungsfreiheit* liegt bei 44 cm horizontal und 22 cm vertikal, wenn die Augen des Benutzers 70 cm vom Eyetracker entfernt sind. Als Distanz, bei der Eyetracking möglich ist, werden 50 - 80 cm angegeben. Kopfbewegungen können im gesamten Sichtfeld der Kamera kompensiert werden mit einem Fehler von 0,2°. Binokulares Tracking bewirkt robustere Toleranz bezüglich Kopfbewegungen, da das Tracking auch dann fortgesetzt werden kann, wenn nur ein Auge im Sichtfeld liegt.

Verliert der Eyetracker den Blick, beispielsweise weil sich der Benutzer vom Eyetracker wegdreht, so dauert es typischerweise 300 ms, bis das Eyetracking wiederhergestellt ist. Im Falle von Lidschlag, wo Beleuchtung und Erfassung von Pupille und CornearefleXionen blockiert sind, beträgt die Wiederherstellzeit 17 ms.

Tabelle 4.1 stellt die Merkmale der beiden Eyetracker einander gegenüber. Beide liefern die Blickposition mit gleicher Messunsicherheit, der X60 mit einer etwas höheren Abtastrate als der Tobii 1750. Der X60 hat zudem die großzügigeren Werte bezüglich Kopfbewegungen: Bei besserer Kopfbewegungskompensation ist die Kopfbewegungsfreiheit größer und die maximale Kopfbewegungsgeschwindigkeit, bei der der Blick noch getrackt werden kann, ist höher. Dies ermöglicht dem Benutzer uneingeschränktere Bewegung, was ein natürlicheres und entspannteres Benutzerverhalten erlaubt. Außerdem ist eine längere Eyetracker-Nutzung ohne erneute Kalibrierung möglich, da der Wert für die Drift beim X60 deutlich besser ist.

Tabelle 4.1: Die im Rahmen dieser Arbeit eingesetzten Eyetracker Tobii 1750 und Tobii X60 und ihre technischen Spezifikationen.

Kenngröße	Tobii 1750	Tobii X60
Räumliche Genauigkeit (accuracy)	0,5°	typischerweise 0,5°
Räumliche Auflösung	0,25°	typischerweise 0,2°
Drift	< 1°	typischerweise 0,1°
Kopfbewegungsfreiheit	30 x 15 x 20 cm bei 60 cm	44 x 22 x 30 cm bei 70 cm
Sichtfeld der Kamera	20 x 15 x 20 cm bei 60 cm	k.A.
Maximaler Blickwinkel	40°	35°
Binokulares Tracking	Ja	Ja
Kopfbewegungskompensationsfehler	< 1°	typischerweise 0,2°
Maximale Geschwindigkeit der Kopfbewegung	ca. 10 cm/s	25 cm/s
Datenrate	50 Hz	60 Hz
Latenz	25-35 ms	maximal 33 ms
Wiederherstellzeit bei Trackverlust	< 100 ms	typischerweise 300 ms
Wiederherstellzeit nach Lidschlag	k.A.	maximal 17 ms
Gewicht	ca. 11 kg	ca. 3 kg

Der im Versuchsaufbau mit dem Tobii X60 genutzte **Monitor** ist ein Dell U2412M. Er hat eine Größe von 24 Zoll (Seitenverhältnis 16:10) und eine Auflösung von 1920 x 1200 Pixeln.

Monitor und Eyetracker sind mithilfe einer Tobii Monitor Mount-Halterung fest miteinander verbunden (Abb. 4.3). Diese Halterung besteht aus einem beweglichen Arm, der an die Tischplatte montiert ist. Der Monitor ist an die Haltungsvorrichtung montiert. Damit der Eyetracker unterhalb des Monitors platziert werden kann, ist in die Halterung eine Metallvorrichtung mit einem Fach eingehängt, in das der Eyetracker gestellt wird.

Der Aufbau mit Tobii Monitor Mount bewirkt, dass sich die geometrischen Beziehungen zwischen Monitor und Eyetracker nicht oder kaum verändern. Trotzdem wurden sie vor jeder Datenerhebung überprüft und gegebenenfalls im Konfigurationstool korrigiert.

Abbildung 4.3: Der Versuchsaufbau mit Tobii X60 und 24-Zoll-Monitor verbunden über eine Tobii Monitor Mount-Halterung.

Bei den Datenerhebungen saßen die Versuchspersonen wie in Abb. 4.3 zu sehen vor dem Versuchsaufbau. Der Abstand der Augen vom Monitor bzw. vom Eyetracker betrug ca. 65 cm. Die Unsicherheits-Region von 1°, in der die Blickposition laut Hersteller-Spezifikation bestenfalls zu erwarten ist, ist bei diesem Abstand auf dem Monitor im Durchmesser 11,3 mm. Für die gegebene Auflösung entsprechen 11,3 mm 42 Monitorpixeln.

Als **Rechner** für den X60 wurde ein Standard-PC mit Windows 7 als Betriebssystem in der 64-bit-Edition genutzt. Er verfügt über einen Intel Core i7-2600K-Prozessor mit Taktfrequenz 3,40 GHz, 16 GB RAM und eine NVIDIA GeForce GTX 460-Grafikkarte.

Der X60 wird über ein LAN-Kabel angeschlossen. Abb. 4.4 zeigt die Verarbeitungskette für die Blickdaten. Der Eyetracker führt die Messung der Blickposition durch und übergibt die Messdaten über ein TCP-Socket an einen C#-Prozess. Dieser transformiert die Messdaten in (x,y)-Koordinaten auf dem Monitor und versieht sie mit einem Zeitstempel. Der C#-Prozess reicht dann das Tupel (x_i, y_i, t_i) an einen Java-Prozess weiter. Dieser Java-Prozess wendet dann einen Filteralgorithmus auf die Blickdaten an und gibt das Ergebnis der Blickschätzung als Tupel $(\hat{x}_i, \hat{y}_i, t_i)$ über ein UDP-Socket an den Rechner weiter.

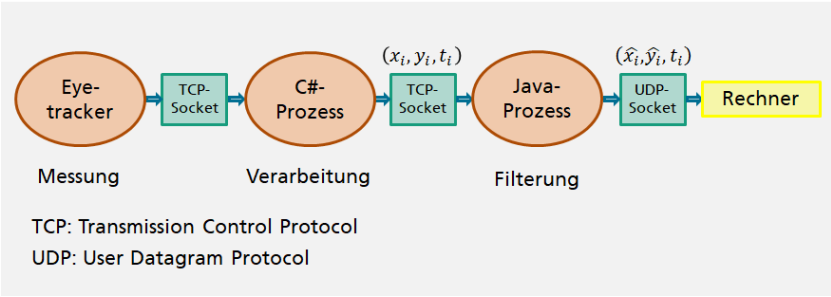


Abbildung 4.4: Verarbeitungskette der Blickdaten beim Tobii X60.

Bei den Untersuchungen aus Kapitel 5 und Abschnitt 6.4 laufen auf diesem Rechner auch die Software-Applikationen, die die Versuchsaufgaben realisieren.

Bei den Untersuchungen aus Abschnitt 6.1 und Abschnitt 6.2 dient der Rechner nur als Blickrechner für die Unterstützung und Anbindung des X60-Eyetrackers. Die Software-Applikationen, die die Versuchsaufgaben realisieren, sind ABUL-Derivate und laufen unter Linux auf einem separaten Videorechner, an den auch Computermaus und Computertastatur angeschlossen sind. Der Blickrechner ist über ein LAN-Kabel mit dem Videorechner verbunden und liefert darüber die Blickdaten für die Interaktion.

4.3 Peripherie-Geräte: Computermaus, Computertastatur und Fußtaste

Als Computermaus wird eine optische Maus nach Stand der Technik genutzt. D.h., wann immer in den Untersuchungen die *Mauseingabe* für Selektionsoperationen vorkommt, handelt es sich stets um die traditionelle Zeige-Klick-Interaktion mit Einzel-Klick der linken Maustaste.

Als Computertastatur wird eine Standard-PC-Tastatur genutzt.

Als Fußtaste wird der *USB Foot Switch 2 Single* von Scythe genutzt [Scy21] (Abschnitt 5.1.3).

4.4 EEG-Erfassung

Bei den Experimenten, die die Blick+EEG-Interaktion untersuchen, wurde für die EEG-Erfassung ein actiCHamp Recorder mit 32 Elektroden eingesetzt [Bra21]. Dabei wurden 28 aktive Elektroden gemäß dem internationalen 10-5-System auf Pz referenziert. Die übrigen drei wurden zur Erfassung des EOG um die Augen platziert gemäß dem Vorschlag von Schlögl u. a. [Sch07].

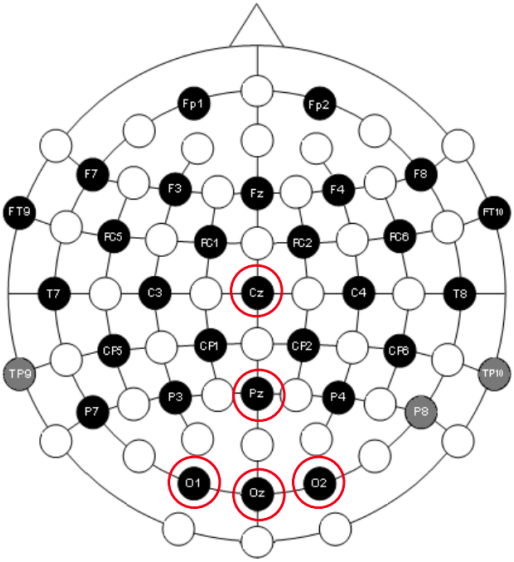


Abbildung 4.5: Die Platzierung der Elektroden bei der EEG-Datenaufzeichnung (Abschnitt 5.2).

4.5 Versuchsort

Für die überwiegende Anzahl an Untersuchungen fand die Datenerhebung am Fraunhofer IOSB statt. Das Blickmesslabor ist ein ruhig gelegener Raum, mit ausreichend Platz für Versuchsperson und Versuchsleiter und kontrollierbaren Lichtverhältnissen.

Für zwei Untersuchungen fanden Datenerhebungen mit Videoanalyseexperten statt. Diese wurden an verschiedenen Standorten der Bundeswehr durchgeführt. Dafür wurde der X60-Versuchsaufbau des Blickmesslabor an diese Orte verbracht.

5 Identifikation geeigneter Blickinteraktionstechniken für die Bewegtobjektselektion

Die Konzepterstellung in Kapitel 3 identifizierte als Ergebnis mehrere explizite und eine implizite blickbasierte Interaktionstechnik als potenziell leistungsstarke und belastungsarme Alternativen zur Mauseingabe für die Bewegtobjektselektion. Dieses Kapitel beschreibt die grundlegenden Untersuchungen, die im Rahmen der vorliegenden Arbeit für diese blickbasierten Alternativen durchgeführt wurden. Abschnitt 5.1 beschreibt die Untersuchungen zu expliziten blickbasierten Interaktionstechniken, Abschnitt 5.2 die Untersuchungen zur impliziten blickbasierten Interaktionstechnik Blick+EEG.

5.1 Explizite blickbasierte Bewegtobjektselektion

Dieser Abschnitt beschreibt vier Untersuchungen. Sie bauen aufeinander auf und dienen dem Zweck, diejenige explizite Blickinteraktionstechnik zu identifizieren, die am besten für die Bewegtobjektselektion geeignet ist. Diese wird dann auch für die Interaktion mit Überflugvideos eingesetzt und weiter evaluiert (siehe Kapitel 6).

Jede der vier Untersuchungen betrachtet eine oder mehrere Blickinteraktionstechniken und evaluiert sie in einer Nutzerstudie. Alle Untersuchungen evaluieren im Vergleich außerdem die Mauseingabe als Stand der Technik.

Mit folgenden Maßen wird die Qualität der Interaktionstechniken beschrieben, mit der die Versuchspersonen die Bewegtobjektselektionsaufgabe bewältigen:

- Selektionsfehlerquote (oder Selektionstrefferquote) als Maß der Effektivität
- Selektionszeit als Maß der Effizienz
- Fragebogen zur Einzelbewertung der Interaktionstechniken nach DIN EN ISO 9241-411 als Maß der Zufriedenstellung

Als Testaufgaben nutzen alle Untersuchungen visuelle Stimuli mit abstrakten Selektionsobjekten. Die Selektionsobjekte sind einfarbige Kreise, die sich vor einem einfarbigen Hintergrund bewegen. Variiert werden die Merkmale Objektgröße und Objektgeschwindigkeit, wobei sowohl typische Werte aus Bildfolgen von Überflügen als auch aus verwandter Fachliteratur berücksichtigt werden.

Verwandte Arbeiten, die Blickeingabe ohne Kopfstabilisierung betrachten, nutzten oft eine Objektgröße von 4-5°, sowohl zur Selektion unbewegter [Bed09, Ver08, Mon05, Sta95][1] als auch bewegter Objekte [Ram19, Vel15a]. Manche Arbeiten nutzen eine kleinere sichtbare Objektgröße in Kombination mit einer selektierbaren Objektgröße von 4-5° [Min04, Min06, Mon05]. Verwandte Arbeiten zur Selektion bewegter Objekte mit Mauseingabe betrachten für bessere Selektierbarkeit Objektvergrößerungen auf 2° (unsichtbare Vergrößerung) [Rag20] bis 9,64° Objektgröße (Schweiflänge bei [Has11] für Geschwindigkeit von 500 Pixel/s; für 800 Pixel/s 14,63°; vgl. Formel (2.5), S. 53). Es ist daher angemessen, diese Objektgrößen bzw. selektierbaren Objektgrößen bei kleineren sichtbaren Objektgrößen für die Gestaltung der Testaufgaben zu berücksichtigen.

Die Selektionsobjekte werden den Versuchspersonen (in fast allen Fällen) durch visuelle Hervorhebung verdeutlicht, d.h. der in der Praxis der Bildfolgenanalyse ggf. aufwendige Suchvorgang vor der Objektselektion entfällt.

[1] verwendete Eyetracker laut Hersteller mit räumlicher Genauigkeit von 0,5°

Wir folgen hier Ware u. a. [War86], die Hervorhebung aus dem Grund nutzen, dass der Suchvorgang zusätzliches Rauschen in den Ergebnissen erzeugen würde. Nachteilig daran ist, dass dies die externe Validität der Aufgabe beeinträchtigt, denn das Hervorheben verdeckt möglicherweise „den wahren Wert der Detektion der Blickrichtung als Mittel für die Selektion“ [War86]. Andererseits kann man argumentieren, dass sich der Zeitpunkt des Hervorhebens als Zeitpunkt der Einsicht des Benutzers interpretieren lässt, dass genau dieses Objekt Zielobjekt ist und selektiert werden muss.

Grundsätzlich streben die Versuchsdesigns in diesem Kapitel nach bestmöglicher Kontrolle der Variablen. Dies geht naturgemäß zu Lasten der externen Validität. Mehr Gewicht bekommt die externe Validität in den Untersuchungen in Kapitel 6, wo die Testaufgaben auf Basis von Bildfolgen von Überflügen gestaltet sind und zumindest zum Teil die Versuchspersonen ausgebildete Videoauswerteexperten sind.

In allen vier Untersuchungen erhielten die Versuchspersonen die Instruktion, die Bewegtobjektselektion stets so schnell und so genau wie möglich durchzuführen.

Die erste Untersuchung in Abschnitt 5.1.1 vergleicht *Blick+Taste*, *MAGIC pointing liberal* und *Mauseingabe* anhand einer abstrakten Testaufgabe, die einen Überflug simuliert. Ziel ist herauszufinden, ob die Benutzer mit den blickbasierten Techniken eine mindestens ähnlich gute Leistung erbringen wie mit der Mauseingabe. Ausgehend von den Ergebnissen wurden alle folgenden Untersuchungen geplant und gestaltet.

Die zweite Untersuchung in Abschnitt 5.1.2 vertieft den Vergleich zwischen der MAGIC pointing-Technik und der Mauseingabe. Sie vergleicht die beiden Realisierungen *MAGIC pointing konservativ* und die im Rahmen dieser Arbeit neu vorgeschlagene Variante *MAGIC button* mit der *Mauseingabe*.

Die dritte Untersuchung in Abschnitt 5.1.3 vertieft den Vergleich zwischen blickbasierten Techniken mit Tastenauslösung und der Mauseingabe. Sie betrachtet *Blick+Taste*, *Blick+Fußtaste* und *Mauseingabe*.

Die vierte Untersuchung in Abschnitt 5.1.4 betrachtet *Blick+Taste* in einer *Längsschnittstudie über 6 Monate*, und ermittelt die Entwicklung der Leistung und Zufriedenstellung der Versuchspersonen bei wöchentlichem Training.

5.1.1 Vergleich Blick+Taste, MAGIC pointing liberal, Mauseingabe

Dieses Experiment und seine Ergebnisse wurden bei der ACHI 2013 veröffentlicht [Hil13a].

Verglichen werden drei Interaktionstechniken bei der Aufgabe der Bewegtobjektselektion: Blick+Taste, MAGIC pointing liberal und Mauseingabe.

Die *Mauseingabe* erfolgt als traditionelle Zeige-Klick-Interaktion mit Einzel-Klick der linken Maustaste.

Die Selektion mit der *Blick+Taste*-Interaktion erfolgt, indem der Benutzer das Objekt anblickt und währenddessen die ENTER-Taste des NumPad drückt.

MAGIC pointing wird in Form des *liberalen* Ansatzes getestet, im Folgenden mit *MAGIC-lib* bezeichnet. MAGIC-lib zeigte in den Experimenten seiner Erfinder die beste Leistungsfähigkeit bei der Selektion statischer Objekte [Zha99]. Die Kontrolle über den Mauszeiger wechselt zwischen Eyetracker und Computermaus. Anfänglich liegt die Kontrolle beim Eyetracker, d.h., der Mauszeiger bewegt sich mit der Blickposition; das Eyetrackersignal wird gefiltert mit dem I-DT-Algorithmus [Sal00] (vgl. Abschnitt 2.4.1). Der Mauszeiger springt mit jeder Fixation, die mindestens 100 ms dauert. Will der Benutzer eine Selektion vollenden, so greift er die Maus, die dabei unmittelbar die Kontrolle über den Mauszeiger gewinnt. Der Benutzer bewegt den Mauszeiger manuell auf das Zielobjekt und selektiert wie gewohnt mit linkem Mausklick. Sobald die Maus für 100 ms nicht bewegt wird, wechselt die Kontrolle über den Mauszeiger wieder zum Eyetracker.

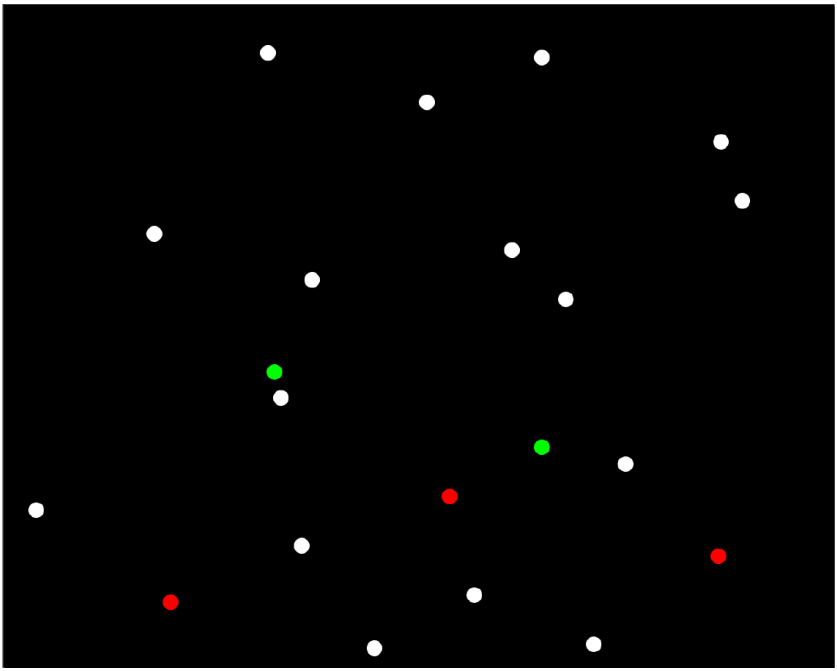


Abbildung 5.1: Visueller Stimulus der Testaufgabe.

5.1.1.1 Testaufgabe

Als visueller Stimulus für die Testaufgabe wurde eine Videosequenz von 3 Minuten Länge gestaltet. Abb. 5.1 zeigt den visuellen Stimulus. Die Gestaltung abstrahiert die Gegebenheiten eines realen Überflugvideos mit Fahrzeugen als Zielobjekten bezüglich ihrer sichtbaren Größe, ihrer Geschwindigkeit sowie der unterschiedlich großen Abstände zwischen den Objekten.

Die Selektionsobjekte sind als weiße Kreise mit einer sichtbaren Größe von 24 Pixeln Durchmesser (0,63 cm auf dem Monitor bzw. 0,60° Sehwinkel bei einem Augenabstand von 60 cm) gestaltet. Alle Objekte haben eine unsichtbare

Vergrößerung des selektionssensitiven Bereichs auf 160 Pixel (4,01° Sehwinkel), siehe Abb. 5.2.

Die Selektionsobjekte überlappen nicht in ihren sichtbaren Bereichen, können aber in ihren unsichtbaren Selektionsbereichen überlappen. Fällt eine Selektionsposition in den Selektionsbereich von mehr als einem Objekt, wird immer das Objekt selektiert, dessen Mittelpunkt der eingegebenen Selektionsposition am nächsten liegt.

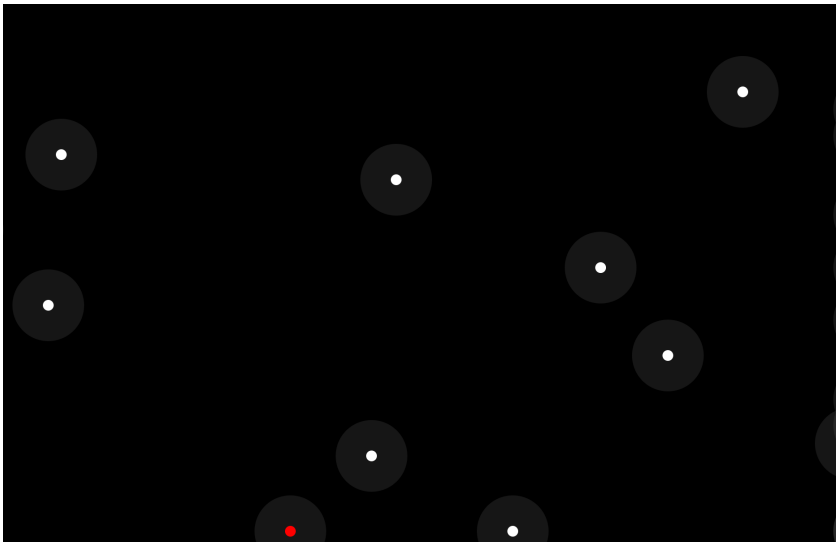


Abbildung 5.2: Sichtbare und selektierbare Objektgröße.

Alle 294 Objekte kommen am linken Monitorrand ins Bild, bewegen sich linear mit konstanter Geschwindigkeit über den Monitor und verlassen am rechten Monitorrand das Bild. Die Geschwindigkeiten betragen 115 px/s, 122 px/s, 128 px/s und 135 px/s (entspricht 2,88°/s, 3,05°/s, 3,20°/s, 3,37°/s bzw. 3,02 cm/s, 3,20 cm/s, 3,36 cm/s, 3,54 cm/s); jedes Objekt ist zwischen 9 und 11 Sekunden

auf dem Monitor zu sehen. Um etwas Zeitdruck bei der Selektion zu erzeugen, nimmt die Anzahl gleichzeitig sichtbarer Objekte über die Aufgabendauer hinweg zu.

Ebenso wie die sichtbare Objektgröße wurden auch die Objektgeschwindigkeiten angelehnt an Geschwindigkeiten in Überflugvideos gewählt. Abb. 5.3 zeigt eine Objektbewegung über einen Ausschnitt von 30 Sekunden aus Überflug-Videodatenmaterial. Dargestellt sind die Bild-zu-Bild-Distanzen der Positionen eines bewegten Fahrzeugs. Dabei überlagern sich die Eigenbewegung des Fahrzeugs und die Bewegung des Sensors.

Über weite Strecken sind die Distanzen gering und betragen deutlich unter 10 Pixel. Bei einer Bildrate von 25 Hz entspricht 1 Pixel Bild-zu-Bild-Distanz einer Geschwindigkeit von 25 Pixel/s. Die gelegentlich auftretenden großen Amplituden von bis zu 48 Pixeln (entspricht einer Geschwindigkeit von 1200 Pixel/s) sind auf starke, atmosphärisch hervorgerufene Sensorbewegungen zurückzuführen. Die Distanzen während der ruhigen Phasen betragen im Mittel ± 1 Standardabweichung 3,6 ± 2,3 Pixel, also 90 ± 58 Pixel/s. Für die Testaufgabe wurden basierend auf diesen Werten „typische“ Geschwindigkeiten so gewählt, dass sie zwischen diesem Mittelwert und dem Mittelwert + 1 Standardabweichung liegen.

Zum Zielobjekt, das selektiert werden muss, wird ein Objekt dann, wenn es sich für 1 Sekunde rot färbt; dies ist der Fall für 91 der 294 Objekte. Gelingt der Versuchsperson eine erfolgreiche Selektion, färbt sich das Objekt permanent grün. Für die Blick+Taste-Interaktion ist diese visuelle Rückmeldung der einzige Hinweis darauf, welches Objekt selektiert wurde; bei den beiden anderen Interaktionstechniken hat der Benutzer zusätzlich die visuelle Rückmeldung der Selektionsposition über den Mauszeiger. Eine fälschlicherweise erfolgte Selektion kann nicht rückgängig gemacht werden, sondern wird als falsch positive Selektion (Falschalarm, engl. *false alarm*) gezählt.

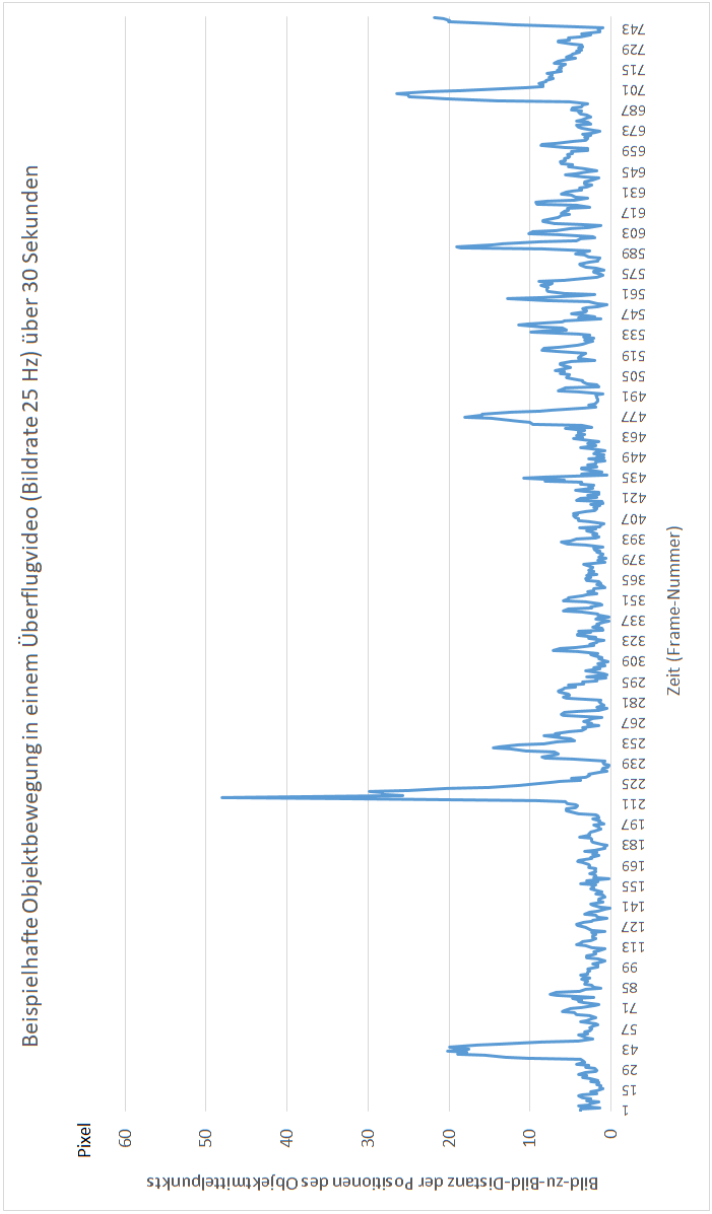


Abbildung 5.3: Typische Objektbewegung in einem Überflugvideo.

5.1.1.2 Versuchsaufbau, Versuchspersonen und Versuchsablauf

Die Versuchssoftware wurde als JAVA-Anwendung unter Windows 7 programmiert. Zur Darstellung der Animationen der Testaufgabe wurde die Trident 1-Bibliothek[1] von Kirill Grouchnikov verwendet. Sie stellt eine einfache Interpolation von Klassenattributen der x- und y-Koordinaten eines Objekts zwischen definierten Start- und Endwerten bereit. Trident erlaubt daher eine einfache Realisierung linearer Animationen. Angezeigt wird die Testaufgabe auf dem Tobii 1750-Monitor (17 Zoll, Auflösung 1280 x 1024 Pixel).

Verwendete Eingabegeräte sind der Tobii 1750 (Abschnitt 4.1), eine gewöhnliche USB-Computertastatur mit NumPad und eine gewöhnliche optische USB-Computermaus. Die Mauszeigergeschwindigkeit wurde auf 8/11 unter Windows 7 gesetzt.

An der Untersuchung nahmen 20 Versuchspersonen teil. Bei zwei von ihnen war das Ergebnis der Eyetracker-Kalibrierung zu schlecht (>1,5° Sehwinkel), sodass sie bereits die Trainingsaufgabe nicht durchführen konnten. Es wurden also Daten von 18 Versuchspersonen erfasst (16 männlich, 2 weiblich; Alter zwischen 18 und 54 Jahren, Altersdurchschnitt ± 1 Standardabweichung 28,6 ± 11,1 Jahre. Alle verfügten über normale oder auf normal korrigierte Sicht, zwei trugen eine Brille, drei Kontaktlinsen. Alle waren Kollegen oder Studenten, also keine Videoauswerteexperten. Alle waren erfahrene Benutzer der Computermaus, niemand hatte zuvor MAGIC pointing gekannt oder genutzt, neun hatten bereits geringe Erfahrung mit der Nutzung von Blick+Taste mit dem Tobii 1750 aus einem Systemtest, bei dem große statische Fenster selektiert werden mussten [Hil13c].

Das Versuchsdesign nutzte ein vollständiges, ausbalanciertes Within-Subjects-Design[2]. Jede Versuchsperson führte die Testaufgabe je einmal mit jeder der drei Techniken durch. Um Ermüdungs- und Lerneffekte zu kontrollieren, wurden die Versuchspersonen in 6 Versuchsgruppen mit je

[1] https://www.pushing-pixels.org/2009/06/21/trident-part-1-hello-world.html
[2] auch: repeated measures design oder dependent measures design [She20]

3 Versuchspersonen eingeteilt, wobei jede Versuchsgruppe eine der 6 möglichen Technik-Reihenfolgen durchführte.

Der Versuchsablauf war wie folgt. Als Erstes wurde in einer allgemeinen Einführung die Bewegtobjektselektion erläutert. Danach wurde die erste getestete Interaktionstechnik erläutert, gefolgt von der Standard-9-Punkt-Kalibrierung des Tobii 1750 im Falle der beiden blickbasierten Techniken. Danach konnten sich die Versuchspersonen anhand einer Trainingsaufgabe mit 50 Objekten mit der Aufgabe und der Interaktionstechnik vertraut machen und gegebenenfalls Fragen stellen. Für die Durchführung der Aufgabe wurden die Versuchspersonen instruiert, so schnell wie möglich und mit so wenig Fehlern wie möglich zu selektieren. Nach der Trainingsaufgabe absolvierten die Versuchspersonen eine erneute Kalibrierung im Falle der blickbasierten Techniken. Darauf folgte die Durchführung der Testaufgabe mit 294 Objekten. Abschließend bewerteten die Versuchspersonen subjektiv auf einer 5-Punkt-Skala (1: beste Bewertung, 5: schlechteste Bewertung) die Merkmale Selektionsgeschwindigkeit, Selektionsgenauigkeit, Aufgabenangemessenheit und Nutzerfreundlichkeit.

5.1.1.3 Forschungsfragen, Metriken und Ergebnisse

Dieses Experiment adressierte folgende Forschungsfragen:

- Frage 1: Welche Leistung erzielen die Versuchspersonen mit den blickbasierten Interaktionstechniken im Vergleich zur Mauseingabe? Ist die Leistung so gut, dass sie eine Alternative zur manuellen Mausselektion sein können?
- Frage 2: Wie leistungsfähig sind die Benutzer mit Blick+Taste (Zeigen erfolgt ausschließlich mit dem Blick) verglichen mit MAGIC pointing liberal (grobes Zeigen mit dem Blick, feines Zeigen manuell)?
- Frage 3: Wie ist die subjektive Bewertung der Aspekte Selektionsgeschwindigkeit, Selektionsgenauigkeit, Aufgabenangemessenheit und Nutzerfreundlichkeit?

Hierfür wurden folgende Metriken verwendet:

- Selektionsfehlerquote (falsch negative Selektionen (Verpasser[1], engl. *miss*) und falsch positive Selektionen (Falschalarm, engl. *false alarm*)): Sie berechnet sich als Anzahl fehlerhafter Selektionen / Gesamtaufgabenumfang (91 Objektselektionen).
- Selektionszeit: Die Selektionszeit wird berechnet als die zeitliche Differenz zwischen der Rotfärbung eines Objekts und seiner erfolgreichen Selektion.

Abb. 5.4 zeigt die Ergebnisse der Selektionsfehlerquote. Eine einfaktorielle Varianzanalyse (ANOVA) mit Messwiederholung (Signifikanzniveau $\alpha = 0{,}05$) zeigt signifikante Unterschiede zwischen den Interaktionstechniken ($F(2;52) = 17{,}41$; $p<0{,}001$). Post-hoc-Analyse mit Bonferronikorrektur zeigt signifikante Unterschiede ($p<0{,}001$) zwischen MAGIC pointing und Maus sowie zwischen MAGIC pointing und Blick+Taste. MAGIC pointing ist also signifikant schlechter als die beiden anderen Interaktionstechniken. Der Unterschied zwischen Mauseingabe und Blick+Taste ist nicht signifikant. Im Mittel machten die Versuchspersonen mit MAGIC pointing etwa doppelt so viele Fehler wie mit den beiden anderen Techniken.

Abb. 5.5 zeigt die Ergebnisse für falsche negative und falsch positive Selektionen getrennt. Während falsch positive Selektionen für alle drei Techniken praktisch kaum vorkommen, werden mit allen Techniken einige Zielobjekte verpasst. Auch für die Verpassten zeigt sich, dass MAGIC pointing signifikant schlechter abschneidet ($p<0{,}001$) als die beiden anderen Techniken.

Abb. 5.6 zeigt die Ergebnisse für die Selektionszeit. Die Unterschiede zwischen den Techniken sind signifikant mit ($F(2;50) = 510{,}7$; $p<0{,}001$). Post-hoc-Analyse mit Bonferronikorrektur zeigt hochsignifikante Unterschiede ($p<0{,}001$) zwischen Blick+Taste und jeder der beiden anderen Techniken, außerdem signifikante Unterschiede ($p<0.01$) zwischen Mauseingabe und MAGIC pointing. Mit Blick+Taste selektierten die Versuchspersonen im Mittel mehr als doppelt so schnell verglichen mit MAGIC pointing und annähernd doppelt so schnell wie mit der Maus.

[1] Bezeichnung vgl. Dorsch u. a. [Dor20, S.1631] unter **Signalentdeckungstheorie**

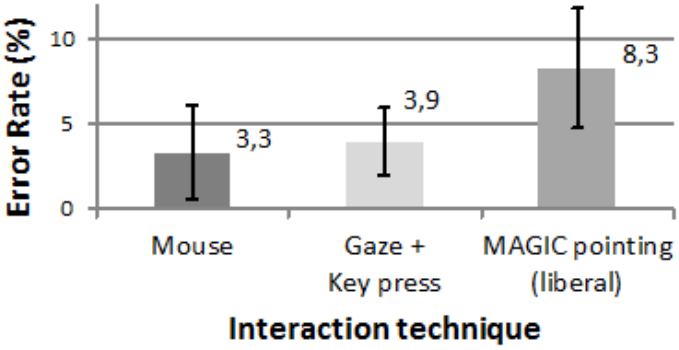


Abbildung 5.4: Selektionsfehlerquote als Funktion der Interaktionstechnik.

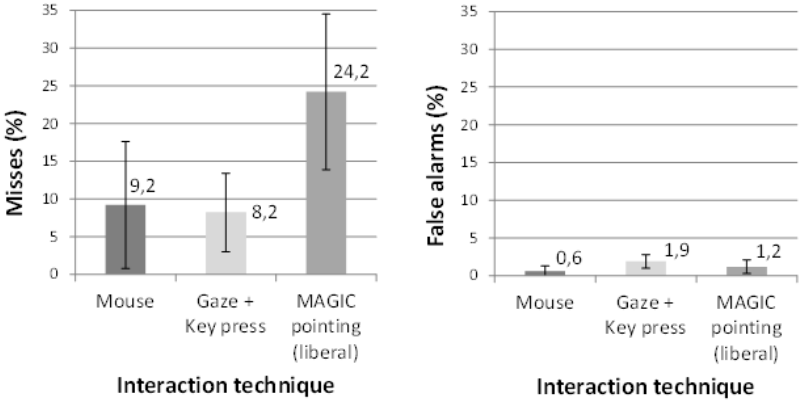


Abbildung 5.5: Verpasser und Falschalarme als Funktion der Interaktionstechnik.

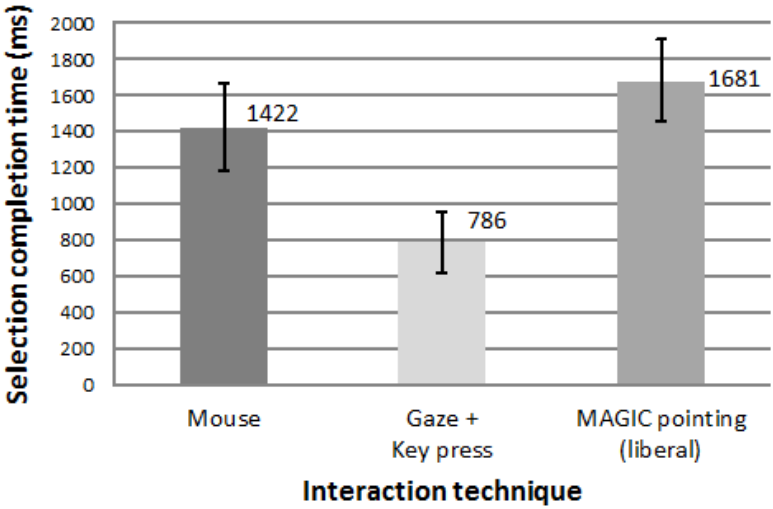


Abbildung 5.6: Selektionszeit als Funktion der Interaktionstechnik.

Abb. 5.7 zeigt die Ergebnisse der subjektiven Bewertung. MAGIC pointing erzielte für alle Merkmale mit Abstand die schlechteste Bewertung. Blick+Taste erzielte für Selektionsgeschwindigkeit, Aufgabenangemessenheit und Nutzerfreundlichkeit mit Abstand die beste Bewertung. Bei der Selektionsgenauigkeit erzielen Blick+Taste und Maus gleich gute Bewertungen. Nach ihrer bevorzugten Technik gefragt, entschieden sich 16 der 18 Versuchspersonen für Blick+Taste, die anderen 2 für die Mauseingabe.

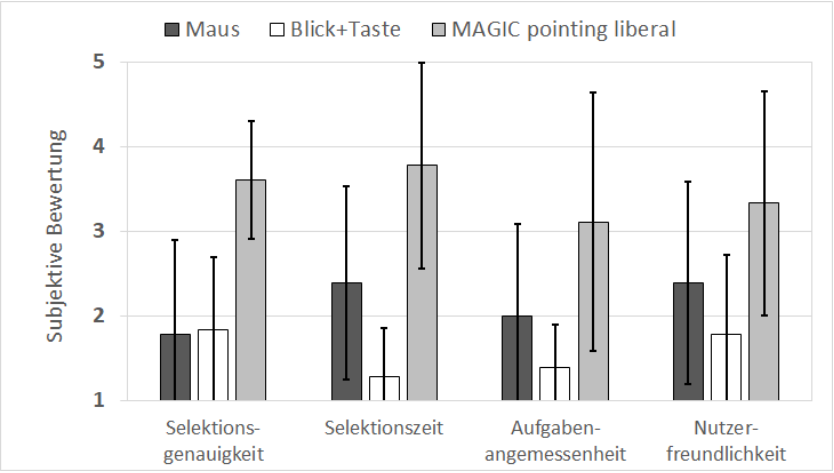


Abbildung 5.7: Subjektive Bewertung als Funktion der Interaktionstechnik. 1: sehr gute Bewertung, 5: sehr schlechte Bewertung.

Bezüglich der **Forschungsfragen** ergeben sich folgende Antworten.

Frage 1: Die Versuchspersonen erzielen im Vergleich zur Mauseingabe mit Blick+Taste eine ähnliche Effektivität (Selektionsfehlerquote), mit MAGIC pointing schneiden sie deutlich schlechter ab. Bezüglich der Effizienz (Selektionszeit) übertreffen die Versuchspersonen mit Blick+Taste die Mauseingabe erheblich; mit MAGIC pointing ist die Selektionszeit länger als mit der Mauseingabe. Die Ergebnisse zeigen, dass Blick+Taste eine Alternative zur Mauseingabe sein kann, da bei vergleichbar guter Selektionsfehlerquote deutlich schneller selektiert werden konnte.

Frage 2: Die Versuchspersonen erzielten mit Blick+Taste deutlich bessere Ergebnisse als mit MAGIC pointing.

Frage 3: Blick+Taste erzielt die beste Bewertung in allen Kategorien.

5.1.1.4 Fazit

Blick+Taste zeigte sich als sehr schnelle und effektive Methode, um bewegte Objekte zu selektieren. Zudem bekam sie die beste subjektive Bewertung und wurde von 88% der Versuchspersonen favorisiert. Die Ergebnisse der subjektiven Bewertung für die Selektionsgenauigkeit und die Selektionsgeschwindigkeit bestätigen die objektiv gemessene Selektionsfehlerquote bzw. Selektionszeit. Blick+Taste zeigt sich als leistungsfähige Alternative zur Mauseingabe.

MAGIC pointing in der liberalen Realisierung ist keine Alternative zur Mauseingabe. Ein Grund dafür mag in der in einer Querschnittstudie naturgemäß begrenzten Trainingszeit liegen. Im Vergleich zu Blick+Taste ist der Gesamtmechanismus der MAGIC pointing-Interaktionstechnik deutlich komplexer, sodass hier mutmaßlich längeres Training zu besserem Erfolg führen könnte.

Zudem beklagten einige der Versuchspersonen die permanente Sichtbarkeit des Mauszeigers. Während das bei der Mauseingabe niemand als störend erwähnte, wurde es bei MAGIC pointing als ablenkend erlebt (vgl. Visuelles Feedback der Blickposition Abschnitt 2.4.2.3). Dies leuchtet ein, da bei MAGIC pointing der Mauszeiger stets im Fixationsbereich visualisiert wird und dadurch mit den auszuwertenden Objekten um die visuelle Wahrnehmung konkurriert. Bei der Mauseingabe liegt der Mauszeiger nur nach bewusster manueller Verschiebung durch den Benutzer im Fixationsbereich. Möglicherweise hat diese Ablenkung auch die längere Selektionszeit von MAGIC pointing im Vergleich zur Mauseingabe bewirkt (vgl. die Ergebnisse zu Blick+Taste+Blickvisualisierung in Abschnitt 6.2.2.4).

5.1.2 Vergleich MAGIC pointing konservativ, MAGIC Button, Mauseingabe

Dieses Experiment und seine Ergebnisse wurden zum Großteil bei der ETRA 2014 veröffentlicht [Hil14a].

Verglichen werden drei Interaktionstechniken bei der Aufgabe der Bewegtobjektselektion: MAGIC pointing konservativ, MAGIC button und Mauseingabe.

Die *Mauseingabe* erfolgt als traditionelle Zeige-Klick-Interaktion mit Klick der linken Maustaste.

Rein grundsätzlich hat MAGIC pointing das Potenzial, genauso hohe Selektionsgenauigkeit zu erzielen wie die Mausinteraktion. Denn am Ende jeder MAGIC pointing-Interaktion steht ja eine verkürzte Mausinteraktion.

In unserer ersten Nutzerstudie (Abschnitt 5.1.1) hat MAGIC pointing in der liberalen Realisierung im direkten Vergleich mit der Mauseingabe schlechter abgeschnitten. Einige der Versuchspersonen hatten aber insbesondere die permanente Mauszeigeranzeige im Fixationsbereich (außer dann, wenn der Benutzer die Maus bewegt) als ablenkend und störend empfunden. Daher untersucht dieses Experiment MAGIC pointing-Varianten, bei denen sich der Mauszeiger weniger proaktiv, sondern mehr so wie bei der traditionellen Mauseingabe verhält.

Bei der Variante *MAGIC pointing konservativ* (wie die liberale vorgestellt von Zhai u. a. [Zha99]) springt der Mauszeiger erst dann in den aktuellen Fixationsbereich, wenn der Benutzer die Maus bewegt. Im vorliegenden Experiment wurde die Originalimplementierung an zwei Stellen verändert. Erstens wurde nicht der „intelligente Offset“ implementiert, der die Richtungsunsicherheit beim Erscheinen des Mauszeiger minimieren will [Zha99]. Stattdessen wird der Mauszeiger an die vom Eyetracker geschätzte Blickposition platziert, d.h. so nahe am Objekt wie möglich.

Zweitens wurden zwei Schwellenwerte implementiert, um Überaktivität des Mauszeigers zu verhindern für den Fall, dass der Benutzer die Maus versehentlich bewegt. Zum einen wird der Mauszeiger nur dann neu platziert, wenn die letzte Mausbewegung länger als 25 ms zurücklag. Zum anderen wird eine Blickposition erst dann als neu interpretiert und der Mauszeiger dorthin platziert, wenn die Distanz zur vorangehenden Blickposition 150 Pixel oder mehr beträgt. Auf diese Weise wird verhindert, dass der Mauszeiger alle 25 ms springt, was passiert, falls der Benutzer die Maus mit Bewegungsunterbrechung verschiebt.

Wenn zuletzt eine Sakkade detektiert wurde, wird der Distanzschwellenwert von 150 Pixel auf 20 Pixel reduziert, denn nach einer Sakkade folgt mit hoher Wahrscheinlichkeit eine Korrektursakkade, die den Blick noch näher an das Zielobjekt bringt. Auf diese Weise positioniert das System den Mauszeiger noch etwas näher an das Zielobjekt. Falls die zuletzt detektierte Augenbewegung eine Fixation repräsentiert, wird der Distanzschwellenwert wieder auf 150 Pixel gesetzt, denn jetzt ist die Wahrscheinlichkeit, dass eine Sakkade folgt, höher.

Als weitere Variante wird *MAGIC button* untersucht. Diese Variante wurde im Rahmen der vorliegenden Arbeit (inspiriert durch die Variante MAGIC Touch [Dre09]) gestaltet. MAGIC button zeichnet sich dadurch aus, dass der Benutzer das Springen des Mauszeigers in den aktuellen Fixationsbereich durch Drücken der rechten Maustaste bewusster steuert. Dadurch kann der Benutzer die Maus unbeabsichtigt bewegen, ohne dass der Mauszeiger unerwünscht springt.

Bei MAGIC button muss der Benutzer für das Springen eine separate Aktion durchführen, die nicht Bestandteil des internalisierten Zeigen-und-linke-Maustaste-drücken-Prozesses ist. Auf diese Weise gibt MAGIC button dem Benutzer bessere Kontrolle über das Zeigerspringen. Der Preis dafür ist, dass jede Selektion zwei Tastendrucke erfordert (erst rechts zum Springen, dann links zum Selektieren), wodurch die Finger mehr belastet sind.

5.1.2.1 Testaufgabe

Als visueller Stimulus für die Testaufgabe wird die 3-minütige Testaufgabe des Experiments aus Abschnitt 5.1.1 etwas abgewandelt. Abb. 5.8 zeigt den visuellen Stimulus. Die Selektionsobjekte sind hier als hellgraue, hellgelbe und hellblaue Kreise auf mittelgrauem Hintergrund gestaltet, um farblich der Realität von Überflugvideos etwas näher zu sein. Die Größe der Objekte beträgt jetzt 30 Pixel (0,71° Sehwinkel; es gibt hier keinen unsichtbar vergrößerten Selektionsbereich für die Objekte). Zum Zielobjekt, das selektiert werden muss, wird ein Objekt dann, wenn sich sein Mittelpunkt für 1 Sekunde rot färbt. Gelingt der Versuchsperson eine erfolgreiche Selektion, färbt sich das Objekt permanent grün. In allen übrigen Merkmalen wie Objektgeschwindigkeiten und Aufgabenumfang entspricht die Testaufgabe der Testaufgabe aus Abschnitt 5.1.1.

Da die Erfahrung aus dem Experiment mit MAGIC pointing liberal in Abschnitt 5.1.1 gezeigt hatte, dass MAGIC pointing schwierig zu erlernen ist, wurden zudem drei Trainingsaufgaben mit aufsteigendem Schwierigkeitsgrad gestaltet. Abb. 5.9 und Abb. 5.10 zeigen die visuellen Stimuli. Sie sind inspiriert vom in der DIN CEN ISO/TS 9241-411 vorgeschlagenen Tipptest mit mehreren Richtungen. Für unsere Zwecke wurde die Aufgabe etwas abgewandelt durch das Hinzufügen eines Startobjekts in der Mitte, mit dessen Selektion jede Einzelaufgabe (engl. *Trial*) beginnt. Jedes Trial besteht aus der Selektion eines Punktepaares: Zuerst wird das Startobjekt selektiert, dann das Zielobjekt[1].

Abb. 5.9 zeigt, wie ein Trial abläuft. Links ist die Situation vor Trialstart zu sehen: Der Startbutton in der Mitte (Seitenlänge 50 Pixel) ist rot eingefärbt und signalisiert dadurch, dass er als nächstes zu selektieren ist; die Zielobjekte (Durchmesser 30 Pixel) sind grau und unmarkiert. Sobald der Benutzer den Startbutton selektiert hat, wechselt seine Farbe zu Grau und eines der

[1] Vgl. die Realisierung von Zhang u. a. [Zha07]: Zweck des Startkreises ist, dass er sicherstellt, dass die Reihenfolge, in der die Zielobjekte zur Selektion angeboten werden, im Voraus unbekannt bleiben kann; beim Tipptest mit mehreren Richtungen aus der DIN CEN ISO/TS 9241-411 ist dies nicht gegeben.

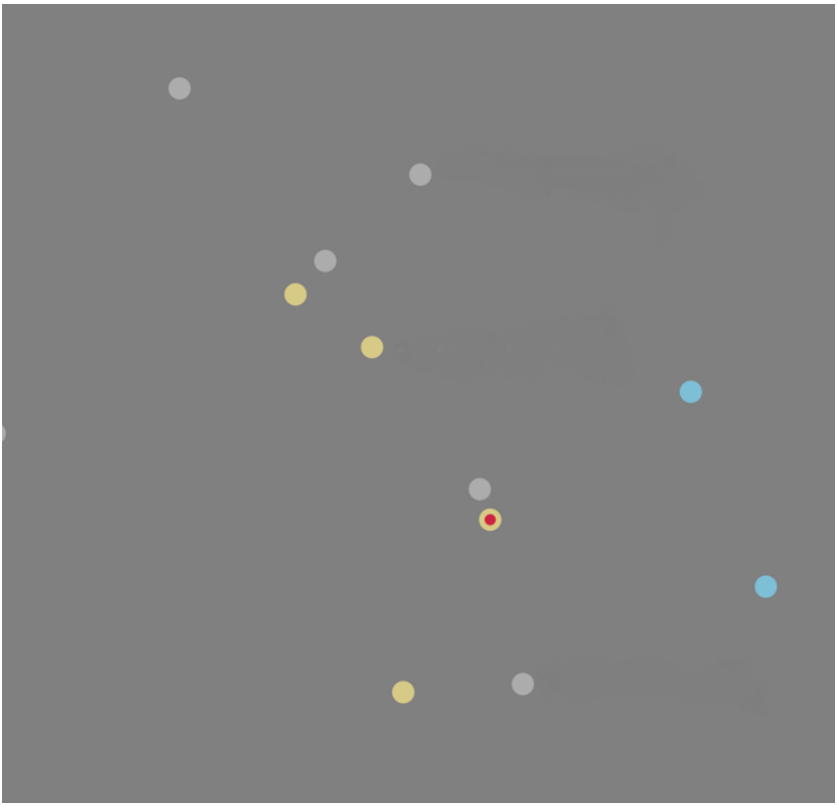


Abbildung 5.8: Visueller Stimulus der Testaufgabe.

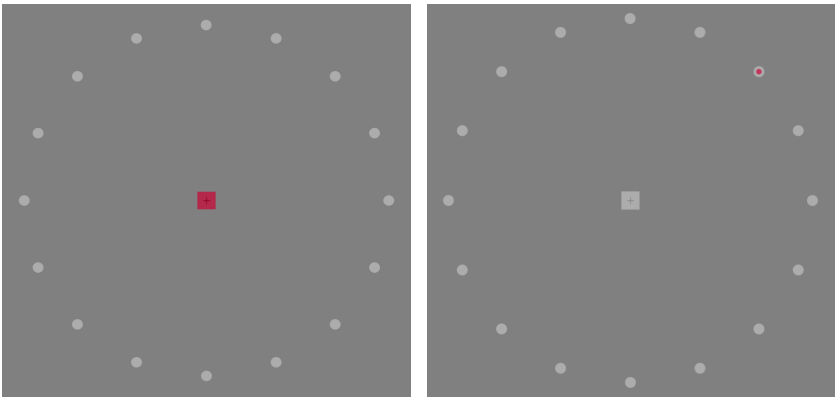
zwölf Zielobjekte wird mit einem roten Punkt markiert und muss so schnell wie möglich selektiert werden; gelingt dies innerhalb des Time-Outs von 3 Sekunden, färbt sich das Zielobjekt grün, falls nicht, färbt es sich grau und der Startbutton wird wieder rot und zeigt den Beginn des nächsten Trials an. Jedes Trial besteht also bestenfalls aus zwei Selektionen, wobei die Zeitmessung für die Selektionszeit mit der Selektion des Startbuttons beginnt und mit der Selektion des Zielobjekts endet; da der Time-Out bei 3 Sekunden liegt, sind wiederholte Selektionsversuche möglich.

Jede Trainingsaufgabe umfasst insgesamt 32 Trials, wobei jedes der 16 Zielobjekte je zweimal selektiert werden muss; die Reihenfolge ist dabei zufällig. Bei Trainingsaufgabe 1 (Abb. 5.9b) sind die Zielobjekte statisch und haben einen Abstand von 500 Pixeln vom Startbutton.

Bei Trainingsaufgabe 2 (Abb. 5.10a) bewegen sich die Zielobjekte im Abstand von 500 Pixeln im Kreis. Die Kreisbewegung beginnt, sobald der Startbutton selektiert wurde und beschreibt initial zufällig eine Links- bzw. Rechtsdrehung; die initiale Geschwindigkeit beträgt zufällig entweder 115 Pixel/s oder 135 Pixel/s (die langsamste bzw. schnellste Geschwindigkeit der Testaufgabe). Wird die Selektion des markierten Zielobjekts verpasst, so ändern sich Bewegungsrichtung und Geschwindigkeit zufällig.

Bei Trainingsaufgabe 3 (Abb. 5.10a) verläuft die Objektbewegung radial. Initial befinden sich die Objekte in einem Abstand von 250 Pixeln vom Startbutton. Sobald dieser selektiert wurde, bewegt sich jedes Objekt mit zufällig zugewiesener Geschwindigkeit von 115 Pixel/s oder 135 Pixel/s bis zu einem Abstand von maximal 500 Pixeln nach außen. Dort wechselt die Geschwindigkeit wieder zufällig und die Objekte bewegen sich bis zum Abstand von 250 Pixeln nach innen.

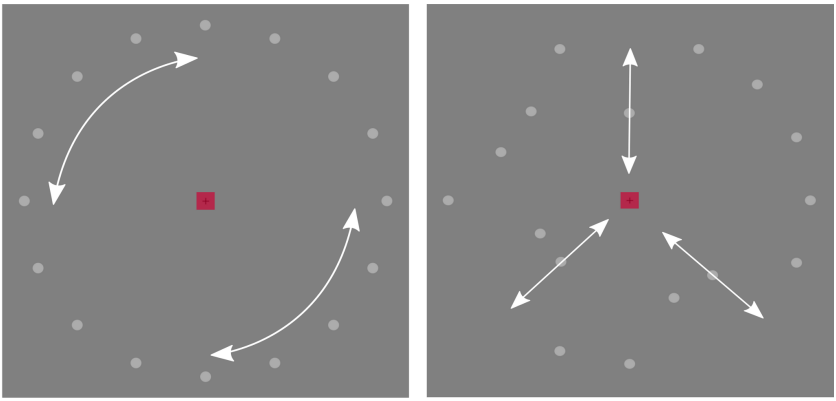
Nach Absolvieren der Trainingsaufgaben haben die Versuchspersonen 96 Selektionen mit der Interaktionstechnik trainiert. Zudem wurde die Selektionstechnik in alle Richtungen (nach oben, unten, links, rechts sowie schräg) trainiert. Dies wurde als nützlich erachtet, da die fortlaufend erforderlichen Selektionen bei der Testaufgabe keine Richtung ausschließen. Zudem trainieren die Versuchspersonen die Minimal- und Maximalgeschwindigkeit der Testaufgabe.



(a) Startbild jedes Trials: Start-Knopf rot, Objekte grau.

(b) Zielobjekt rot markiert.

Abbildung 5.9: Trainingsaufgabe 1 mit statischen Zielobjekten.



(a) Trainingsaufgabe 2: Zielobjekte mit Kreisbewegung.

(b) Trainingsaufgabe 3: Zielobjekte mit radialer Bewegung.

Abbildung 5.10: Trainingsaufgaben 2 und 3 mit bewegten Zielobjekten.

5.1.2.2 Versuchsaufbau, Versuchspersonen und Versuchsablauf

Die Versuchssoftware baut auf der JAVA-Anwendung des Experiments aus Abschnitt 5.1.1 auf (Windows 7, Trident 1-Bibliothek zur Darstellung der Animationen [Gil12]). Angezeigt werden die Trainingsaufgaben und die Testaufgabe auf einem 24-Zoll-Monitor (Auflösung 1920 x 1200 Pixel). Verwendete Eingabegeräte sind der Tobii X60 und die Comfort-Mouse 2000 for Business 3 (1000dpi) von Microsoft.

Die Mauszeigergeschwindigkeit wurde auf 8/11 unter Windows 7 gesetzt. Die Eyetracker-Rohdaten wurden gefiltert mit dem I-VT-Algorithmus [Sal00] (vgl. Abschnitt 2.4.1). Der Schwellenwert (Velocity Threshold) wurde auf $v_{max} = 50°/s$ gesetzt. Abb. 5.11 zeigt den Versuchsaufbau (vgl. a. Abschnitt 4.2).

Die 12 Versuchspersonen (11 männlich, 1 weiblich; Alter zwischen 21 und 27 Jahre, Altersdurchschnitt ± 1 Standardabweichung 23,5 ± 2,4 Jahre) verfügten alle über normale oder auf normal korrigierte Sicht, drei trugen eine Brille, zwei Kontaktlinsen. Alle waren Kollegen oder Studenten, also keine Videoauswerteexperten. Alle waren erfahrene Benutzer der Computermaus. Eine Versuchsperson hatte erste Erfahrung mit Blickinteraktion bei der Untersuchung aus Abschnitt 5.1.1 gesammelt. Andere Vorerfahrung mit Blickinteraktion gab es nicht.

Das Versuchsdesign nutzte ein vollständiges, ausbalanciertes Within-Subjects-Design. Jede Versuchsperson führte die Testaufgabe je einmal mit jeder der drei Techniken durch. Um Ermüdungs- und Lerneffekte zu kontrollieren, wurden die Versuchspersonen in 6 Versuchsgruppen mit je 2 Versuchspersonen eingeteilt, wobei jede Versuchsgruppe eine der 6 möglichen Technik-Reihenfolgen durchführte.

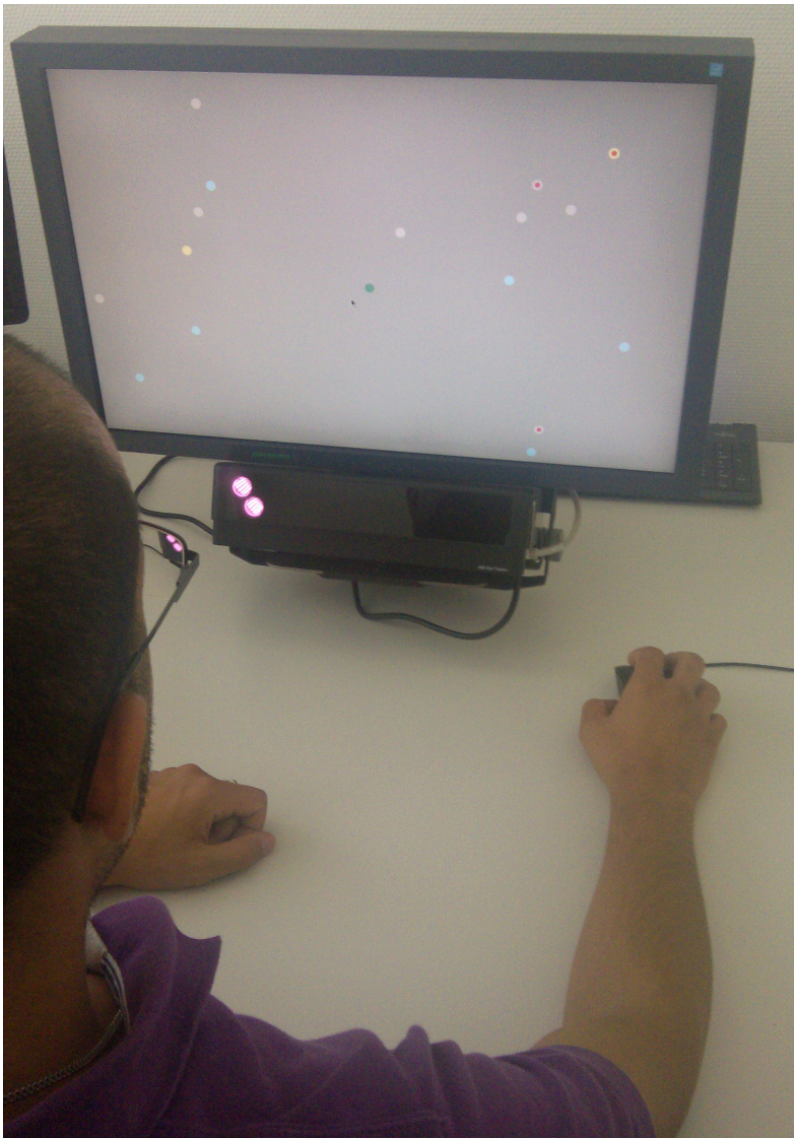


Abbildung 5.11: Versuchsaufbau.

Der Versuchsablauf war wie folgt. Als Erstes wurde in einer allgemeinen Einführung die Bewegtobjektselektion erläutert. Danach absolvierten die Versuchspersonen für jede Interaktionstechnik folgende Schritte. Zunächst wurde die zu testende Interaktionstechnik erläutert, gefolgt von der Standard-9-Punkt-Kalibrierung des Tobii X60 im Falle der beiden MACIC pointing-Techniken.

Danach absolvierten die Versuchspersonen die drei Trainingsaufgaben sowie eine kurze Version der Testaufgabe als weitere Trainingsaufgabe. Dann absolvierten die Versuchspersonen die Testaufgabe. Abschließend bewerteten die Versuchspersonen die Interaktionstechnik subjektiv mithilfe des DIN EN ISO/TS 9241-411 Standard-Fragebogens zur Bewertung der Zufriedenstellung von Benutzern mit Eingabegeräten auf einer 7-Punkte-Skala (7: beste Bewertung; 1: schlechteste Bewertung).

Da alle Versuchspersonen gänzlich unerfahren waren mit den beiden MAGIC pointing-Varianten und MAGIC pointing eine komplexe Interaktionstechnik ist, wurde das Merkmal *Erforderliche Anstrengung bei der Nutzung* in die zwei Merkmale *Erforderliche physische Anstrengung* und *Erforderliche mentale Anstrengung* aufgeteilt. Außerdem wurde das Merkmal *Ermüdung der Augen* hinzugefügt, da zwei blickbasierte Interaktionstechniken untersucht wurden. Wir folgten hierbei dem Vorbild von Zhang und MacKenzie [Zha07].

Die Versuchsdauer betrug 50 Minuten, wobei auf jede Interaktionstechnik 15 min (inklusive des intensiven Trainings) entfielen.

5.1.2.3 Forschungsfragen, Metriken und Ergebnisse

Dieses Experiment adressierte folgende Forschungsfragen:

- Frage 1: Welche Leistung erzielen die Versuchspersonen mit den MAGIC pointing-Interaktionstechniken im Vergleich zur Mauseingabe? Ist die Leistung so gut, dass sie eine Alternative zur reinen, manuellen Mausselektion sein können?
- Frage 2: Wie leistungsfähig sind die Benutzer mit den beiden unterschiedlichen MAGIC pointing-Varianten?

- Frage 3: Wie ist die subjektive Bewertung der Interaktionstechniken?

Hierfür wurden folgende Metriken verwendet:

- Selektionsfehlerquote: Sie berechnet sich als Anzahl fehlerhafter Selektionen / Gesamtaufgabenumfang (= 91 Objektselektionen).
- Selektionszeit: Sie berechnet sich als die zeitliche Differenz zwischen der Rotfärbung eines Objekts und seiner erfolgreichen Selektion.

Abb. 5.12 zeigt die Ergebnisse der Selektionsfehlerquote. Eine ANOVA mit Messwiederholung (Signifikanzniveau $\alpha = 0{,}05$) zeigt signifikante Unterschiede zwischen den Interaktionstechniken ($F(2;34) = 11{,}96$; $p<0{,}001$). Post-hoc-Analyse mit Bonferronikorrektur zeigte hochsignifikante Unterschiede ($p<0{,}001$) zwischen Maus und MAGIC konservativ sowie MAGIC button und MAGIC konservativ. Der Unterschied zwischen Maus und MAGIC button war nicht signifikant.

Abb. 5.13 zeigt die Ergebnisse für die Selektionszeit. Die Unterschiede zwischen den Techniken sind signifikant mit ($F(2;32) = 26$; $p<0{,}001$). Post-hoc-Analyse mit Bonferronikorrektur zeigt hochsignifikante Unterschiede ($p<0{,}001$) zwischen Mauseingabe und MAGIC konservativ sowie einen signifikanten Unterschied ($p<0{,}05$) zwischen MAGIC button und MAGIC konservativ. Der Unterschied zwischen Maus und MAGIC button war nicht signifikant.

Abb. 5.14 zeigt die Ergebnisse der subjektiven Bewertung. Bei einigen Merkmalen sind die Bewertungen für die drei Techniken sehr ähnlich. Insgesamt erzielt jedoch MAGIC konservativ im Mittel am häufigsten die schlechteste Bewertung. Deutlich schlechter ist sie vor allem für die Merkmale Gleichmäßigkeit bei der Nutzung, Erforderliche mentale Anstrengung, Genauigkeit, Allgemeine Zufriedenheit und Nutzung der Interaktionstechnik insgesamt. Auch bei der Ermüdung der Augen schneidet MAGIC konservativ etwas schlechter ab als die beiden anderen Techniken. Diese Bewertungen zeigen, dass MAGIC konservativ eine komplexe Interaktionstechnik ist.

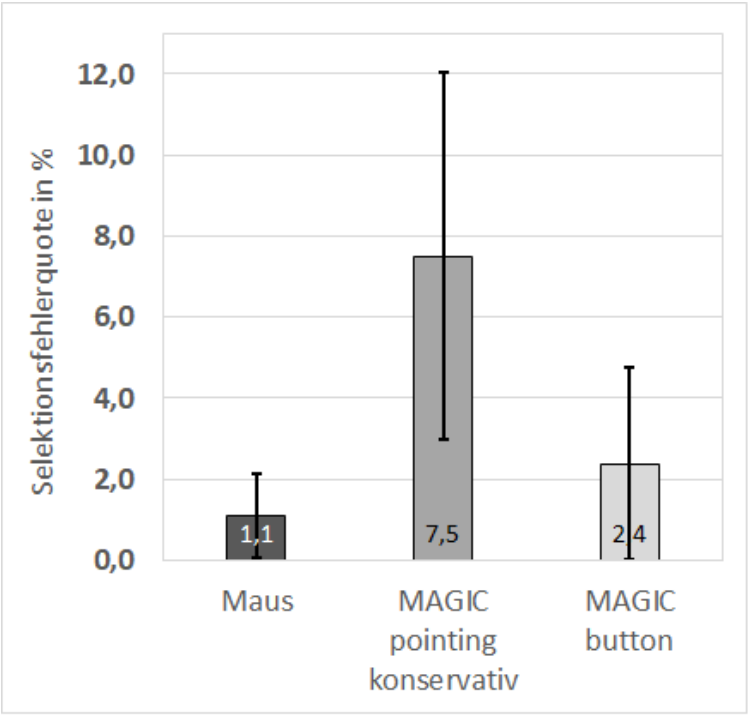


Abbildung 5.12: Selektionsfehlerquote als Funktion der Interaktionstechnik.

MAGIC button bekommt für kein Merkmal die schlechteste Bewertung, wohl aber im Mittel die beste für die Erforderliche physische Anstrengung und Allgemeine Zufriedenheit, und sie liegt bei der Benutzungsgeschwindigkeit weit vor den beiden anderen (obwohl die Mauseingabe objektiv schneller war).

Auch die Mauseingabe wird in keinem Merkmal am schlechtesten bewertet. Am besten ist sie im Mittel bei der Erforderlichen mentalen Anstrengung sowie bei der Nutzung der Interaktionstechnik insgesamt. Dieses Ergebnis mag auf die große Erfahrung der Benutzer im Vergleich zu den MAGIC pointing-Varianten zurückzuführen sein.

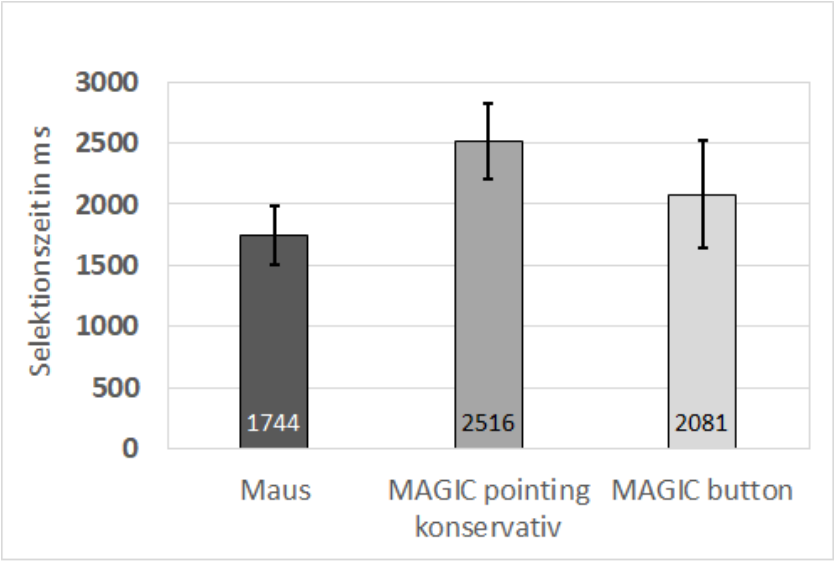


Abbildung 5.13: Selektionszeit als Funktion der Interaktionstechnik.

Nach ihrer bevorzugten Technik gefragt, entschieden sich 8 der 12 Versuchspersonen für MAGIC button, 3 für MAGIC konservativ und 1 für die Mauseingabe.

Die Befürworter von MAGIC konservativ beschrieben die Technik als sehr intuitiv und schätzten an ihr, dass sie weniger manuelle Aktion erfordert als die beiden anderen.

Die Befürworter von MAGIC button schätzten vor allem die Kontrolle über den Mauszeiger. Wenige Versuchspersonen äußerten, dass sie das viele Rechts-Links-Mausklicken irritiert habe und es auf die Dauer möglicherweise für die Finger ermüdend werden könne. Das Ergebnis der subjektiven Bewertung für Ermüdung der Finger kann in diese Richtung interpretiert werden.

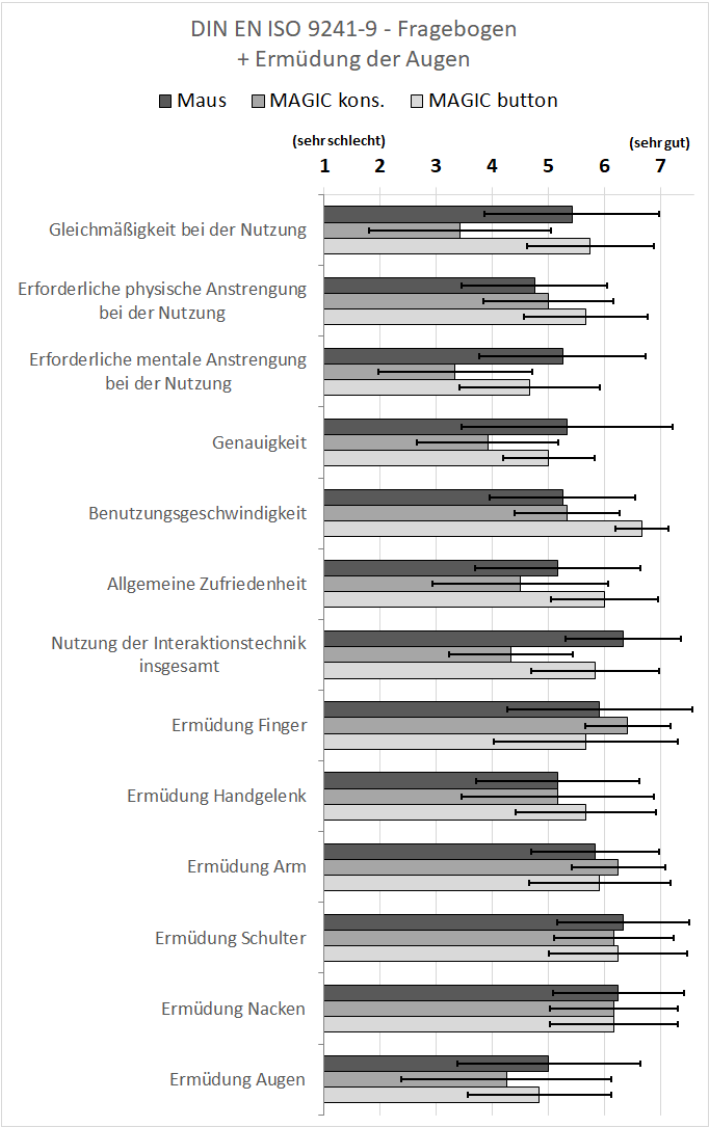


Abbildung 5.14: Subjektive Bewertung.

Interessant ist, dass die mit Abstand beste subjektive Bewertung der Benutzungsgeschwindigkeit für MAGIC button nicht mit den objektiv gemessenen Selektionszeiten übereinstimmt. Ähnliches berichten auch Zhai u. a. [Zha99] und Drewes u. a. [Dre09] in ihren MAGIC pointing-Untersuchungen: Die Benutzer fühlen sich mit einer Interaktionstechnik gut und bewerten sie daher positiv.

Bezüglich der **Forschungsfragen** ergeben sich folgende Antworten.

Frage 1: Die Versuchspersonen erzielten im Vergleich zur Mauseingabe mit MAGIC button eine vergleichbare Effektivität (Selektionsfehlerquote), schnitten mit MAGIC konservativ jedoch deutlich schlechter ab. Bezüglich der Effizienz (Selektionszeit) kommen die Versuchspersonen mit MAGIC button immerhin in die Nähe der Mauseingabe, sind mit MAGIC konservativ jedoch erheblich langsamer.

MAGIC konservativ zeigt sich nicht als Alternative zur Mauseingabe. MAGIC button mag dann eine Alternative sein, wenn der Benutzer lange und viel selektieren muss und die Belastung von Kognition und Motorik zwischendurch etwas verlagern will.

Frage 2: Die Versuchspersonen erzielten mit MAGIC button deutlich bessere Ergebnisse als mit MAGIC konservativ.

Frage 3: MAGIC button kann bezüglich der Zufriedenstellung im Vergleich zur Maus nicht nur mithalten, sondern wird in einigen wichtigen Merkmalen besser bewertet. MAGIC konservativ erzielt insgesamt die schlechteste Zufriedenstellung.

5.1.2.4 Fazit

Erstens war die Mauseingabe schneller und effektiver als die MAGIC pointing-Varianten. Allerdings wurde sie nur von einer der zwölf Versuchspersonen für die gestellte Aufgabe der Bewegtobjektselektion favorisiert. Ganz offensichtlich bevorzugten subjektiv fast alle Versuchspersonen eine Interaktionstechnik, mit der der Mauszeiger deutlich weniger manuell verschoben werden muss.

Zweitens zeigte sich, dass MAGIC pointing in der konservativen Realisierung keine Alternative zur Mauseingabe ist. Immerhin favorisierten drei Versuchspersonen diese Interaktionstechnik. Möglicherweise wäre mit mehr Training über einen längeren Zeitraum eine bessere Leistung möglich. Vergleicht man allerdings die Ergebnisse mit denen von MAGIC pointing liberal (Abschnitt 5.1.1.3 bzw. Abschnitt 5.1.1.4), so sind die Selektionsfehlerquoten im Vergleich zur Mauseingabe ähnlich schlechter. Die benötigte Selektionszeit war jedoch für MAGIC konservativ im Vergleich zur Mauseingabe noch deutlich länger als für MAGIC liberal.

Drittens zeigte sich, dass die im Rahmen dieser Arbeit neu eingeführte Interaktionstechnik MAGIC button als einzige der betrachteten MAGIC pointing-Varianten subjektiv positive Bewertungen erzielte. Betrachtet man zudem die Selektionsfehlerquoten, so sind diese im Vergleich zur Mauseingabe ähnlich gut wie die, die Blick+Taste verglichen mit der Maus erzielt hatte (vgl. Abb. 5.4). Möglicherweise könnte Training über einen längeren Zeitraum auch die Selektionszeit von MAGIC button noch verbessern. Mit Trainingsverbesserung wäre MAGIC button eine Alternative zur Mauseingabe, wenn der Benutzer die Menge an Armbewegungen reduzieren möchte und stattdessen in Kauf nimmt, eine Weile lang die Finger mehr zu belasten.

Betrachtet man die Ergebnisse der beiden Experimente, die MAGIC pointing-Varianten untersuchten (Abschnitt 5.1.1 und Abschnitt 5.1.2), so zeigte sich keine als effektiver und schneller als die Mauseingabe. Aus diesem Grund wurden im Rahmen dieser Arbeit keine weiteren Untersuchungen zu MAGIC pointing durchgeführt.

5.1.3 Vergleich Blick+Taste, Blick+Fußtaste, Mauseingabe

Dieses Experiment und seine Ergebnisse wurden bei der ETRA 2016 veröffentlicht [Hil16b].

Verglichen werden drei Interaktionstechniken bei der Aufgabe der Bewegtobjektselektion: Blick+Taste, Blick+Fußtaste und Mauseingabe.

Da Blick+Taste im initialen Experiment (Abschnitt 5.1.1) im Vergleich zur Mauseingabe gut abgeschnitten hatte, wurde ein weiteres Experiment durchgeführt, um herauszufinden, ob sich das Ergebnis bestätigen ließe. Zusätzlich wurde mit Blick+Fußtaste eine Tasten-Alternative untersucht, die zu einer blickbasierten Interaktionstechnik führt, die keinerlei manuelle Aktion des Benutzers erfordert.

Die *Mauseingabe* erfolgt als traditionelle Zeige-Klick-Interaktion mit Klick der linken Maustaste.

Die Selektion mit der *Blick+Taste*-Interaktion erfolgt, indem der Benutzer das Objekt anblickt und währenddessen die ENTER-Taste des NumPad auf der Computertastatur drückt.

Die Selektion mit der *Blick+Fußtaste*-Interaktion erfolgt, indem der Benutzer das Objekt anblickt und währenddessen die Fußtaste (USB Foot Switch 2 Single, vgl. Abschnitt 4.3) drückt.

5.1.3.1 Testaufgabe

Die Testaufgabe wurde als kontrolliertes Szenario gestaltet, das ein abstraktes Trial-Versuchsparadigma nutzt, bei dem Punktepaare selektiert werden müssen (vgl. Tipptests aus [DIN14] bzw. Abwandlung mit drei Richtungen horizontal, vertikal, diagonal von [Zha99]). Jedes Trial besteht aus einem statischen Startkreis und einem bewegten Zielobjekt. Abb. 5.15 zeigt beispielhaft den Ablauf eines Trials.

1. Der Startkreis wird angezeigt (Abb. 5.15a).
2. Der Benutzer selektiert den Startkreis. Sobald er auf den Startkreis zeigt, wechselt dieser seine visuelle Gestaltung (hellgrau mit grünem Mittelpunkt, Abb. 5.15b, vgl. a. Abschnitt 2.4.2.3 zu visuellem Feedback für das Zeigen mit Blick). Der Selektionszeitpunkt des Startkreises bestimmt den Startzeitpunkt des Trials.

3. Sobald der Benutzer den Startkreis erfolgreich selektiert hat, wird dieser nicht mehr angezeigt. Gleichzeitig wird das Zielobjekt angezeigt (Abb. 5.15c).
4. Der Benutzer selektiert das Zielobjekt. Sobald er auf das Zielobjekt zeigt, wechselt dieses seine visuelle Gestaltung (hellgrau mit grünem Mittelpunkt, Abb. 5.15d). Der Selektionszeitpunkt des Zielobjektes bestimmt den Endzeitpunkt des Trials.

Die sichtbare Größe von Startkreisen und Zielobjekten betrug 100 Pixel (2,7 cm bzw. ca. 2,38° bei 65 cm Abstand der Augen vom Monitor). Die selektierbare Größe betrug 200 Pixel (5,4 cm bzw. 4,75°).

Bei der Gestaltung der Trials wurden fünf Variablen variiert. Dies diente dazu, möglichst viele denkbare Selektionssituationen zu berücksichtigen. Die variierten Faktoren und ihre Stufen waren wie folgt:

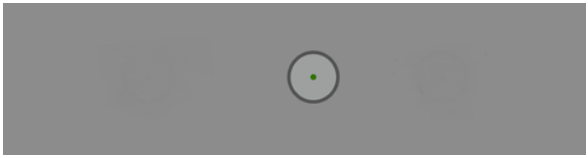
- Position des Startkreises (4 Stufen): Oben links, oben rechts, unten links, unten rechts (Abb. 5.16).
- Startposition des Zielobjekts und Bewegungsrichtung relativ zur Startkreisposition (2 Stufen): Startposition nah mit Bewegung von der Startkreisposition weg, Startposition entfernt mit Bewegung zur Startkreisposition hin (Abb. 5.17).
- Bewegungsrichtung des Zielobjekts (3 Stufen): horizontal, diagonal, vertikal (Abb. 5.17).
- Bewegungsgeschwindigkeit des Zielobjekts (5 Stufen): 200 Pixel/s, 350 Pixel/s, 500 Pixel/s, 650 Pixel/s, 800 Pixel/s (entspricht ca. 4,8°/s, 8,3°/s, 11,9°/s, 15,5°/s, 19°/s bzw. 5,4 cm/s, 9,5 cm/s, 13,5 cm/s, 17,6 cm/s, 21,6 cm/s).
- Bewegungsmuster des Zielobjekts (2 Stufen): Linear, wellenförmig (mit Amplitude von 20 Pixeln orthogonal zur Hauptbewegungsrichtung).

Insgesamt ergibt sich so eine Anzahl Trials von $4 * 3 * 2 * 5 * 3 = 240$.

(a) Startbild jedes Trials: Der Startkreis wird angezeigt, grau mit dunkelgrauem Mittelpunkt.

(b) Der Benutzer zeigt auf den Startkreis (visuelle Rückmeldung durch Wechsel der Farbgebung zu hellgrau mit grünem Punkt) und selektiert ihn.

(c) Mit erfolgreicher Selektion des Startkreises verschwindet dieser vom Bildschirm. Gleichzeitig wird das Zielobjekt angezeigt und beginnt unmittelbar mit seiner Bewegung. Der weiße Pfeil dient der Veranschaulichung der Bewegungsrichtung, er wird dem Benutzer nicht anzeigt.



(d) Der Benutzer zeigt auf das bewegte Zielobjekt (visuelle Rückmeldung durch Wechsel der Farbgebung zu hellgrau mit grünem Punkt) und selektiert es. Bei erfolgreicher Selektion, verschwindet das Zielobjekt von der Anzeige und ein neuer Startkreis erscheint (Abb. 5.15a) und zeigt den Beginn des nächsten Trials an.

Abbildung 5.15: Ablauf eines Trials.

Die gewählten (konstanten) Geschwindigkeiten wurden zum einen inspiriert durch die Geschwindigkeiten, die in Videodatenmaterial zu erwarten sind. Abb. 5.3 zeigte, dass Geschwindigkeiten von bis zu 500 Pixel/s (20 Pixel Bild-zu-Bild-Distanz des Objektmittelpunkts bei 25 Hz Bildrate) häufig vorkommen und gelegentlich deutlich überschritten werden (vgl. S. 135). Die genaue Wahl der höheren Geschwindigkeiten wurde in Anlehnung an Hasan u. a. [Has11] getroffen, die 500, 650 und 800 Pixel/s untersuchten. 350 und 200 Pixel/s wurden hinzugefügt, um äquidistante Unterschiede der Geschwindigkeiten zu erhalten.

Die Versuchsgestaltung stellt sicher, dass die Bewegtobjektselektionen nur innerhalb des rechteckigen Bereichs stattfinden, den die vier Startkreispositionen aufspannen. Abb. 5.17 zeigt beispielhaft für den Startkreis links oben die zugehörigen sechs Startpositionen der Zielobjekte. Für die anderen Startkreise werden die Startpositionen der Zielobjekte entsprechend gespiegelt.

Zielobjekte, die nah beim Startkreis starten, bewegen sich 400 Pixel weit, bevor sie vom Monitor verschwinden. Zielobjekte, die entfernt starten, bewegen sich 600 Pixel weit. Je nach Bewegungsgeschwindigkeit sind die Zielobjekte unterschiedlich lange sichtbar und selektierbar. Die kürzeste Zeitdauer liegt bei 0,5 s und gilt für nah startende Zielobjekte (Pfadlänge 400 Pixel) und eine Geschwindigkeit von 800 Pixel/s. Die längste Zeitdauer liegt bei 3 s und gilt für entfernt startende Zielobjekte (Pfadlänge 600 Pixel) und eine Geschwindigkeit von 200 Pixel/s.

Die Präsentation der 240 Selektionsbedingungen ist randomisiert. Gelingt einer Versuchsperson eine Selektion nicht, so wird ein Zielobjekt mit denselben Charakteristika erneut präsentiert. Die Anzahl Versuche pro Zielobjekt ist jedoch auf drei begrenzt. Die höchste Anzahl Selektionen, die eine Versuchsperson mit einer Interaktionstechnik durchführt, liegt also bei $3*240 = 720$ Trials.

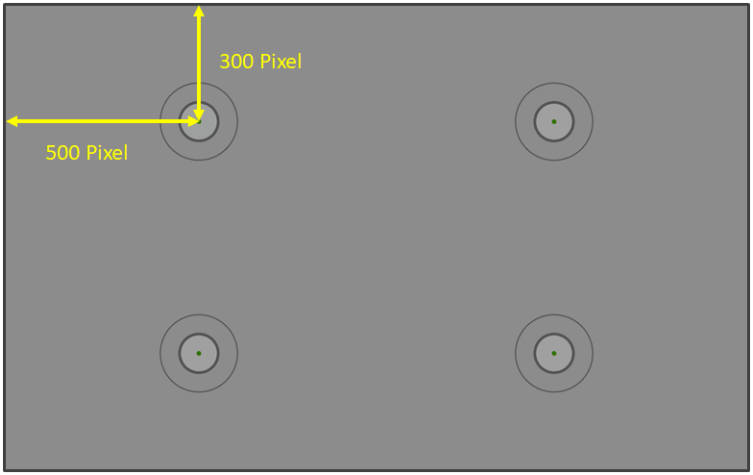


Abbildung 5.16: Die vier Positionen des Startkreises.

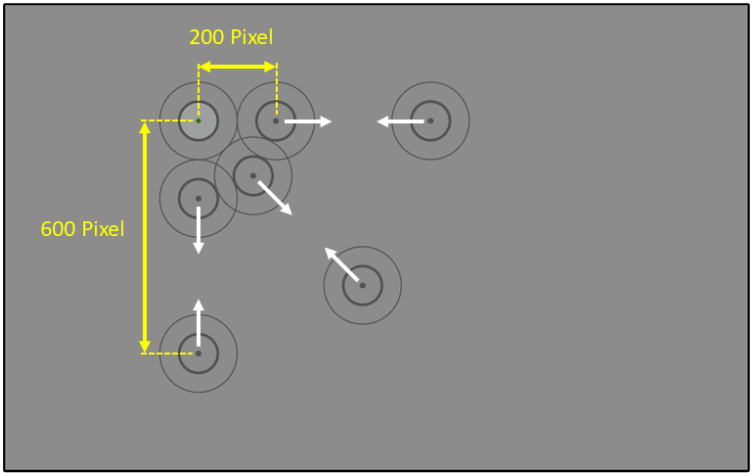


Abbildung 5.17: Startkreisposition links oben mit allen Zielobjekt-Startposition und Bewegungsrichtungen.

5.1.3.2 Versuchsaufbau, Versuchspersonen und Versuchsablauf

Die Versuchssoftware zur Präsentation der Testaufgaben wurde als JAVA-Anwendung unter Windows 7 (64bit) implementiert. Die Animationen wurden mithilfe der auf der Programmiersprache JAVA aufbauenden Software „Processing“ implementiert [Pet15]. Der Versuchsaufbau ist wie in Abb. 4.3.

Zur Blickerfassung wird der Tobii X60 verwendet (vgl. Abschnitt 4.2). Die Blickrohdaten werden gefiltert mit dem Real-Time Saccade Detection and Fixation Smoothing-Algorithmus [Kum08] (vgl. Abschnitt 2.4.1). Der Schwellenwert S zur Trennung von Fixationen und Sakkaden wurde experimentell vor dem Experiment bestimmt und beträgt 15 Pixel für Blick+Taste und 20 Pixel für Blick-Fußtaste. Da dieser Algorithmus eine zusätzliche Latenz von 1 Blickdatensample bewirkt, beträgt die Latenz, mit der das Blicksignal geliefert wird, jetzt 33 ms (X60) + 17 ms (Blickfilterung) = 50 ms.

Der manuelle Tastendruck bei Blick+Taste erfolgte mit der ENTER-Taste des NumPad einer gewöhnlichen Computertastatur. Als Fußtaste wurde ein USB Foot Switch 2 Single genutzt (vgl. Abschnitt 4.3). Die Mauseingabe erfolgte mit einer Comfort-Mouse 2000 for Business 3 (1000 dpi) von Microsoft. Die Mauszeigergeschwindigkeit wurde auf 8/11 unter Windows 7 gesetzt.

Die 12 Versuchspersonen (9 männlich, 3 weiblich; Alter zwischen 21 und 32, Altersdurchschnitt ± 1 Standardabweichung 23,8 ± 3,0 Jahre) verfügten alle über normale oder auf normal korrigierte Sicht, zwei trugen eine Brille. Alle waren Studenten, also keine Videoauswerteexperten. Alle waren erfahrene Benutzer der Computermaus. Sechs Versuchspersonen hatten bereits einmal an einem Experiment mit Blick+Taste teilgenommen, darüber hinaus jedoch keine Erfahrung mit Eyetracking. Keine der Versuchspersonen hatte Erfahrung mit Blick+Fußtaste.

Das Versuchsdesign nutzte ein vollständiges, ausbalanciertes Within-Subjects-Design. Jede Versuchsperson führte die Testaufgabe je einmal mit jeder der drei Techniken durch. Um Ermüdungs- und Lerneffekte zu kontrollieren,

wurden die Versuchspersonen in 6 Versuchsgruppen mit je 2 Versuchspersonen eingeteilt, wobei jede Versuchsgruppe eine der 6 möglichen Technik-Reihenfolgen durchführte. Jede der 6 Versuchsgruppen umfasste eine Versuchsperson mit Erfahrung in Blick+Taste und eine ohne diese Erfahrung.

Der Versuchsablauf war wie folgt. Als Erstes wurde in einer allgemeinen Einführung die Testaufgabe erläutert. Danach wurde die Standard-9-Punkt-Kalibrierung des Tobii X60 durchgeführt. Die Kalibrierung wurde jeweils unmittelbar vor den blickbasierten Techniken wiederholt. Vor den Testaufgaben konnten die Probanden mit einer Untermenge der Testaufgaben trainieren und sich mit der jeweiligen Interaktionstechnik vertraut machen. Die Versuchspersonen wurden instruiert, so schnell und so genau wie möglich zu selektieren. Sie wurden zudem instruiert, bei jeder Selektion das Zeigegerät, d.h. die Blickposition bzw. den Mauszeiger, auf die kleinen Punkte im Zentrum der Startkreise bzw. der Zielobjekte zu richten.

5.1.3.3 Forschungsfragen, Metriken und Ergebnisse

Dieses Experiment adressierte folgende Forschungsfragen:

- Frage 1: Welche Leistung erzielen die Versuchspersonen mit den blickbasierten Interaktionstechniken im Vergleich zur Mauseingabe? Ist die Leistung so gut, dass sie eine Alternative zur Mauseingabe sein können?
- Frage 2: Wie leistungsfähig sind die Benutzer mit den beiden unterschiedlichen blickbasierten Techniken?
- Frage 3: Wie wirkt sich insbesondere die Geschwindigkeit der Objekte auf die Leistung aus?

Hierfür wurden folgende Metriken verwendet:

- Selektionstrefferquote: Sie berechnet sich als Anzahl erfolgreicher Selektionen / Anzahl durchgeführter Selektionen; die Anzahl durchgeführter Selektionen beträgt zwischen 240 und 3∗240 = 720.

- Selektionsgenauigkeit: Sie berechnet sich als Abstand (euklidische Distanz) zwischen Selektionsposition und Zentrum des Zielobjekts.
- Selektionszeit: Sie berechnet sich als Zeitdauer zwischen dem Selektionszeitpunkt des Startkreises und dem Selektionszeitpunkt des Zielobjekts.

Abb. 5.18 zeigt die Ergebnisse der *Selektionstrefferquote* für alle zehn Kombinationen aus Bewegungsgeschwindigkeit und Bewegungsmuster („L" steht für linear, „W" für wellenförmig). Die Ergebnisse sind für alle drei Techniken sehr gut für Geschwindigkeiten von 200 Pixel/s und 350 Pixel/s. Sie nehmen für alle Techniken ab Geschwindigkeit 650 Pixel/s stark ab. Während bei 800 Pixel/s mit Blick+Taste noch 30% (Linear) bzw. 20% (Wellenförmig) der Zielobjekte selektiert wurden, gelingt kaum noch eine Selektion mit den beiden anderen Techniken.

Eine ANOVA (Signifikanzniveau $\alpha = 0{,}05$) zeigte statistisch signifikante Unterschiede für die Geschwindigkeiten von 650 Pixel/s und 800 Pixel/s sowie für die Bedingung W500. Eine Post-hoc-Analyse mit Bonferroni-Korrektur zeigte signifikante Unterschiede für Blick+Taste und Mauseingabe ($p<0{,}01$ für L650 und L800; $p<0{,}05$ für W500, W650 und W800).

Abb. 5.19 zeigt die Ergebnisse der *Selektionsgenauigkeit*. Je kleiner der Wert, desto näher war die Selektion am Zielobjektzentrum. Man kann erkennen, dass die Selektionsgenauigkeit mit Erhöhung der Geschwindigkeit schlechter wird. Die Ergebnisse für die Geschwindigkeiten bis 500 Pixel/s sind ähnlich für alle drei Techniken. Für die beiden schnellsten Geschwindigkeiten ist das Ergebnis mit der Mauseingabe geringfügig besser im Vergleich mit den blickbasierten Techniken. Der größte Unterschied der Mittelwerte ist mit 10 Pixel der zwischen Mauseingabe und Blick+Fußtaste; dies entspricht bei einer Pixelgröße von 0,27 mm einem Unterschied von 2,7 mm.

Betrachtet man die erzielten Werte für Mittelwerte und Standardabweichungen, so sieht man, dass für die blickbasierten Interaktionstechniken bei den hohen Geschwindigkeiten die selektierbare Größe der Zielobjekte von 200 Pixel ausgenutzt wurde, denn die einfache Standardabweichung erreicht bis zu 90 Pixel. Die sichtbare Größe von 100 Pixel genügte nur für die langsamste

Geschwindigkeit von 200 Pixel/s; hier lag die Distanz zum Objektmittelpunkt bei den meisten Trials unter 50 Pixel. Die Selektionsgenauigkeit verschlechtert sich mit zunehmender Geschwindigkeit in etwa linear.

Die Analyse der *Selektionszeit* mit einem zweiseitigen *t*-Test (Signifikanzniveau $\alpha = 0{,}05$) ergab statistisch signifikante Unterschiede bei den Ergebnissen für nah bzw. entfernt vom Startkreis startende Zielobjekte für sehr viele der zehn Bedingungen kombiniert aus Bewegungsmuster und Bewegungsgeschwindigkeit (Unterschiede bei Mauseingabe: $p<0{,}001$ für alle zehn Bedingungen; Blick+Taste: $p<0{,}05$ für L650, $p<0{,}001$ für L800; Blick+Fußtaste: $p<0{,}05$ für W500; $p<0{,}01$ für L200, L650, W650, $p<0{,}001$ für L800, W800). Daher wurden die beiden Gruppen separat analysiert.

Abb. 5.20 zeigt oben die Ergebnisse für nah startende Zielobjekte, unten die für entfernt startende. Beiden gemeinsam ist, dass die Mauseingabe stets die längste Selektionszeit benötigt. Tabelle 5.1 zeigt die statistisch signifikanten Unterschiede (nach ANOVA mit Bonferroni-Korrektur).

Man sieht, dass Blick+Taste bei allen Bedingungen die kürzeste Selektionszeit erzielt. Die Mauseingabe benötigt bei allen Bedingungen die längste Selektionszeit. Im Mittel erzielt Blick+Taste Zeiten zwischen 459 ms und 563 ms, Blick+Fußtaste zwischen 442 ms und 636 ms, Mauseingabe zwischen 479 ms und 771 ms. Es ist zu beachten, dass die sehr kurzen Zeiten der beiden schnellsten Geschwindigkeiten auf einer erheblich geringeren Selektionstrefferquote beruhen als die Zeiten der drei anderen Geschwindigkeiten.

Betrachtet man nur die Geschwindigkeiten 200, 350 und 500 Pixel/s, so ist die Selektionszeit bei Blick+Taste für diese drei ähnlich und liegt im Mittel bei 540 ms. Für Blick+Fußtaste ist die Selektionszeit ebenfalls ähnlich und liegt im Mittel stets um 50 ms höher als Blick+Taste; Blick+Fußtaste kommt also auf 590 ms im Mittel. Bei der Mauseingabe sind deutliche Unterschiede zu erkennen für die nah bzw. entfernt startenden Zielobjekte. Für die nahen werden im Mittel ca. 620 ms benötigt, was kaum schlechter ist als die blickbasierten Techniken. Für die entfernten werden im Mittel ca. 720 ms benötigt, was im Vergleich zu Blick+Taste immerhin 180 ms oder 33% längere Selektionszeit bedeutet.

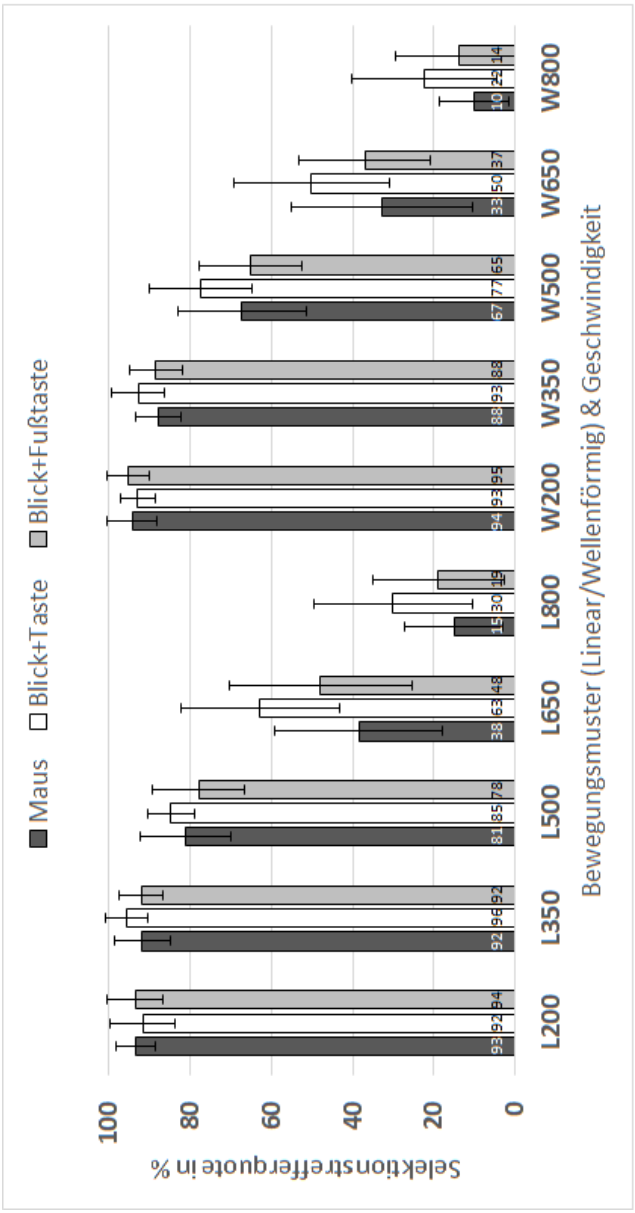


Abbildung 5.18: Selektionstrefferquote als Funktion von Bewegungsmuster und Geschwindigkeit für die drei Interaktionstechniken als Mittelwerte ± 1 Standardabweichung.

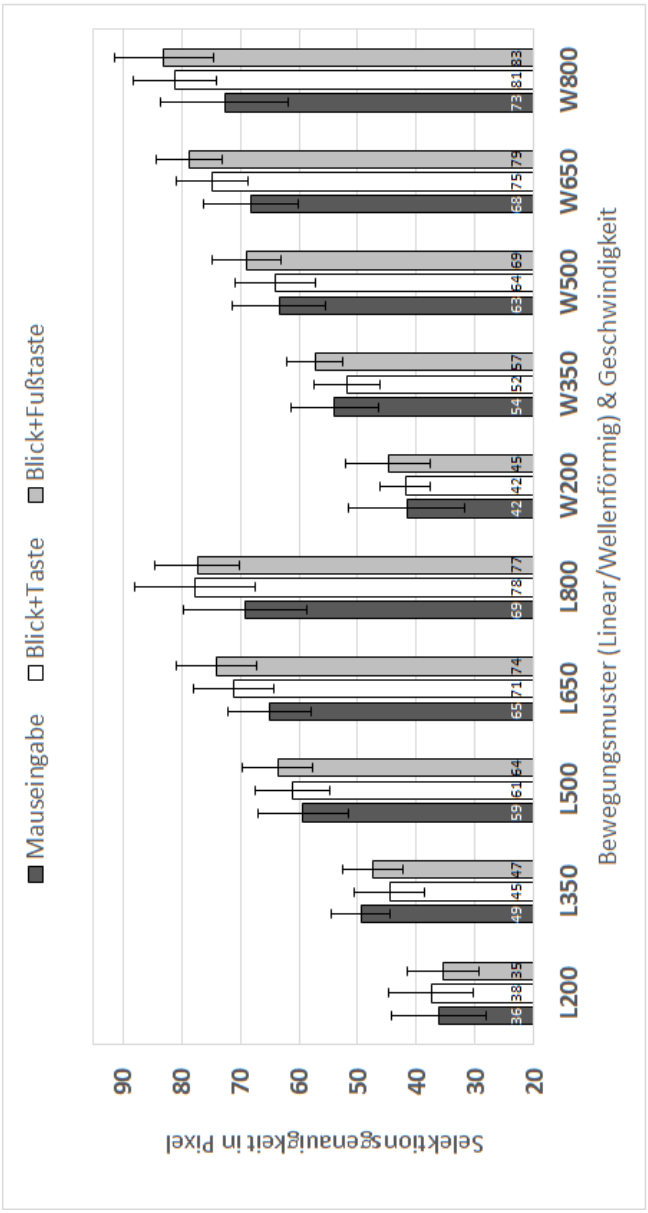


Abbildung 5.19: Selektionsgenauigkeit als Funktion von Bewegungsmuster und Geschwindigkeit für die drei Interaktionstechniken als Mittelwerte ± 1 Standardabweichung.

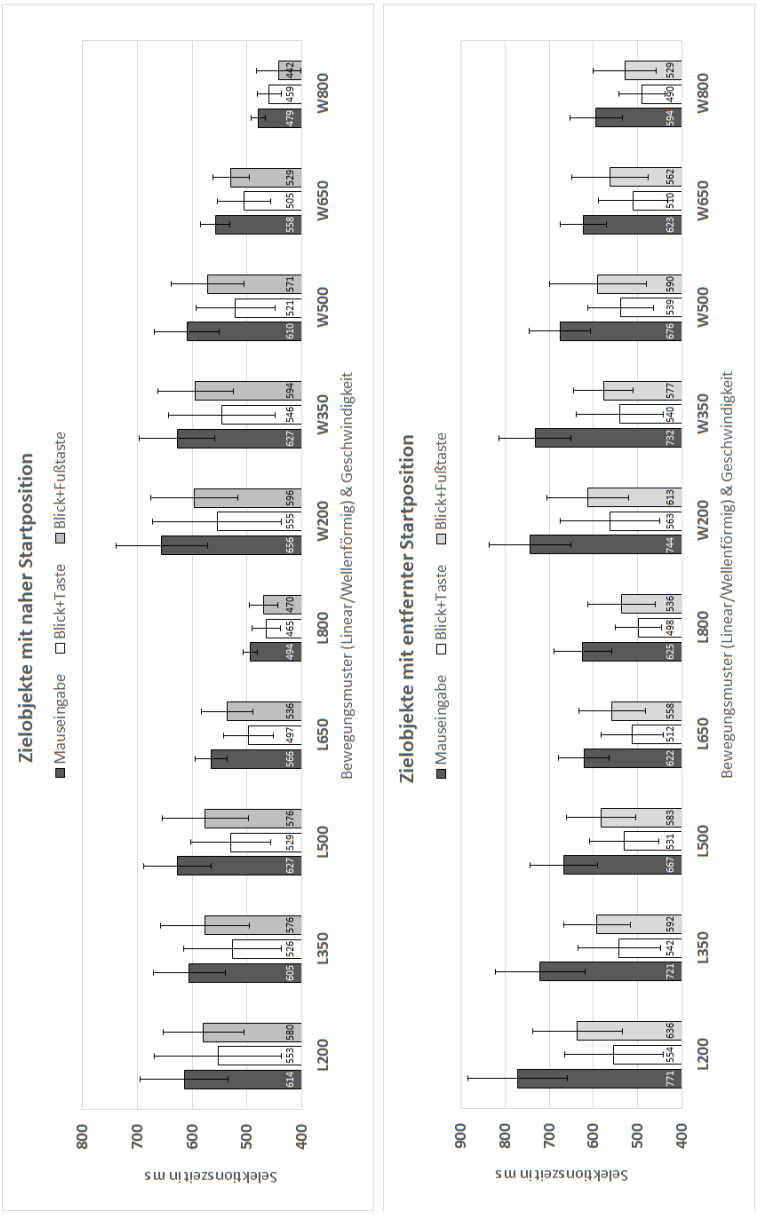


Abbildung 5.20: Selektionszeit als Funktion von Bewegungsmuster und Geschwindigkeit für die drei Interaktionstechniken als Mittelwerte ± 1 Standardabweichung.

Tabelle 5.1: Statistisch signifikante Unterschiede für **Selektionszeit**, separat für nah und entfernt startende Zielobjekte. M steht für Mauseingabe, BT für Blick+Taste, BFT für Blick+Fußtaste.

Nah	**L200**	**L350**	**L500**	**L650**	**L800**
M, BT	$p<0{,}001$				
M, BFT	-	-	$p<0{,}001$		$p<0{,}05$
BT, BFT	-	$p<0{,}01$	$p<0{,}05$	$p<0{,}01$	-
	W200	**W350**	**W500**	**W650**	**W800**
M, BT	$p<0{,}001$				
BT, BFT	-		$p<0{,}05$	-	
M, BFT	$p<0{,}01$	-	$p<0{,}01$	$p<0{,}05$	$p<0{,}01$
BT, BFT	-				
Entfernt	**L200**	**L350**	**L500**	**L650**	**L800**
M, BT	$p<0{,}001$				
M, BFT	$p<0{,}001$				
BT, BFT	$p<0{,}001$	$p<0{,}01$	$p<0{,}001$	$p<0{,}01$	
	W200	**W350**	**W500**	**W650**	**W800**
M, BT	$p<0{,}001$				
M, BFT	$p<0{,}001$				$p<0{,}01$
BT, BFT	$p<0{,}01$	$p<0{,}05$	$p<0{,}001$		$p<0{,}05$

Bezüglich der **Forschungsfragen** ergeben sich folgende Antworten.

Frage 1: Die Versuchspersonen erzielen mit den blickbasierten Interaktionstechniken eine so gute Leistung, dass sie eine Alternative zur Mauseingabe sein können. Die Effektivität (Selektionstrefferquote) ist für Geschwindigkeiten bis 500 Pixel/s für alle drei Interaktionstechniken ähnlich gut. Bei den sehr hohen Geschwindigkeiten (650 und 800 Pixel/s) sowie bei 500 Pixel/s

und wellenförmiger Objektbewegung schneidet Blick+Taste erheblich besser ab als die beiden anderen Interaktionstechniken.

Auch die Selektionsgenauigkeit ist für Geschwindigkeiten bis 500 Pixel/s für alle drei Interaktionstechniken ähnlich gut. Bei höheren Geschwindigkeiten ist die Genauigkeit mit der Computermaus etwas besser als mit den blickbasierten Techniken; die Verbesserung beträgt (im Mittel) maximal 10 Pixel, was 2,7 mm entspricht.

Die Selektionszeit ist am besten mit Blick+Taste (im Mittel 540 ms), dicht gefolgt von Blick+Fußtaste (im Mittel 590 ms). Die Computermaus ist für nah startende Zielobjekte kaum schlechter (im Mittel 620 ms), für entfernt startende mit (im Mittel) 720 ms jedoch um 33% langsamer als Blick+Taste.

Frage 2: Blick+Taste erzielt insbesondere bei der Selektionstrefferquote bei hoher Selektionsschwierigkeit (sehr schnelle Geschwindigkeiten 650 und 800 Pixel/s sowie 500 Pixel/s mit wellenförmiger Objektbewegung) deutlich bessere Ergebnisse als Blick+Fußtaste. Die Selektionszeit ist zudem durchgehend um 50 ms kürzer. Bezüglich aller anderen Ergebnisse erzielen die beiden blickbasierten Interaktionstechniken ähnliche Ergebnisse.

Frage 3: Für alle Interaktionstechniken hat die Geschwindigkeit der Objekte erheblichen Einfluss auf die Selektionstrefferquote und auf die Selektionsgenauigkeit: Je höher die Geschwindigkeit, desto schlechter das Ergebnis.

5.1.3.4 Fazit

Auch in diesem Experiment zeigte sich Blick+Taste als leistungsfähigste Interaktionstechnik. Die Selektionstrefferquote ist für alle Geschwindigkeiten entweder gleich gut wie die der Mauseingabe bzw. erheblich besser bei schnellen Geschwindigkeiten. Die Selektionszeit liegt mit Blick+Taste im Mittel bei 540 ms für alle Versuchsbedingungen. Je nach Startposition der Zielobjektbewegung ist Blick+Taste damit 13% (nah startende Zielobjekte) bzw. 25% (entfernt startende Zielobjekte) schneller als die Mauseingabe.

Die Selektionsgenauigkeit ist für höhere Geschwindigkeiten geringfügig besser mit der Mauseingabe. Am größten ist der Unterschied bei 800 Pixel/s (19°/s

bzw. 21,6 cm/s), wo Mauseingabe (im Mittel) um 2,7 mm (10 Pixel bzw. 0,24°) gegenüber Blick+Fußtaste und um 2,2 mm (8 Pixel bzw. 0,19°) gegenüber Blick+Taste genauer ist. Die Ergebnisse zeigen zudem, dass mit zunehmender Geschwindigkeit die Selektionsgenauigkeit bei allen Interaktionstechniken abnimmt.

Blick+Fußtaste lieferte im Vergleich zu Blick+Taste eine kaum langsamere Selektionszeit und ähnliche Selektionsgenauigkeit. Mit zunehmender Geschwindigkeit ist die Selektionstrefferquote jedoch deutlich geringer. Dies mag darauf zurückzuführen sein, dass die Interaktion mit Fußtaste im Interaktionsalltag der Versuchspersonen am Computer nicht vorkommt. Im Gegensatz dazu sind Benutzer daran gewöhnt, eine Computertastatur zu benutzen, wenn auch im Zusammenhang mit Texteingabe und nicht mit Selektionsoperationen. Bei längerer Nutzung einer Fußtaste könnte es zudem zu Ermüdungserscheinungen im Fuß kommen. Dies müsste in einem längeren Experiment gezielt untersucht werden.

Blick+Taste zeigte sich hier wie im initialen Experiment (vgl. Abschnitt 5.1.1) als leistungsfähige Alternative zur Mauseingabe für die Bewegtobjektselektion. Aus diesem Grund wurde Blick+Taste als Eingabetechnik der Wahl in allen weiteren Untersuchungen betrachtet. Zum einen betrifft dies eine Längsschnittstudie an abstrakten Testaufgaben (Abschnitt 5.1.4) sowie die Untersuchungen der Bewegtobjektselektion in Bildfolgen von Überflugvideos (Abschnitte 6.1 und 6.2)

5.1.4 Längsschnittstudie Blick+Taste

Die vorangehenden Untersuchungen zeigten das Potenzial von Blick+Taste als leistungsfähige Alternative zur Mauseingabe für die Bewegtobjektselektion in Querschnittstudien. Ziel der Längsschnittstudie war herauszufinden, wie sich die Leistung möglicherweise noch verbessert, wenn Blick+Taste über längere Zeit regelmäßig trainiert wird.

4 Versuchspersonen trainierten über 6 Monate hinweg pro Woche je zweimal eine Stunde Blick+Taste.

Die Selektion mit der *Blick+Taste*-Interaktion erfolgt, indem der Benutzer das Objekt anblickt und währenddessen die ENTER-Taste des NumPad auf der Computertastatur drückt.

Ein Leistungstest fand vor Beginn der Trainingsperiode sowie nach 1, 2, 3 und 6 Monaten Training statt.

Um einen Vergleich zur Mauseingabe ziehen zu können, wurde der Leistungstest vor Beginn der Trainingsphase und nach 6 Monaten auch für die Mauseingabe durchgeführt. Die *Mauseingabe* erfolgte als traditionelle Zeige-Klick-Interaktion mit Klick der linken Maustaste.

Um das Training abwechslungsreich zu gestalten, wurden 4 Versuchsaufgaben entworfen. Sie dienten nicht nur zum Training, sondern auch als Testaufgaben für die Leistungstests.

5.1.4.1 Versuchsaufgabe 1: „Circle Select“

Versuchsaufgabe 1 entsprach im Wesentlichen der Testaufgabe aus der Nutzerstudie aus Abschnitt 5.1.3. Unterschiedlich war zum Ersten die visuelle Gestaltung der Objekte. Zum einen war die sichtbare Objektgröße 50 Pixel (entspricht 1,19° bzw. 1,35 mm in unserer Versuchsanordnung). Zum anderen war die Objektfarbe grün (RGB 39, 174, 96), der Objektmittelpunkt schwarz und der Objekthintergrund dunkelgrau (RGB 51, 51, 51) (Abb. 5.21). Denn dies bietet einen guten Kontrast und lässt die Augen wenig ermüden [Gut15].

Zum Zweiten wurden nur vier Objekt-Geschwindigkeiten betrachtet: 200 Pixel/s, 350 Pixel/s, 500 Pixel/s und 650 Pixel/s. Die Gesamtzahl an Trials betrug demzufolge $4 * 2 * 3 * 2 * 4 = 192$. Die Trials wurden in randomisierter Reihenfolge präsentiert, wobei alle Versuchspersonen in jeder Trainings- bzw. Testsession dieselbe Randomisierung vorgelegt bekamen.

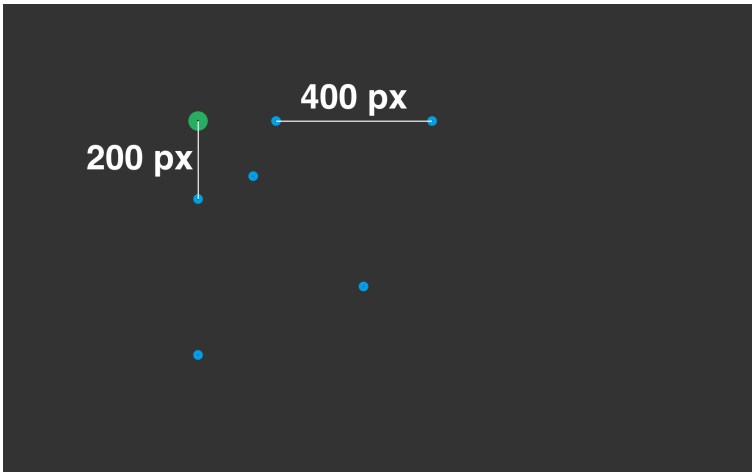


Abbildung 5.21: Versuchsaufgabe 1 der Längsschnittstudie. Startkreisposition links oben mit allen zugehörigen Zielobjekt-Startpositionen, vgl. a. Abb. 5.17.

5.1.4.2 Versuchsaufgabe 2: „Find Unique Shape“

Bei dieser Versuchsaufgabe musste das Zielobjekt nicht nur selektiert, sondern zuvor in einer 3 x 3-Anordnung von Objekten als Zielobjekt identifiziert werden. Solche kombinierten Such- und Selektionsaufgaben mit Distraktoren sind typisch für die Bildfolgenanalyse, aber auch für die MMI allgemein [Bie10].

Auch diese Versuchsaufgabe folgt einem Trial-Paradigma, bei dem Punktepaare selektiert werden müssen. Abb. 5.22 zeigt ein Trial: Zu Beginn wird das Startquadrat angezeigt (grünes Quadrat mit 60 Pixel Seitenlänge, selektierbare Größe 200 Pixel Seitenlänge). Sobald das Startquadrat selektiert wird, verschwindet es vom Bildschirm und die 3 x 3-Objektanordnung erscheint und beginnt sich sofort zu bewegen. Die Objekte sind 5 Sekunden lang sichtbar, bewegen sich linear und mit einer konstanten Geschwindigkeit von 78 Pixel/s (1,86°/ bzw. 2,1 cm/s) und legen eine Distanz von 390 Pixel zurück.

Abbildung 5.22: Versuchsaufgabe 2 der Längsschnittstudie: Beispielhaftes Trial mit Startquadrat (linkes Bild) und Such- und Selektionsaufgabe (rechtes Bild). Zielobjekt ist der rote Kreis links unten.

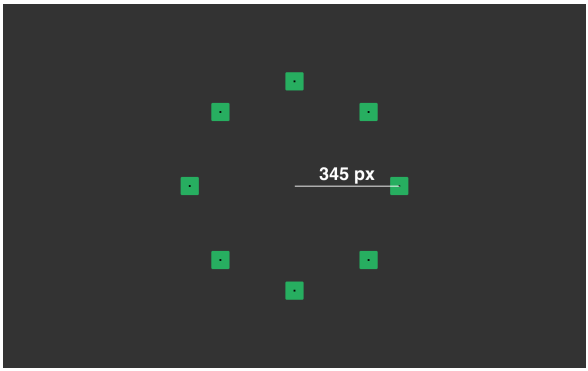


Abbildung 5.23: Versuchsaufgabe 2 der Längsschnittstudie: Startquadratpositionen.

Alle Objekte in der 3 x 3-Anordnung kombinieren in ihrer Erscheinung eine der Formen Kreis oder Quadrat mit einer der Farben Grün oder Rot. Zielobjekt ist das Objekt, dessen Kombination nur genau einmal vorkommt. Die Distanz zwischen den Mittelpunkten benachbarter Objekte beträgt 200 Pixel, sodass jede Selektionsposition eindeutig einem Objekt zugeordnet werden kann.

Abb. 5.23 zeigt die verwendeten 8 Position des (statischen) Startquadrats. Der Aufgabenpool enthält bei Beginn einer Session (Training wie Test) jede der 8 Positionen je 10-mal. In den zugehörigen Trials bewegt sich die 3 x 3-Anordnung je 5-mal vom Startquadrat weg und 5-mal auf das Startquadrat zu. Bei diagonaler Bewegung (für Startquadratpositionen auf 2, 4, 8 und 10 Uhr) wird die 3 x 3-Anordnung um 45° rotiert.

5.1.4.3 Versuchsaufgabe 3: „Catch 'em All“

Versuchsaufgabe 3 simuliert eine Bewegtobjektselektions-Situation wie sie bei einer Überwachungsaufgabe vorkommen kann, wenn sich viele Objekte in unterschiedliche Richtungen bewegen. Die bewegten Zielobjekte kann man sich als stilisierte Gruppe von Menschen vorstellen, die alle in unterschiedliche Richtungen weglaufen.

Wie bei Versuchsaufgabe 1 folgt Versuchsaufgabe 3 einem Trialparadigma ohne Distraktoren. Im Vergleich zu Versuchsaufgabe 1 sind jetzt jedoch zwischen 1 und 18 bewegte Zielobjekte zu selektieren. Abb. 5.24 zeigt beispielhaft die Versuchsaufgabe mit 18 Zielobjekten. Bei Trialbeginn wird im Bildschirmzentrum das Startquadrat angezeigt. Sobald es selektiert wurde, verschwindet es vom Bildschirm und die Zielobjekte erscheinen und beginnen unmittelbar mit ihrer Bewegung. Die Bewegung erfolgt für alle linear und radial nach außen mit einer konstanten Geschwindigkeit von 100 Pixel/s (2,38°/s bzw. 2,7 cm/s).

Für die Aufgabenumfänge von 1 bis 6 Zielobjekten sind alle Zielobjekte initial im inneren konzentrischen Kreis positioniert; für Aufgabenumfänge von 7 bis 18 Zielobjekten werden zwei konzentrische Kreise als Startpositionen genutzt. Ein Trial endet, sobald keine Zielobjekte mehr zu sehen sind, entweder weil alle selektiert wurden oder weil sie den Bildschirm verlassen haben. Die 18 Trials werden in einer Trainings- oder Testsession je 4-mal wiederholt, sodass pro Session 72 Trials zu absolvieren sind. Die Anzahl zu selektierender Objekte beträgt insgesamt also $\frac{18*19}{2} * 4 = 171 * 4 = 684$.

5.1.4.4 Versuchsaufgabe 4: „Air Traffic Control“

Versuchsaufgabe 4 ist inspiriert von der Überwachungsaufgabe eines Fluglotsen, der Flugzeuge auf einem Monitor selektiert, etwa um sich Informationen zu den Flugcharakteristika anzeigen zu lassen [Alo13].

Versuchsaufgabe 4 wurde mit zwei Zielobjektgrößen implementiert. Die große Variante (G) nutzt die Zielobjektgröße der anderen Versuchsaufgaben mit 50 Pixel sichtbarem Durchmesser und 200 Pixel selektierbarer Größe (vgl. Abb. 5.25). Die kleine Variante (K) nutzt einen sichtbaren Durchmesser von 15 Pixeln (0,36° bzw. 0,41 cm) und 60 Pixeln (1,43° bzw. 1,62 cm) selektierbarer Größe (vgl. Abb. 5.26).

Abbildung 5.24: Versuchsaufgabe 3 der Längsschnittstudie. Beispielhaftes Trial mit Startquadrat (linkes Bild) und Aufgabenumfang von 18 Zielobjekten (rechtes Bild).

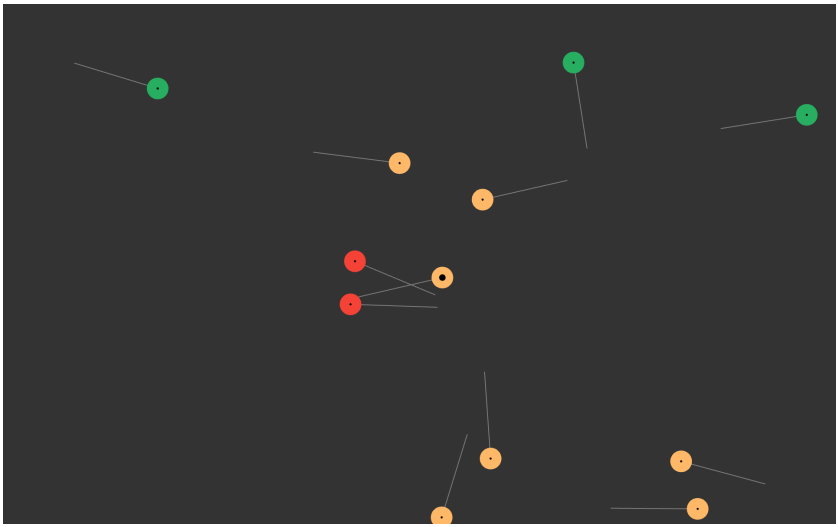


Abbildung 5.25: Versuchsaufgabe 4 der Längsschnittstudie, Variante mit großen Objekten.

Diese Aufgabe folgt nicht dem Trial-Paradigma, sondern es bewegen sich über 6 Minuten lang Objekte vom Bildschirmrand in den Bildschirm hinein, folgen einer linearen Bewegungstrajektorie und verlassen den Bildschirm auf der gegenüberliegenden Bildschirmseite. Ein Objekt ist zwischen 12 und 22 Sekunden lang auf dem Monitor zu sehen. Aufgrund der Bildschirmabmessungen resultiert dies in einer minimalen Geschwindigkeit von 1300 Pixel/22 s = 59 Pixel/s (1,40°/s bzw. 1,59 cm/s) und einer maximalen Geschwindigkeit von 2350 Pixel/12 s = 196 Pixel/s (4,66°/s bzw. 5,3 cm/s).

Jedes Objekt kann einen von drei Zuständen annehmen: „sicher“ (grün), „gefährdet“ (orange) oder „verloren“ (rot). Zielobjekte sind orange. Objekte sind ausschließlich dann selektierbar, wenn sie im Zustand orange sind.

Ein Objekt ändert seine Farbe von grün zu orange, wenn die Distanz zu einem anderen unter 350 Pixel (Variante G) bzw. 260 Pixel (Variante K) fällt. Wird es

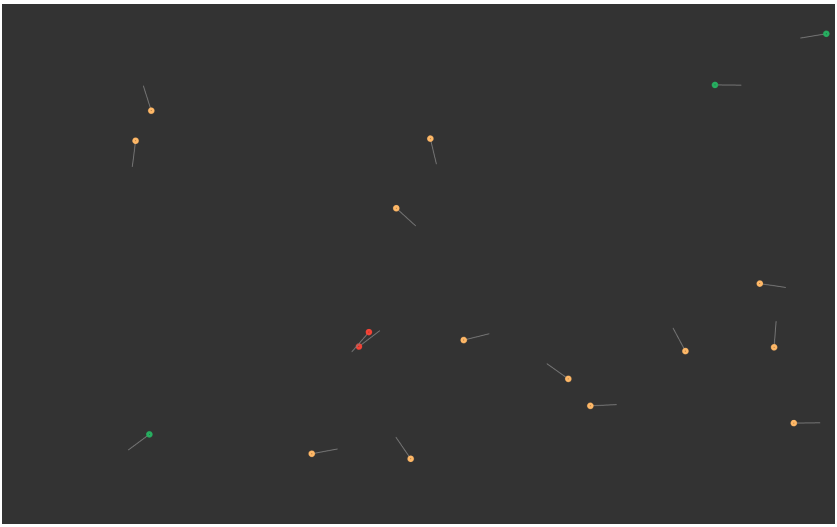


Abbildung 5.26: Versuchsaufgabe 4 der Längsschnittstudie, Variante mit kleinen Objekten.

selektiert, verschwindet es vom Bildschirm. Kommen sich zwei Objekt 100 Pixel (Variante G) bzw. 60 Pixel (Variante K) nah, so ändern sie ihre Farbe zu rot und verschwinden bei der kurz darauf folgenden Kollision vom Bildschirm.

Zu Beginn der Versuchsaufgabe tritt ein Objekt in den Bildschirm ein, alle 2 Sekunden folgt ein weiteres, solange bis 12 (Variante G) oder 18 (Variante K) Objekte zu sehen sind. Für jedes Objekt, das selektiert wird, den Bildschirm verlässt oder mit einem anderen kollidiert, tritt ein neues Objekt in den Bildschirm ein.

Aufgrund der erlaubten Nähe der Objekte können Selektionen mehrdeutig sein. In diesem Fall wird die Selektion auf das Objekt angewandt, dessen Mittelpunkt näher bei der Selektionsposition liegt. Da in vielen Situationen mehrere Optionen für die Selektion oranger Objekte gegeben ist, verändert sich dasselbe Startszenario individuell. Das bedeutet, dass die Versuchspersonen niemals genau dieselbe Aufgabe durchführen, sondern Zielobjekte an verschiedenen Orten auftreten.

5.1.4.5 Versuchsaufbau, Versuchspersonen und Versuchsablauf

Die Versuchssoftware zur Präsentation der Testaufgaben wurde als JAVA-Anwendung unter Windows 7 implementiert. Die Animationen wurden mithilfe der auf der Programmiersprache JAVA aufbauenden Software „Processing“ implementiert [Gut15]. Der Versuchsaufbau ist wie in Abb. 4.3.

Zur Blickerfassung wird der Tobii X60 verwendet (vgl. Abschnitt 4.2). Die Blickrohdaten werden gefiltert mit dem Real-Time Saccade Detection and Fixation Smoothing-Algorithmus [Kum08] (vgl. Abschnitt 2.4.1). Der Schwellenwert zur Trennung von Fixationen und Sakkaden wurde auf die empfohlenen 20 Pixel gesetzt. Da dieser Algorithmus eine zusätzliche Latenz von 1 Blickdatensample bewirkt, beträgt die Latenz, mit der das Blicksignal geliefert wird, jetzt 33 ms (X60) + 17 ms (Blickfilterung) = 50 ms.

Der manuelle Tastendruck bei Blick+Taste erfolgte mit der ENTER-Taste des NumPad einer gewöhnlichen Computertastatur. Die Mauseingabe erfolgte mit einer Comfort-Mouse 2000 for Business 3 (1000 dpi) von Microsoft. Die Mauszeigergeschwindigkeit wurde auf 8/11 unter Windows 7 gesetzt.

Die 4 Versuchspersonen (2 männlich, 2 weiblich; Altersdurchschnitt ± 1 Standardabweichung 22 ± 1,8 Jahre), alle mit normaler Sicht (keine Sehhilfen) waren Bachelor-Studenten[1], also keine Videoauswerteexperten. Alle waren erfahrene Benutzer der Computermaus und ohne jede Erfahrung mit Eyetracking. Sie wurden pro Stunde mit 12 Euro entlohnt.

Das Training für Blick+Taste erfolgte über 6 Monate hinweg pro Woche mit zweimal einer Stunde. Dabei wurden stets alle 4 Versuchsaufgaben trainiert.

Die Versuchspersonen absolvierten zwei Vergleichstests mit Blick+Taste und Mauseingabe, ebenfalls anhand der 4 Versuchsaufgaben. Einer davon fand vor Beginn der Trainingsperiode statt. Er sollte eine Baseline für Blick+Taste liefern für Benutzer ohne jede Erfahrung damit. Der zweite Vergleichstest fand nach Beendigung der Trainingsperiode nach 6 Monaten statt. Dazwischen

[1] Studienfach-Hintergrund in Wirtschaftsingenieurwesen, Informatik, Mathematik bzw. Pädagogik.

fanden Tests ausschließlich für Blick+Taste nach 1, 2 und 3 Monaten Training statt.

Alle Test- und Trainingssessions nutzten unterschiedlich randomisierte Versuchsaufgaben, sodass die Reihenfolge der Trials stets unterschiedlich war. Bei den Trainingssessions durften die Versuchspersonen die Reihenfolge, in der sie die Versuchsaufgaben trainierten, selbst wählen. Bei den Tests war die Reihenfolge festgelegt auf Versuchsaufgabe 1, 2, 3 und abschließend 4.

Der Versuchsablauf für die Tests war wie folgt. Als erstes wurde die Standard-9-Punkt-Kalibrierung des Tobii X60 durchgeführt. Die Versuchspersonen wurden instruiert, so schnell und so genau wie möglich zu selektieren. Dann absolvierten die sie die vier Versuchsaufgaben. Beim Baseline-Test und beim finalen Test wurde ein vollständiges, ausbalanciertes Within-Subjects-Design angewendet. Jede Versuchsperson führte die Versuchsaufgaben je einmal mit Blick+Taste und mit Mauseingabe durch. Je 2 Versuchspersonen begannen mit Blick+Taste bzw. mit Mauseingabe. Am Ende des Tests mit einer Interaktionstechnik bewerteten sie diese subjektiv mit dem Fragebogen zur Einzelbewertung der Norm DIN CEN ISO/TS 9241-411:2014 (vgl. [DIN14], Anhang C.1) auf einer 7-Punkte-Skala (7: beste Bewertung; 1: schlechteste Bewertung). Die Merkmale wurden ergänzt um Ermüdung der Augen wie vorgeschlagen von Zhang u. a. [Zha07].

5.1.4.6 Forschungsfragen, Metriken und Ergebnisse

Dieses Experiment adressierte folgende Forschungsfragen:

- Frage 1: Welche Leistung erzielen die Versuchspersonen mit Blick+Taste mit zunehmendem Training?
- Frage 2: Wie ist die Leistung im Vergleich zur Mauseingabe?
- Frage 3: Wie ist die subjektive Bewertung der Interaktionstechniken?

Metriken hierfür waren Selektionsfehlerquote und Selektionszeit für Versuchsaufgabe 1 bis 3, für Versuchsaufgabe 4 wurde nur die Effektivität als Anzahl Kollisionen bestimmt (je weniger desto besser).

Tabelle 5.2 zeigt für **Versuchsaufgabe 1 (Circle Select)** die Ergebnisse der *Selektionsfehlerquoten*. Bei der Analyse nahmen wir an, dass langsamere Geschwindigkeiten einfacher zu selektieren sind als schnellere und wandten daher einseitige *t*-Tests (Signifikanzniveau $\alpha = 0{,}05$) auf alle Paare von Geschwindigkeiten für Blick+Taste an. Dabei zeigten sich für die Mehrzahl der Vergleiche signifikante Unterschiede. Daher werden die Ergebnisse für alle Geschwindigkeiten separat aufgeführt. Tatsächlich sind die Selektionsfehlerquoten signifikant unterschiedlich für alle Geschwindigkeiten beim Baseline-Test (Spalte „0 [0 h]“). Bei den anderen vier Tests sind die Unterschiede nur noch zwischen der Geschwindigkeit 650 Pixel/s und allen anderen Geschwindigkeiten signifikant.

Bei Mauseingabe sind die Unterschiede beim Baseline-Test signifikant für alle Paare von Geschwindigkeiten außer zwischen 200 und 350 Pixel/s. Beim finalen Test sind wie bei Blick+Taste die Unterschiede nur noch zwischen der Geschwindigkeit 650 Pixel/s und allen anderen Geschwindigkeiten signifikant.

Die Ergebnisse zeigen einen Lerneffekt für alle Geschwindigkeiten, sowohl für Blick+Taste als auch für Mauseingabe. In der Annahme, dass die Ergebnisse mit zunehmender Trainingsdauer besser werden, wurden für alle Geschwindigkeiten einseitige *t*-Test zum Vergleich der Ergebnisse des Baseline-Tests und des finalen Tests durchgeführt. Für Blick+Taste war die Verbesserung bei 200 Pixel/s nicht signifikant, für die anderen war sie signifikant (für 350 Pixel/s mit $p<0{,}05$, für 500 und 650 Pixel/s mit $p<0{,}01$). Die Verbesserung war bereits nach einem Monat Training signifikant ($p<0{,}05$ für 350 und 650 Pixel/s; $p<0{,}01$ für 500 Pixel/s). Für 650 Pixel/s verbesserte sich das Ergebnis mit jedem weiteren Monat Training. Mit Mauseingabe war die Verbesserung signifikant für 500 und 650 Pixel/s ($p<0{,}05$).

Vergleicht man die Ergebnisse für Blick+Taste und Mauseingabe, so sind die Ergebnisse sowohl beim Baseline-Test als auch beim finalen Test ähnlich für alle Geschwindigkeiten. Beim Baseline-Test ist das Ergebnis für Blick+Taste bei 500 Pixel/s im Mittel deutlich besser, aber der Unterschied zur Mauseingabe ist nicht signifikant (zweiseitiger *t*-Test mit $p=0{,}332$); für 650 Pixel/s ist die Mauseingabe im Mittel deutlich besser, aber auch dieser Unterschied ist nicht signifikant (zweiseitiger *t*-Test mit $p=0{,}239$).

Tabelle 5.2: Selektionsfehlerquote in Prozent als Mittelwert (± 1 Standardabweichung) für Versuchsaufgabe 1 „Circle Select". IT steht für Interaktionstechnik, BT für Blick+Taste, M für Mauseingabe.

1: Circle Select		Trainings-Monate [-Stunden h] bis zum Testzeitpunkt				
Bedingung		**0 [0 h]**	**1 [9 h]**	**2 [18 h]**	**3 [27 h]**	**6 [52 h]**
IT	Pixel/s					
BT	200	5,2 (6,5)	2,6 (2,6)	2,6 (2,0)	3.1 (4,0)	2,6 (2,0)
	350	9,9 (4,3)	1,6 (2,0)	3,1 (1,2)	2,6 (3,9)	3,1 (2,1)
	500	19,3 (7,3)	5,2 (4,0)	9,4 (5,5)	5,7 (4,6)	6,8 (3,1)
	650	57,3 (4,0)	38,0 (7,1)	33,9 (13,9)	26,4 (8,3)	24,0 (7,8)
M	200	3,4 (4,0)	-	-	-	0,0 (0,0)
	350	7,1 (9,4)	-	-	-	1,5 (1,8)
	500	28,0 (17,5)	-	-	-	9,6 (11,3)
	650	45,2 (17,0)	-	-	-	24,5 (14,2)

Tabelle 5.3 zeigt die Ergebnisse der *Selektionszeiten* für Versuchsaufgabe 1. Die Ergebnisse für Blick+Taste waren ähnlich für alle Bedingungen, es fanden sich keine signifikanten Unterschiede. Das Training hatte offenkundig keinen Einfluss auf die Selektionszeit, die gemittelt über alle Versuchspersonen bei 474 ± 35 ms lag. Eine Betrachtung der individuellen Resultate enthüllte kein Muster; die geringen Unterschiede mögen von der Tagesform herrühren.

Für die Mauseingabe ist die Selektionszeit beim Baseline-Test für 650 Pixel/s signifikant schneller als für alle anderen, beim finalen Test unterscheiden sich alle Geschwindigkeiten signifikant. Vergleicht man die Ergebnisse des Baseline-Tests mit denen des finalen Tests, so ergibt sich eine signifikante Verbesserung für die Geschwindigkeiten 350 und 500 Pixel/s (p<0,05).

Vergleicht man Blick+Taste und Mauseingabe, so ist Blick+Taste signifikant schneller für alle Geschwindigkeiten beim Baseline-Test sowie für 200, 350 und 500 Pixel/s beim finalen Test.

Tabelle 5.3: Selektionszeit in ms als Mittelwert (± 1 Standardabw.) für Versuchsaufgabe 1 „Circle Select". IT steht für Interaktionstechnik, BT für Blick+Taste, M für Mauseingabe.

1: Circle Select		Trainings-Monate [-Stunden h] bis zum Testzeitpunkt				
Bedingung		**0 [0 h]**	**1 [9 h]**	**2 [18 h]**	**3 [27 h]**	**6 [52 h]**
IT	Pixel/s					
BT	200	490 (107)	497 (48)	469 (22)	467 (20)	497 (51)
	350	489 (102)	486 (43)	456 (23)	466 (24)	497 (54)
	500	475 (89)	477 (41)	475 (21)	464 (27)	489 (54)
	650	465 (78)	479 (39)	450 (27)	468 (26)	489 (56)
M	200	714 (75)	-	-	-	616 (14)
	350	691 (53)	-	-	-	588 (11)
	500	645 (47)	-	-	-	567 (11)
	650	591 (49)	-	-	-	548 (4)

Für **Versuchsaufgabe 2 (Find Unique Shape)** lag die *Selektionsfehlerquote* für Blick+Taste bereits beim Baseline-Test nahe 0%, was in allen folgenden Tests so blieb. Für die Mauseingabe betrug sie im Baseline-Test 4,0 ± 4,5 % und 0% beim finalen Test.

Tabelle 5.4 zeigt die *Selektionszeit*. Für Blick+Taste wird sie über die Trainingsdauer erheblich kürzer. Signifikant schneller waren die Versuchspersonen bereits nach 2 Monaten Training (p<0,01), danach ist die Verbesserung nur noch gering. Betrachtet man die Ergebnisse der Mauseingabe, so ergibt sich ebenfalls eine signifikante Verbesserung (p<0,001). Der Lerneffekt, der durch das Training mit Blick+Taste erzielt wurde, scheint sich auch auf die Schnelligkeit der Mausinteraktion auszuwirken.

Vergleicht man Blick+Taste und Mauseingabe, so sind die Selektionszeiten vor der Trainingsperiode im Mittel um 200 ms schneller für Blick+Taste, nach 6 Monaten Training nur noch um 100 ms. Die Unterschiede sind aber nicht signifikant (p=0,125 beim Baseline-Test, p=0,146 beim finalen Test).

Tabelle 5.4: Selektionszeit in ms als Mittelwert (± 1 Standardabweichung) für Versuchsaufgabe 2 „Find Unique Shape“. IT steht für Interaktionstechnik, BT für Blick+Taste, M für Mauseingabe.

2: Find Un. Sh.		Trainings-Monate [-Stunden h] bis zum Testzeitpunkt				
Bedingung		**0 [0 h]**	**1 [9 h]**	**2 [18 h]**	**3 [27 h]**	**6 [52 h]**
IT	Pixel/s					
BT	78	2063 (200)	1776 (284)	1442 (112)	1351 (90)	1333 (125)
M	78	2258 (142)	-	-	-	1441 (168)

Bei **Versuchsaufgabe 3 (Catch ’em All)** erzielten die Versuchspersonen mit Blick+Taste für 1 bis 8 Zielobjekte 0% Selektionsfehlerquote bei allen Tests. Tabelle 5.5 zeigt die Ergebnisse der *Selektionsfehlerquote* für die Anzahl von 9 bis 18 Zielobjekten. Beim finalen Test war sie für alle Bedingungen < 1%. Dieses Ergebnis wurde für eine Anzahl bis 14 Zielobjekte bereits nach einem Monat Training erreicht, für 15 und 16 Zielobjekte nach zwei Monaten. Die Verbesserung beim finalen Test im Vergleich zum Baseline-Test war signifikant für 13 bis 16 Zielobjekte mit p<0,05, für 17 und 18 Zielobjekte mit p<0,01.

Mit Mauseingabe war die Selektionsfehlerquote für 1 bis 8 Zielobjekte ebenso bereits beim Baseline-Test bei 0%. Beim finalen Test waren die Ergebnisse für alle Bedingungen verbessert (signifikant für 12, 14, 15 und 18 Zielobjekte).

Blick+Taste und Mauseingabe erzielten beim Baseline-Test ähnliche Ergebnisse (keiner der Unterschiede war signifikant). Beim finalen Test war Blick+Taste signifikant besser für 17 und 18 Zielobjekte.

Die *Selektionszeit* für Blick+Taste lag beim Baseline-Test zwischen 502 ± 56 ms für ein Zielobjekt und 6093 ± 831 ms für 18 Zielobjekte (entspricht pro Zielobjekt im Durchschnitt 338 ± 46 ms); der Mittelwert ± 1 Standardabweichung über alle 18 Zielobjektmengen betrug 385 ± 37 ms. Dieses Ergebnis verbesserte sich konstant über die Trainingsperiode. Beim finalen Test lagen die Selektionszeiten bei 470 ± 102 ms für ein Zielobjekt und bei 5777 ± 384 ms für 18 Zielobjekte (entspricht pro Zielobjekt im Durchschnitt 321 ± 21 ms);

der Mittelwert ± 1 Standardabweichung über alle 18 Zielobjektmengen betrug 335 ± 35 ms. Die Gesamtzeit für alle 18 Trials reduzierte sich signifikant (p<0,01) von 63,6 ± 3,4 s vor Trainingsbeginn auf 56,0 ± 3,8 s nach 6 Monaten.

Mit der Mauseingabe lag die Selektionszeit beim Baseline-Test zwischen 730 ± 98 ms für ein Zielobjekt und 6698 ± 707 ms für 18 Zielobjekte (pro Zielobjekt 372 ± 10 ms); der Mittelwert ± 1 Standardabweichung über alle 18 Zielobjektmengen betrug 449 ± 85 ms. Beim finalen Test lagen die Werte zwischen 554 ± 100 ms und 7062 ± 475 ms (pro Zielobjekt 392 ± 26 ms); der Mittelwert ± 1 Standardabweichung über alle 18 Zielobjektmengen betrug 421 ± 43 ms. Die Gesamtzeit war praktisch gleich, 70,9 ± 3,7 s beim Baseline-Test, 70,7 ± 6,5 s beim finalen Test.

Im Vergleich zur Mauseingabe ist Blick+Taste bei beiden Tests signifikant schneller (p<0,001), beim Baseline-Test um 16%, beim finalen Test um 26%.

Bei **Versuchsaufgabe 4 (Air Traffic Control)** wurde die Effektivität aufgrund des Versuchsaufgabendesigns über die *Anzahl der Kollisionen* als Maß erfasst. Tabelle 5.6 zeigt die Ergebnisse. Die Anzahl nimmt für Blick+Taste für beide Objektgrößen G und K über die Trainingsperiode ab. Die Verbesserung zwischen Baseline-Test und finalem Test beträgt im Mittel 30% für Bedingung G, 40% für Bedingung K (Unterschiede signifikant mit p<0,05). Die Verbesserungen waren für Bedingung G bereits nach einem Monat Training signifikant, für Bedingung K nach zwei Monaten.

Auch mit Mauseingabe verbesserte sich das Ergebnis für beide Bedingungen signifikant (p<0,01). Die Ergebnisse waren für beide Interaktionstechniken ähnlich beim Baseline-Test. Beim finalen Test schnitt Blick+Taste signifikant besser ab (p<0,05 für Bedingung G, p<0,01 für Bedingung K).

Die Anzahl Objekte der Versuchsaufgaben lag mit Blick+Taste im Mittel ± 1 Standardabweichung bei 836 ± 59 für Bedingung G und bei 759 ± 42 für Bedingung K. Für die Mauseingabe waren es 759 ± 42 Objekte bei Bedingung G und 767 ± 57 bei Bedingung K. Wie oben bemerkt, rührt die unterschiedliche Anzahl daher, dass jede getätigte Selektion den weiteren Verlauf der Aufgabe beeinflusst, da jedes selektierte Objekt durch ein neues ersetzt wird.

Tabelle 5.5: Selektionsfehlerquote in Prozent als Mittelwert (± 1 Standardabweichung) für Versuchsaufgabe 3 „Catch 'em All". IT steht für Interaktionstechnik, BT für Blick+Taste, M für Mauseingabe.

3: Catch 'em All		Trainings-Monate [-Stunden h] bis zum Testzeitpunkt				
Bedingung		**0 [0 h]**	**1 [9 h]**	**2 [18 h]**	**3 [27 h]**	**6 [52 h]**
IT	Objektanzahl					
BT	9	2,1 (2,7)	0.0			
	10	4,4 (5,9)	0,0	0,6 (1,3)	0.0	
	11	3,4 (5,4)	0.6 (1,1)	0.0		
	12	9,4 (8,8)	0.5 (1.0)	0,5 (1,0)	0.0	
	13	7,7 (5,7)	0.0			
	14	11,6 (8,3)	0.0			0,4 (0,9)
	15	11,7 (8,3)	1,3 (1,6)	0.0	0.4 (0.8)	0.0
	16	12,2 (8,9)	2,0 (3,0)	0.0	0.4 (0.8)	0.0
	17	19,1 (8,7)	2,9 (4,2)	2,9 (2,1)	0,0	0.4 (0.7)
	18	18,8 (8,3)	9,4 (7,1)	2,4 (3,3)	2,4 (2,1)	0,3 (0,7)
M	9	6,3 (7,3)	-	-	-	0,7 (1,4)
	10	6,3 (6,0)	-	-	-	0,6 (1,3)
	11	9,7 (8,0)	-	-	-	2,8 (2,2)
	12	12,5 (9,0)	-	-	-	4,2 (5,1)
	13	9,6 (6,1)	-	-	-	4,3 (5,1)
	14	12,5 (6,8)	-	-	-	8,0 (6,3)
	15	12,5 (6,5)	-	-	-	6,7 (8,2)
	16	14,1 (8,9)	-	-	-	7,8 (8,2)
	17	18,8 (12,7)	-	-	-	15,1 (8,4)
	18	22,6 (7,0)	-	-	-	16,3 (9,7)

Tabelle 5.6: Kollisionen in Prozent als Mittelwert (± 1 Standardabweichung) für Versuchsaufgabe 4 „Air Traffic Control". IT steht für Interaktionstechnik, BT für Blick+Taste, M für Mauseingabe.

4: Air Traffic Control		Trainings-Monate [-Stunden h] bis zum Testzeitpunkt				
Bedingung		**0 [0 h]**	**1 [9 h]**	**2 [18 h]**	**3 [27 h]**	**6 [52 h]**
IT	Objektgröße					
BT	G	11,1 (1,0)	9,5 (0,6)	8,6 (0,8)	9,0 (0,9)	7,7 (1,3)
M	G	12,5 (1,1)	-	-	-	10,0 (0,9)
BT	K	11,1 (1,8)	8,9 (1,1)	7,4 (1,1)	6,9 (1,3)	6,7 (0,3)
M	K	11,2 (0,6))	-	-	-	9,2 (0,7)

Tabelle 5.7 zeigt die Ergebnisse der subjektiven Bewertung, die die Zufriedenstellung der Versuchspersonen mit den Interaktionstechniken dokumentiert. Betrachtet man Blick+Taste vor Beginn der Trainingsperiode und nach 6 Monaten Training, so sind die Ergebnisse ähnlich. Eine signifikante Verbesserung zeigte sich für die Ermüdung der Augen.

Vergleicht man Blick+Taste und Mauseingabe, so sind im Mittel fast alle Merkmale für Blick+Taste besser bewertet. Besonders auffallend sind die Unterschiede für die Erforderliche Anstrengung, die Benutzungsgeschwindigkeit, die Allgemeine Zufriedenheit, die Ermüdung des Handgelenks und die Ermüdung des Arms.

Einzige Ausnahme ist die Ermüdung der Augen, die im Mittel etwas besser für die Mauseingabe bewertet wurde. Die Ermüdung der Augen ist jedoch beim Baseline-Test auch für die Mauseingabe, mit der die Versuchspersonen langjährige Erfahrung haben, schlecht bewertet; im finalen Test sind die Ergebnisse für beide Interaktionstechniken deutlich besser.

Tabelle 5.7: Subjektive Bewertungen mit dem Fragebogen zur Einzelbewertung der DIN CEN ISO/TS 9241-411:2014 auf einer 7-Punkte-Skala (7: beste Bewertung, 1: schlechteste Bewertung) als Mittelwert (± 1 Standardabweichung). Signifikant bessere Resultate zwischen Blick+Taste und Mauseingabe innerhalb eines Tests[a] sind **fett** gedruckt. Signifikante Verbesserungen bei Blick+Taste zwischen Baseline-Test und finalem Test sind ***fett kursiv*** gedruckt.

Subjektive Bewertung	0 Monate/0 Stunden		6 Monate/52 Stunden	
	BT	M	BT	M
Gleichmäßigkeit bei der Nutzung	6,5 (0,6)	5,8 (1,3)	6,5 (0,6)	5,5 (1,0)
Erforderliche Anstrengung	5,8 (1,0)	3,0 (1,4)	**6,5 (0,6)**	4,0 (0,8)
Genauigkeit	5,0 (0,8)	4,3 (2,1)	6,0 (0,0)	5,0 (2,4)
Benutzungsgeschwindigkeit	6,8 (0,5)	4,3 (2,1)	**7,0 (0,0)**	3,8 (1,0)
Allgemeine Zufriedenheit	6,0 (0,0)	4,5 (1,7)	**6,0 (0,0)**	4,5 (0,6)
Nutzung insgesamt	6,3 (0,5)	5,5 (1,3)	7,0 (0,0)	5,8 (1,5)
Ermüdung Finger	6,3 (1,5)	5,3 (1,5)	7,0 (0,0)	5,8 (1,3)
Ermüdung Handgelenk	**7,0 (0,0)**	4,0 (0,8)	6,8 (0,5)	4,8 (3,1)
Ermüdung Arm	**7,0 (0,0)**	3,8 (1,5)	6,5 (1,0)	4,0 (2,2)
Ermüdung Schulter	7,0 (0,0)	5,5 (1,9)	6,5 (0,6)	5,0 (1,8)
Ermüdung Nacken	5,0 (1,4)	5,5 (2,4)	6,0 (0,8)	5,3 (1,7)
Ermüdung Augen	3,0 (0,0)	3,8 (2,2)	***4,8 (0,5)***	5,5 (1,7)

[a] Zweistichproben *t*-Test bei abhängigen Stichproben, Signifikanzniveau $\alpha = 0{,}05$.

Bezüglich der **Forschungsfragen** ergeben sich folgende Antworten.

Frage 1: Die Ergebnisse zeigen, dass Blick+Taste eine effektive und effiziente Interaktionstechnik für die Bewegtobjektselektion ist.

Die *Effektivität* verbessert sich signifikant über die Trainingsperiode hinweg. Der Zeitpunkt, zu dem sich der Lerneffekt stabilisiert, hängt stark davon ab, wie schwierig die Aufgabe ist, d.h. wie viele Objekte zu selektieren sind

und wie hoch ihre Geschwindigkeit ist. Bei Versuchsaufgabe 2 (Find Unique Shape), wo die Geschwindigkeit mit 78 Pixel/s (1,86°/s bzw. 2,1 cm/s) vergleichsweise niedrig war und 5 Sekunden Zeit für die Selektion gegeben waren, war die Selektionsfehlerquote von Beginn an nahe 0%.

Bei Versuchsaufgabe 1 (Circle Select) waren die Selektionsfehlerquoten für Geschwindigkeiten beim Baseline-Test noch vergleichsweise hoch. Für Geschwindigkeiten bis 500 Pixel/s (11,9°/s bzw. 13,5 cm/s) sind sie jedoch bereits nach 1 Monat Training niedrig. Für 650 Pixel/s (15,5°/s bzw. 17,6 cm/s) verbesserte sich das Ergebnis über die gesamte Trainingsperiode hinweg konstant immer weiter.

Vergleicht man die Ergebnisse des Baseline-Tests mit denen des Experiments aus Abschnitt 5.1.3, so ist die Selektionsfehlerquote im Baseline-Test etwas schlechter. Ein Grund dafür ist möglicherweise, dass dort jede der Bedingungen bis zu dreimal vorgelegt wurde, während hier für jede Bedingung nur ein Versuch gestattet war. Mit Training übertreffen die Ergebnisse der Längsschnittstudie die der Querschnittstudie aus Abschnitt 5.1.3 nach 1 Monat Training für 200 bis 500 Pixel/s, nach 3 Monaten auch für 650 Pixel/s.

Bei Versuchsaufgabe 3 (Catch 'em All) mit 100 Pixel/s (2,38°/s bzw. 2,7 cm/s) Geschwindigkeit zeigte sich, dass bis zu 8 Zielobjekte fehlerlos bereits ohne Training selektiert werden können. Nach 2 Monaten gelingt 0% Selektionsfehlerquote für bis zu 16 Zielobjekte, und nach 6 Monaten für bis zu 18 Zielobjekte.

Bei Versuchsaufgabe 4 (Air Traffic Control) war eine Stabilisierung des Lerneffekts nach 2 Monaten zu beobachten; die Ergebnisse verbesserten sich jedoch bis zum Ende der Trainingsperiode geringfügig weiter.

Die *Effizienz* in Form der Selektionszeit verbesserte sich ebenfalls für komplexere Aufgabenstellungen mit zunehmender Trainingsdauer. Bei Versuchsaufgabe 2 (Find Unique Shape) betrug die Verbesserung beim finalen Test 35%. Auch bei Versuchsaufgabe 3 (Catch 'em All) war für alle Zielobjektmengen eine Verbesserung zu sehen; die Gesamtselektionszeit war beim finalen Test 12% kürzer als beim Baseline-Test.

Im Gegensatz dazu erzielten die Versuchspersonen bei Versuchsaufgabe 1 (Circle Select) von Beginn an eine sehr kurze Selektionszeit von ca. 500 ms, die sich durch Training nicht weiter verkürzte. Da Circle Select nur die Selektion genau eines Zielobjekts verlangt und keine Distraktoren vorhanden sind, ist sie kognitiv deutlich simpler als die anderen Versuchsaufgaben. Die anderen Versuchsaufgaben umfassen Suche (Find Unique Shape) bzw. eine Strategie, welches Zielobjekt aus einer Gruppe von vielen als nächstes selektiert werden soll, damit die Aufgabenbearbeitung möglichst optimal gelingt (Catch 'em All, Air Traffic Control). Die verbesserten Selektionszeiten beinhalten daher auch Verbesserungen im Suchverhalten bzw. bei der Strategie.

Bei Versuchsaufgabe 1 (Circle Select) ist das Versuchsdesign so gestaltet, dass tatsächlich fast nur die reine Selektionszeit gemessen wird. Die einzige Unsicherheit im Vergleich zu Fitts-Law-Aufgaben (vgl. z. B. [Ver08]), die die Selektionszeit für statische Objekte messen, ist, dass dort die Zielobjektposition vor Beginn der Selektion bekannt ist. Bei Circle Select kann eine von sechs Positionen vorkommen, d.h., die Planung der Interaktion kann bei statischen Zielobjekten im Voraus geschehen, bei Circle Select erst beim Erscheinen des Zielobjekts. Die durchgehend im Mittel etwa bei 475 ms liegende Selektionszeit der Längsschnittstudie bestätigt das Ergebnis der Querschnittstudie aus Abschnitt 5.1.3, wo die Selektionszeit bei 540 ms lag.

Frage 2: Blick+Taste erwies sich als vielversprechende Alternative zur Mauseingabe. Die *Effektivität* war für Circle Select und Find Unique Shape ähnlich gut. Für die Aufgaben mit größerer Zielobjektmenge und mehr zu leistenden Selektionen (Catch 'em All und Air Traffic Control) war sie mit Blick+Taste beim finalen Test sogar besser.

Die *Effizienz* war für Blick+Taste für alle Versuchsaufgaben besser. Die Selektionszeit war für Blick+Taste im Mittel bei Versuchsaufgabe 1 (Circle Select) zwischen 14% (200 Pixel/s, 4,75°/s, 2,7 cm/s) und 23% (650 Pixel/s, 15,5°/s, 17,6 cm/s) kürzer, bei Versuchsaufgabe 2 (Find Unique Shape) um 7% und bei Versuchsaufgabe 3 (Catch 'em All) um 20% (Gesamtselektionszeit für einen Durchlauf aller Zielobjektmengen von 1 bis 18).

Frage 3: Die Zufriedenstellung der Versuchspersonen mit Blick+Taste war sehr hoch. Bereits beim Baseline-Test, als die Versuchspersonen zwar Mauseingabe-Experten, aber ohne jede Erfahrung mit Blickinteraktion waren, erzielte Blick+Taste sehr gute Bewertungen, die für die Mehrzahl der Merkmale besser waren als für die Mauseingabe. Einzig für das Merkmal Ermüdung der Augen war die Bewertung für Blick+Taste schlecht; für Mauseingabe war es trotz Erfahrung kaum besser.

Auch beim finalen Test blieben die Bewertungen für Blick+Taste sehr gut. Im Vergleich mit der Mauseingabe wurden die Merkmale Erforderliche Anstrengung, Benutzungsgeschwindigkeit und Allgemeine Zufriedenheit signifikant besser bewertet. Bei der Ermüdung der Augen zeigte Blick+Taste eine signifikante Verbesserung gegenüber dem Baseline-Test. Dieses Ergebnis subjektiver Bewertungen bestätigt den Lernerfolg durch das Training, der sich auch in den objektiven Maßen für Effektivität und Effizienz gezeigt hatte.

5.1.4.7 Fazit

Da die Längsschnittstudie mit nur 4 Versuchspersonen durchgeführt wurde, ist ihre statistische Aussagekraft limitiert und die Ergebnisse müssen als vorläufig angesehen werden.

Mit dieser Einschränkung lässt sich folgendes Fazit ziehen: Die Längsschnittstudie zu Blick+Taste bestätigte zum einen das Ergebnis der Querschnittstudien aus Abschnitt 5.1.1 und Abschnitt 5.1.3, wonach diese Interaktionstechnik bezüglich Effektivität und Effizienz eine Alternative zur Mauseingabe für die Bewegtobjektselektion ist. Zum anderen zeigte die Längsschnittstudie, dass je nach Komplexität der Aufgabe Training zum Teil erhebliche Verbesserungen bezüglich Effektivität und Effizienz bringt, wobei sich der Lerneffekt in den meisten Fällen nach 1 bis 3 Monaten stabilisiert. Zudem brachte Training eine signifikante Verbesserung bezüglich der Ermüdung der Augen.

5.2 Implizite blickbasierte Bewegtobjektselektion: Blick+EEG

Als implizite blickbasierte Interaktionstechnik wird die multimodale Kombination aus *Blick+EEG* betrachtet. Aufgrund des Zusammenhangs des ereigniskorrelierten Potenzials der P300 mit relevanten und selten auftretenden Zielreizen (Abschnitt 2.4.4) und der Verfügbarkeit nicht-intrusiver EEG-Messung ist es denkbar, EEG zur Selektionsauslösung zu nutzen.

Im Vergleich zu den expliziten blickbasierten Objektselektionstechniken geht der Ansatz *Blick+EEG* noch einen Schritt weiter, was die Ziele der Erhöhung der Geschwindigkeit und die Reduzierung der kognitiven und manuellen Belastung für einen Videobildauswerter angeht. Die räumliche Komponente der Selektion (das Zeigen auf das Objekt) wird wie bei den expliziten Techniken aus Abschnitt 5.1 (Ausnahme MAGIC pointing) ausschließlich mittels der Blickrichtung bestimmt. Die zeitliche Komponente der Selektion (die Selektionsauslösung) wird durch EEG-Signale bestimmt. Die Nutzung von EEG bedeutet für den Benutzer, dass er eine Selektion vollständig ohne manuelle Aktion durchführt.

Der Zweck eines solchen Ansatzes liegt darin begründet, dass Bildfolgenanalyse eine herausfordernde Aufgabe ist. Der Videoanalyseexperte kann relevante Ereignisse oder Objekte aus verschiedenen Gründen verpassen. Beispiele sind physikalische Überlastung, wenn viele sehr schnelle Objekte gleichzeitig oder kurz hintereinander zu selektieren sind, mentale Überbelastung aufgrund von Ablenkung (physikalisch oder auditiv) oder Müdigkeit, oder auch Zögern, wenn die Entscheidung „Zielobjekt“ versus „Kein Zielobjekt“ schwierig zu treffen ist.

In solchen Fällen mag der menschliche Beobachter wie gewohnt die Szene durchmustern. Er nimmt Ereignisse aber nicht bewusst wahr und verpasst es so, sie aktiv zu markieren und weiterzuleiten. Oder er nimmt Ereignisse bewusst wahr, zögert aber zu lange, sodass eine Markierung z. B. mit der Computermaus oder auch mit Blick+Taste nicht mehr möglich ist, weil das Objekt die Szene bereits verlassen hat. In solchen Situationen liegt die Blickposition

möglicherweise trotzdem in der Nähe des Ortes, an dem das Ereignis auftrat, d.h. die räumliche Information über das Ereignis ist aus der Blickposition extrahierbar. Die zeitliche Information ist jedoch verloren.

Zwar gab es zur Zeit der Experimentdurchführung Untersuchungen, die Blick+EEG kombinierten (vgl. Abschnitt 2.4.4). Allerdings war die Möglichkeit der Objektselektion ohne Selektionsauslösung mittels mentaler Vorstellung einer motorischen Aktion nicht betrachtet worden. Die Untersuchungen zu Blick+EEG in der vorliegenden Arbeit verfolgten daher das Ziel, die prinzipielle Machbarkeit der Nutzung von Blick+EEG für die Objektselektion festzustellen. Denn die Extraktion der P300 aus dem mehrkanaligen EEG ist algorithmisch herausfordernd.

Die Untersuchungen erfolgten daher nicht in Form einer echtzeitfähigen Blick+EEG-Eingabe für die Interaktion mit einer GUI, bei der der Benutzer das Ergebnis seiner Selektionen visuell rückgemeldet bekommt. Stattdessen wurden anhand abstrakter Versuchsparadigmen (mit geometrischen Formen als Zielobjekten) Blick- und EEG-Daten aufgezeichnet, um das Zielobjekt räumlich-zeitlich zu lokalisieren. Abstrakter wird dieser Vorgang auch als räumlich-zeitliche Ereignis-Lokalisation bezeichnet. Ziel ist herauszufinden, ob diese Ereignis-Lokalisation zur Echtzeiteingabe anstelle einer expliziten Interaktion geeignet sein könnte. Dazu werden die beiden Datensätze offline einzeln ausgewertet und dann für die räumlich-zeitliche Ereignis-Lokalisierung zusammengeführt. Aufgrund des Grundlagencharakters der Untersuchungen wurden neben bewegten Objekten/Ereignissen auch statische Objekte/Ereignisse betrachtet.

5.2.1 Grundsätzliche Machbarkeit

Dieses Experiment und seine Ergebnisse wurden bei der IUI 2013 veröffentlicht [Put13].

Das Ziel war, die grundsätzliche Machbarkeit für Blick+EEG zu überprüfen.

Abbildung 5.27: Testparadigma 1: Zählaufgabe.

5.2.1.1 Testaufgaben

Insgesamt wurden drei Testparadigmen genutzt, eines davon war eine Zählaufgabe, zwei waren Zielobjekt-Lokalisierungsaufgaben.

Abb. 5.27 zeigt den visuellen Stimulus für Paradigma 1, die Zählaufgabe. Zielobjekt ist ein Quadrat mit Seitenlänge 100 Pixel (2,38° bzw. 2,7 cm), zentriert auf grauem Hintergrund. Dieses Quadrat wechselt seine Farbe zwischen hellgrau und rot. Solange das Quadrat hellgrau ist, ist es einfach das Objekt, auf das der Benutzer seine visuelle Aufmerksamkeit ausrichtet. Färbt sich das Quadrat rot – für 2 Sekunden, zwischen je 2 Rotfärbungen wird das Quadrat für mindesten 1,5 Sekunden wieder hellgrau –, so zeigt dies ein Ereignis an bzw. das Objekt wird zum Zielobjekt. Dieses Versuchsparadigma ist ein sehr abstraktes Modell der Überwachung eines statischen Objekts (beispielsweise ein Gebäude), an dem ein relevantes, berichtenswertes Ereignis geschieht (etwa das Eintreten einer Person). Die Versuchspersonen werden instruiert, das Quadrat kontinuierlich zu fixieren (Ausrichtung des Blicks auf das Fixationskreuz in der Mitte) und zu zählen, wie oft das Quadrat sich rot gefärbt hat.

Abb. 5.28 und Abb. 5.29 zeigen die Paradigmen 2 bzw. 3, beides Zielobjekt-Lokalisierungsaufgaben. Paradigma 2 ist inspiriert vom Tipptest mit mehreren Richtungen aus [DIN14], abgewandelt nach Zhang u. a. [Zha07] (vgl. S. 146). Es umfasst 16 statische Zielobjekte (100 Pixel Seitenlänge bzw. 2,38°)

kreisförmig angeordnet um ein Quadrat von 100 Pixel Seitenlänge. Diese Größe wurde gewählt, weil wir annahmen, dass sie bei der vom Hersteller angegeben Genauigkeit des X60-Eyetracker ausreichen sollte, dass bei Blick auf die Fixationskreuze die Blickposition trotz Messunsicherheit auf das zugehörige Zielobjekt fällt; verwandte Arbeiten zu blickbasierter Selektion statischer Objekte nutzen Objektgrößen ab 2° (vgl. Abschnitt 3.1.1). Die Distanz zwischen dem Mittelpunkt des Quadrats und den Mittelpunkten der Objekte betrug 500 Pixel.

Ereignisse, die eines der Kreisobjekte als Zielobjekt hervorheben, werden durch eine rote Markierung von 2 Sekunden Dauer angezeigt. Zwischen je zwei Hervorhebungen liegen mindestens 1,5 Sekunden. Die Versuchspersonen wurden instruiert, kontinuierlich das Fixationskreuz des Quadrats zu fixieren und im Falle einer roten Hervorhebung ihren Blick so schnell wie möglich auf das Fixationskreuz dieses Zielobjekts zu richten. Die Fixation des Zielobjekts sollte so lange beibehalten werden, bis der Beobachter sicher war, das Fixationskreuz als Symbol wahrgenommen zu haben. Danach sollte der Blick so schnell wie möglich wieder auf das Quadrat ausgerichtet werden.

Paradigma 3 (Abb. 5.29) erweitert Paradigma 2 auf bewegte Zielobjekte. Jetzt bewegen sich die Kreise radial mit zwei unterschiedlichen Geschwindigkeiten (115 Pixel/s, 135 Pixel/s, vgl. a. Abb. 5.10a), die sich an den Umkehrpunkten zufällig änderten. Ereignisse in Form roter Hervorhebung konnten an den Umkehrpunkten (Distanz 250 Pixel bzw. 500 Pixel) sowie in der Mitte dazwischen auftreten. Die Aufgabeninstruktion war dieselbe wie bei Paradigma 2.

Der Vorteil dieser simplen Szenarien liegt darin, dass die Distanz zum Zielobjekt für jedes Trial kontrolliert werden kann. Dadurch werden störende Einflüsse wie überlappende Ereignisse oder unterschiedliche Strategien bei der Durchmusterung der Zielobjekte verhindert. Die klare, abstrakte Gestaltung spiegelt zudem gewissermaßen die Fähigkeit erfahrener Videoanalyseexperten wider, relevante visuelle Information aus einem verrauschten Videostrom zu extrahieren.

Abbildung 5.28: Testparadigma 2: Objektselektion mit statischen Zielobjekten.

Abbildung 5.29: Testparadigma 3: Objektselektion mit bewegten Zielobjekten. Links: Startsituation; Mitte: Objekte haben ihre Bewegung begonnen; rechts: Hervorhebung eines bewegten Kreises als Zielobjekt.

5.2.1.2 Versuchsaufbau, Versuchspersonen und Versuchsablauf

Die Versuchssoftware wurde als JAVA-Anwendung unter Windows 7 programmiert. Zur Darstellung der Animationen der Testaufgabe wurde die Trident 1-Bibliothek verwendet.

Zur Blickerfassung wurde der Tobii X60 ohne Kopfstabilisierung verwendet (vgl. Abschnitt 4.2).

Zur Erfassung des EEG wurde ein actiCHamp Recorder wie in Abschnitt 4.4 beschrieben eingesetzt.

Die Synchronisation zwischen Versuchsaufgabe, Blickerfassung und EEG erfolgte über eine Fotodiode, die von einer von der Software gesteuerten Markierung in der linken oberen Monitorecke Lichtsignale empfängt (vgl. Abb. 5.27 bis 5.29 das kleine Quadrat in der linken oberen Ecke). Die Markierung änderte ihre Farbe zwischen schwarz (kein Ereignis) und weiß (Ereignis = rote Hervorhebung). Auf diese Weise wurde die Synchronisierung der Datenströme auf Framelevel sichergestellt.

Abb. 5.30 zeigt den Versuchsaufbau.

Die 11 Versuchspersonen (9 männlich, 2 weiblich; Alter zwischen 26 und 33, Altersdurchschnitt 28,6 Jahre) verfügten alle über normale oder auf normal korrigierte Sicht. Alle waren Studenten oder Kollegen. Zwei hatte Erfahrung mit Eyetracking, keiner mit EEG-Messung.

Jede Versuchsperson absolvierte insgesamt 5 Testaufgaben. Alle begannen mit Paradigma 1, wobei die Anzahl zu zählende Ereignisse (Rotfärbungen) 10 betrug. Dann folgten Paradigma 2 und 3 je zweimal abwechselnd, wobei mit der statischen Variante begonnen wurde. Jede dieser Aufgaben umfasste 32 Ereignisse, d.h. jedes Zielobjekt wurde zweimal rot hervorgehoben. Die Reihenfolge der Rotfärbung erfolgte randomisiert. Jede Versuchsperson absolvierte insgesamt 128 Trials. Vor der Absolvierung der Testaufgaben erfolgte ein Training in derselben Reihenfolge der Aufgaben, aber mit reduzierter Anzahl Ereignisse (7 für Paradigma 1, 8 für Paradigmata 2 und 3).

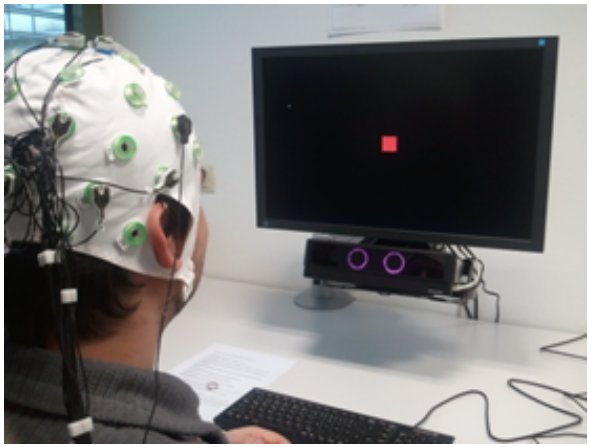


Abbildung 5.30: Versuchsaufbau.

5.2.1.3 Forschungsfragen, Metriken und Ergebnisse

Dieses Experiment adressierte folgende Forschungsfragen:

- Frage 1: Ist die Ereignis-Lokalisation mit Blick+EEG überhaupt möglich?
- Frage 2: Wenn ja, mit welcher Trefferquote, räumlichen Genauigkeit und Geschwindigkeit?

Im Rahmen der vorliegenden Arbeit wurde die Analyse der Blickdaten durchgeführt, die zum Gesamtsystem die *räumliche Lokalisation der Ereignisse* beiträgt. Eine räumliche Lokalisation eines Zielobjekts gilt als erfolgreich, wenn der Beobachter das Zielobjekt fixiert hat, unmittelbar nachdem es rot hervorgehoben wurde. Dafür wird für jedes Trial – bestehend aus (1) Fixation des Quadrats (2) Sakkade zum hervorgehobenen Zielobjekt (3) Fixation des Zielobjekts – bestimmt, ob und zu welchem Zeitpunkt der Blick das Zielobjekt zu fixieren beginnt.

Die Analyse erfolgte in mehreren Schritten:

1. Klassifikation der Blickrohdaten in Fixationen und Sakkaden.
2. Berechnung der ersten Fixation auf jedem Zielobjekt unter Berücksichtigung der Sakkadenreaktionszeit (SRT)
3. Berechnung der Selektionstrefferquote, SRT (Saccadic Reaction Time) und FOT (Fixation Onset Time).

Die Klassifikation der Blickdaten in Fixationen und Sakkaden erfolgte mithilfe des I-VT-Algorithmus [Sal00], s. a. Abschnitt 2.4.1. Die minimale Fixationsdauer wurde auf 100 ms gesetzt. Als Schwellenwert v_{max} wurde 50°/s genutzt, da die Datenaufzeichnung ohne Kopfstabilisierung erfolgte. Durch die erlaubten natürlichen Kopfbewegungen erhöht sich das Rauschen des gelieferten Blicksignals und es wurde erwartet, dass das Tragen der EEG-Kappe mit ihrer Verkabelung dieses Rauschen noch vergrößern könnte.

Auf Basis der erhaltenen Fixationsmenge wurde dann für jedes Trial die erste Fixation auf dem Zielobjekt bestimmt. Als ihr Startzeitpunkt (Fixation Onset Time *FOT*) wurde der Endzeitpunkt der Sakkade vom Quadrat zum Zielobjekt genutzt. Lag die Fixationsposition auf dem Zielobjekt, wurde dies als Treffer gezählt.

Damit eine Fixation als die erste Fixation auf dem Zielobjekt galt, durfte die unmittelbar davor liegende Sakkade eine bestimmte Sakkadenreaktionszeit sowie eine bestimmte Sakkadendauer nicht unterschreiten. So sollte sichergestellt werden, dass die Blickbewegung den typischen physiologischen Eigenschaften entsprach und nicht nur zufällig das Zielobjekt traf.

Die Sakkadenreaktionszeit wurde als Differenz zwischen dem Startzeitpunkt der Rotfärbung des Zielobjekts und dem Endzeitpunkt der letzten Fixation auf dem Quadrat berechnet. Lag die Sakkadenreaktionszeit unter 150 ms, wurde der Zielobjektselektionsversuch als gescheitert gezählt. Dieser Wert wurde gewählt als mittlerer Wert, den Gezeck u. a. [Gez97] für schnelle reguläre Sakkaden angeben.

Die Sakkadendauer wurde als Differenz zwischen der FOT und dem Startzeitpunkt der Sakkade (= Endzeitpunkt der letzten Fixation auf dem Quadrat)

berechnet. Die Sakkadendauer kann mithilfe der Formel (2.2) (vgl. S. 32) modelliert und berechnet werden. Dieser Wert wurde für jedes Trial bestimmt. Lag eine aus den Daten berechnete Sakkadendauer unter dem zugehörigen Modell-Wert, wurde der Selektionsversuch als physiologisch unmöglich und daher gescheitert gezählt.

Tabelle 5.8 zeigt die Ergebnisse für die Trefferquote. Die Ergebnisse bei Zugrundelegung der Objektgröße von 2,38° (linke Ergebnisspalte) betragen im Mittel 49,6% für Paradigma 2 mit den statischen Zielobjekten und 38,9% für Paradigma 3 mit den bewegten Zielobjekten.

Betrachtet man eine virtuelle Vergrößerung der Zielobjekte auf 3° (126 Pixel), verbessern sich die Ergebnisse erheblich auf 76,2% (statisches Paradigma 2) bzw. 66,0% (bewegtes Paradigma 3). Im Falle von Paradigma 3 muss dieser Wert noch etwas korrigiert werden, da bei einer Objektgröße von 3° die Objekte mit 12,3% ihrer Fläche überlappen können, wenn sie sich am inneren Umkehrpunkt (Distanz 250 Pixel) befinden. Im Falle der Testaufgabe war gegeben, dass 12 der 32 Zielobjekte mit einem weiteren Objekt überlappten, 2 mit zwei weiteren. Wenn man dies berücksichtigt, so muss man die Trefferquote von 66,0% um 12,3% ∗ (12/32) + 24,6% ∗ (2/32) = 6,1% reduzieren.

Die Trefferquote verbessert sich auf 86,3% (statisches Paradigma 2) bzw. 70,9% (bewegtes Paradigma 3), wenn man die Ergebnisse jeder Versuchsperson mit ihrem individuellen Kalibrieroffset korrigiert. Dieser Kalibrieroffset wurde aus Paradigma 1 berechnet, indem die mittlere Distanz zwischen Blickposition und Quadratmittelpunkt bestimmt wurde.

Tabelle 5.9 zeigt die Ergebnisse für Sakkadenreaktionszeit (SRT) und Fixationsstart auf dem Zielobjekt (FOT). Für die statischen Zielobjekte sind beide geringfügig kürzer. Dies könnte darauf hindeuten, dass die Ereignis-Lokalisierung für bewegte Zielobjekte die schwierigere Aufgabe ist bzw. die automatische Klassifikation von Fixationen und Sakkaden hier schwieriger ist.

Tabelle 5.8: Trefferquote in Prozent als Mittelwert (± 1 Standardabweichung) bei der räumlichen Lokalisation (Selektion) von Ereignissen (Objekten).

-	Zielobjektgröße		
	2,38°	3°	3° und Kalibrierkorrektur
Paradigma 2 (statisch)	49,6 (28,1)	76,2 (17,4)	86,3 (10,3)
Paradigma 3 (bewegt)	38,9 (25,7)	66,0 (20,3)	77,0 (11,5)
Konservativ korrigiert	-	59,9 (20,3)	70,9 (11,5)

Tabelle 5.9: Sakkadenreaktionszeit (SRT) und **Fixationsstart auf dem Zielobjekt (FOT) in ms** als Mittelwert (± 1 Standardabweichung).

-	SRT	FOT
Statische Zielobjekte (Paradigma 2)	251 (62)	375 (86)
Bewegte Zielobjekte (Paradigma 3)	285 (90)	436 (113)

Die Ergebnisse der *zeitlichen Ereignislokalisierung* im EEG sind im Detail in Putze u. a. [Put13] beschrieben. Sie erfolgte auf Basis der EEG-Daten. Diese wurden dafür in kurze Abschnitte segmentiert. Die Lokalisierung wurde als Zwei-Klassen-Klassifikationsproblem behandelt, bei dem erkannt werden musste, ob ein EEG-Datensegment ein *trial* oder ein *Nicht-trial* repräsentiert. Ein *trial* ist ein Segment, das unmittelbar auf eine rote Hervorhebung folgt, ein *Nicht-trial* ein Segment, das keine oder nur geringe Überlappung mit einem *trial* hat.

Für die Klassifikationsaufgabe wurden Segmente von 200, 300, 400 und 500 ms Dauer unmittelbar nach der roten Hervorhebung als Repräsentation für *trials* ausgeschnitten. Segmente derselben Dauer wurden jeweils vor und nach den *trial*-Segmenten ausgeschnitten und repräsentierten *Nicht-trials*. Die Segmentdauern liegen innerhalb plausibler physiologischer Grenzen neuronaler

Reaktion der P300. Sie erlauben zudem die Betrachtung eines Trade-offs zwischen Erkennungsakkuratheit und Erkennungslatenz: Ein kürzeres Segment resultiert in schnellerer Reaktionszeit, was besonders bei zeitkritischen Ereignisse relevant ist; es resultiert aber auch in weniger Daten, die für die Klassifikation zur Verfügung stehen.

Zudem wurden zwei Erkennungsmodelle betrachtet, ein Personen-abhängiges und ein Personen-unabhängiges. Da es in EEG-Daten große individuelle Unterschiede gibt, war anzunehmen, dass das Personen-abhängige Modell bessere Klassifikationsergebnisse liefern würde.

Außerdem wurden unterschiedlich große Mengen an Elektroden für die Analyse betrachtet, zum einen alle 28 Elektroden (*Full*), zum anderen die 5 Elektroden, die mit visueller Wahrnehmung und Ereignisdetektion assoziiert sind (*Reduced*, vgl. Cz, Pz, Oz, O1 und O2 in Abb. 4.5). Als Klassifikationsmodell wurde eine Support Vector Machine mit Radial Basis Function kernels verwendet [Put13].

Die erzielte Klassifikationsakkuratheit lag mit der reduzierten Elektrodenmenge *Reduced* sowohl für Paradigma 2 als auch für Paradigma 3 für alle Segmentlängen bei ca. 91%, und zwar sowohl für das Personen-abhängige als auch für das Personen-unabhängige Erkennungsmodell. Die Tatsache, dass 91% bereits mit 200 ms Segmentlänge möglich sind, lässt für ein zukünftiges Online-Interaktionsystem eine schnelle Reaktionszeit möglich erscheinen. Die Ergebnisse mit der Elektrodenmenge *Full* liegen etwas höher und erreichen das beste Ergebnis von 96% mit 500 ms Segmentlänge.

Bezüglich der **Forschungsfragen** ergeben sich damit folgende Antworten.

Frage 1: Die Ergebnisse zeigen, dass die räumlich-zeitliche Ereignis-Lokalisierung Personen-unabhängig möglich ist, sowohl für statische als auch für bewegte Zielobjekte.

Frage 2: Die zeitliche Lokalisierung ist mit sehr geringer Latenz von 200 ms nach Auftreten eines Ereignisses mit 91% Akkuratheit möglich. Die räumliche

Lokalisierung ist für statische Zielobjekte mit 86,3% und für bewegte Zielobjekte mit 70,9% möglich, wenn die Zielobjektgröße 3° beträgt und der individuelle Kalibrieroffset zur Korrektur der Blickposition genutzt wird.

Unter der Annahme, dass zeitliche und räumliche Lokalisierung voneinander unabhängig sind, ist die kombinierte räumlich-zeitliche Lokalisierung für statische Zielobjekte mit 78,5% (91% ∗ 86,3%) und für bewegte Zielobjekte mit 64,5% (91% ∗ 70.9%) möglich.

Während die zeitliche Lokalisierung eines Ereignisses mit einer Latenz von 200-300 ms nach Eintrittszeitpunkt des Ereignisses möglich ist, wird die räumliche Komponente für statische Objekte 375 ± 86 ms nach Eintrittszeitpunkt des Ereignisses geliefert, für bewegte Objekte nach 436 ± 133 ms.

5.2.1.4 Fazit

Die Untersuchung zeigte, dass Personen-unabhängige räumlich-zeitliche Ereignis-Lokalisierung mit Blick+EEG möglich ist. Während die zeitliche Lokalisierung 200-300 ms nach Eintritt eines Eregnisses möglich war, erfolgte die räumliche Lokalisierung kurz später nach 375 ± 86 ms für statische bzw. nach 436 ± 133 ms für bewegte Objekte. Für eine Systemeingabe kämen noch 100 ms als typische minimale Fixationsdauer hinzu, um sicherzustellen, dass es sich um die erste Fixation nach Eintritt des Ereignisses handelt.

Verglichen mit den oben betrachteten expliziten Interaktionstechniken Blick+Taste oder Blick+Fußtaste ist die resultierende Selektionszeit vergleichbar. Mit Blick+EEG wird der Benutzer jedoch von jeglicher motorischer Aktion der Extremitäten sowie bewusster kognitiver Aktion entlastet. Die Selektionsfehlerquote (gegeben die vergleichsweise langsame Objektgeschwindigkeit) liegt hingegen deutlich höher, was jedoch mit der geringeren Objektgröße (3° versus > 4°) erklärt werden kann.

5.2.2 Blick+EEG bei einfachen, simulierten Überwachungsaufgaben

Dieses Experiment und seine Ergebnisse wurden bei der ICMI/GazeIn 2014 veröffentlicht [Hil14b].

Es baut auf der in Abschnitt 5.2.1 beschriebenen Untersuchung auf, betrachtet aber weniger artifizielle experimentelle Paradigmen, die die Situation realer Videoüberwachungsaufgaben besser simulieren. Dies spiegelt sich in den genutzten visuellen Stimuli und subtileren Hinweisreizen für die Ereignisse sowie darin, dass realistischeres Beobachtungsverhalten gefordert ist.

5.2.2.1 Testaufgaben

Insgesamt wurden zwei Testaufgaben gestaltet, bei denen ein Zielobjekt beobachtet werden musste. Bei Testaufgabe 1 war das Zielobjekt statisch, bei Testaufgabe 2 ein bewegtes Objekt.

Abb. 5.31 bzw. Abb. 5.32 zeigen die visuellen Stimuli. Es sind abstrakte Stimuli, die jedoch gemäß der visuellen Parameter von Überflugvideos gestaltet sind bezüglich Farbgebung (blasse Farben mit geringem Kontrast zwischen Hintergrund und Zielobjekt), Objektgrößen sowie der Präsenz von Distraktoren (vgl. die Bilder in [Hei10]). Das Zielobjekt ist ein hellgrauer Kreis (Durchmesser 0,72°, entspricht in der Versuchsanordnung 30 Pixel bzw. 0,81 cm).

In Testaufgabe 1 ist das Zielobjekt statisch innerhalb einer 9 x 5-Anordnung von Distraktoren platziert. Die Abstände zwischen den Mittelpunkten aller Objekte betragen 3,16°. Farben und Formen variieren, sodass einige dem Zielobjekt gleichen, die meisten sich jedoch in Farbe und/oder Form unterscheiden.

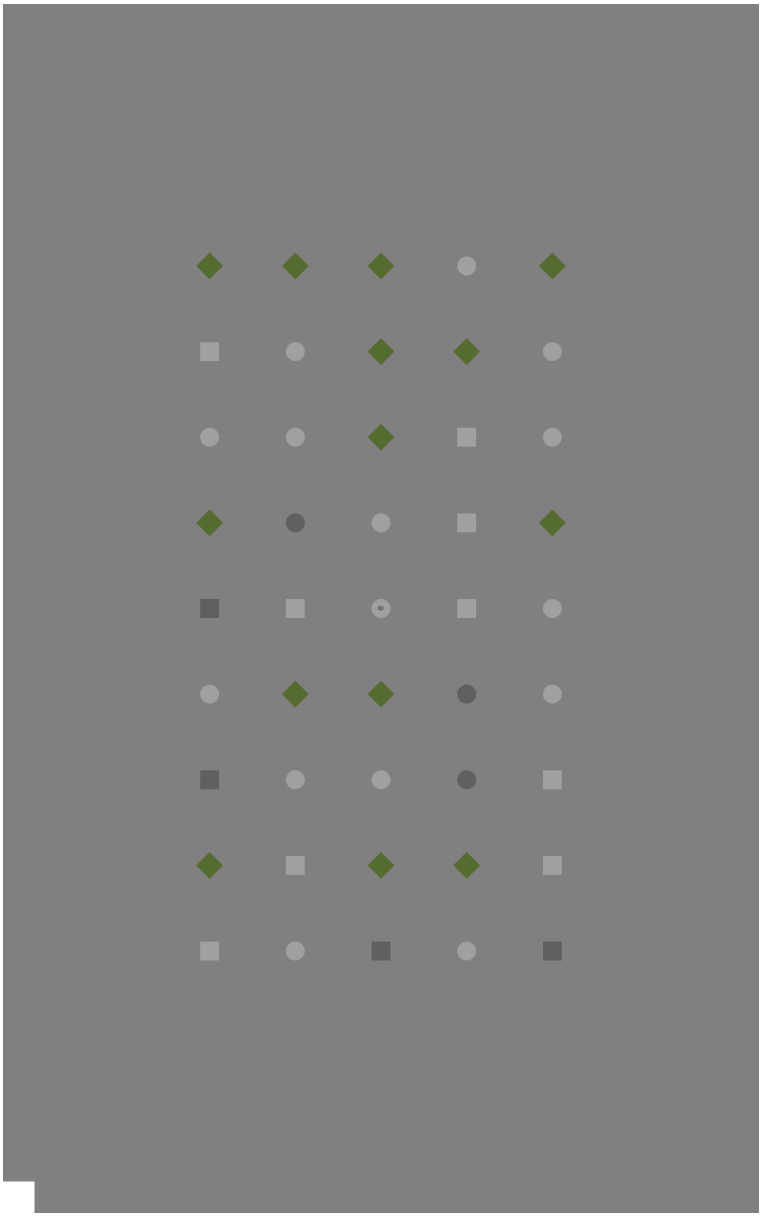


Abbildung 5.31: Testaufgabe 1: Zählaufgabe an statischem Zielobjekt.

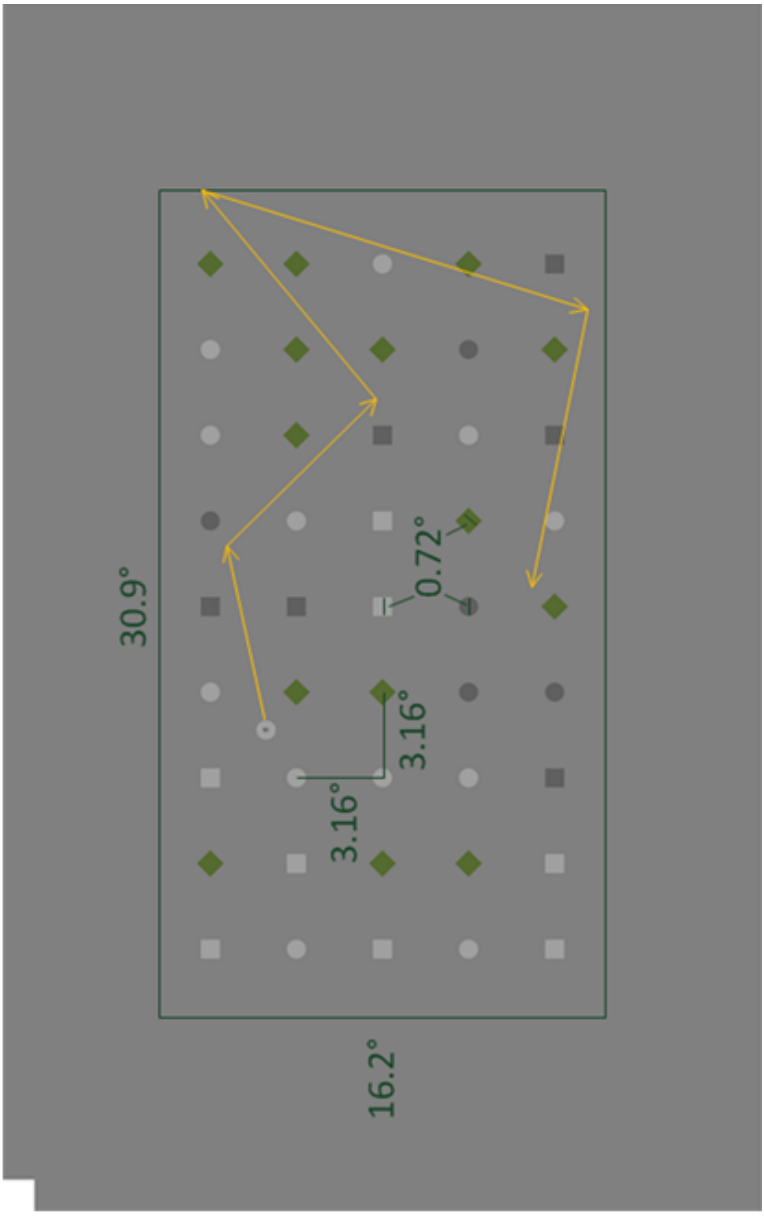


Abbildung 5.32: Testaufgabe 2: Zählaufgabe an bewegtem Zielobjekt.

In Testaufgabe 2 bewegt sich das Zielobjekt linear mit konstanter Geschwindigkeit von 70 Pixel/s (1,67°/s bzw. 1,89 cm/s) oder 100 Pixel/s (2,38°/s bzw. 2,7 cm/s) durch die 9 x 5-Anordnung; die Bewegung erfolgt innerhalb eines Bereichs von 30,9° x 16,2° um die Bildschirmmitte. Die Geschwindigkeit ändert sich zufällig (oder bleibt dieselbe), wann immer das Zielobjekt seine Richtung ändert. Die gelben Pfeile in Abb. 5.32 zeigen einen beispielhaften Bewegungspfad; eine Geschwindigkeitsänderung kann immer am Startpunkt jedes Pfeils auftreten. Die Anordnung der Distraktoren unterscheidet sich von der in Testaufgabe 1.

Die Ereignisse sind mithilfe von zwei Hinweisstufen realisiert. Die erste Stufe ist die Anzeige unterschiedlicher Buchstaben auf dem Zielobjekt. Buchstaben erscheinen zufällig für 2 Sekunden, wobei zwischen je zwei Buchstabenanzeigen 2 Sekunden ohne Buchstabe liegen. Diese erste Stufe entspricht der Anzeige eines Ereignisses durch rote Hervorhebung der Machbarkeitsuntersuchung aus Abschnitt 5.2.1 (vgl. Abb. 5.28 und 5.29). Die Buchstabenanzeige nutzt die Schriftart Arial fett, dunkelgrau mit Fontgröße 16. Dies stellt gute Lesbarkeit sicher und realisiert gleichzeitig einen subtileren Hinweisreiz als ein klar hervorstechender roter Punkt.

Die zweite Stufe implementiert ein sogenanntes Oddball-Paradigma (vgl. a. S.85). Wir unterscheiden 'e' als Zielbuchstabe von verschiedenen Distraktor-Buchstaben. Letztere umfassen die Buchstaben 'h', 'p', 'n' (alle aufgrund ihrer Gestalt einfach unterscheidbar von 'e') und 'c' (ähnliche Gestalt wie 'e', um sicherzustellen, dass die Versuchspersonen die angezeigten Buchstaben genau fixieren). Der menschliche Beobachter muss jetzt mithilfe seiner Kognition entscheiden, ob ein Buchstabe der Zielbuchstabe 'e' ist oder nicht. Die Erwartung ist, dass dies die zeitliche Lokalisierung eines Ereignisses deutlich schwieriger macht im Vergleich zum einfachen Farbwechsel (rote-Punkt-Hervorhebung) der Machbarkeitsuntersuchung aus Abschnitt 5.2.1; denn ein Farbwechsel kann auch präattentiv ohne kognitiven Aufwand wahrgenommen werden.

Zusammengenommen kann man die beiden Stufen bei einer realen Videoauswerteaufgabe folgendermaßen interpretieren. Bei Testaufgabe 1 repräsentiert

das Zielobjekt ein Haus, das zu überwachen ist. Die Anzeige eines Buchstabens kann interpretiert werden als Person, die an einem Fenster des Hauses erscheint. Unterschiedliche Buchstaben repräsentieren unterschiedliche Personen, 'e' repräsentiert die Zielperson. Bei Testaufgabe 2 repräsentiert das Zielobjekt ein fahrendes Fahrzeug, das überwacht werden muss. Die Anzeige eines Buchstabens kann interpretiert werden als Person, die an einem Fenster des Fahrzeugs erscheint. Unterschiedliche Buchstaben repräsentieren wieder unterschiedliche Personen, 'e' die Zielperson.

Da die Untersuchung die Selektion von Ereignissen ohne manuelles Eingreifen des Benutzers, d.h. ohne bewusste Aktion untersucht, wurde die Dauer der Testaufgaben insgesamt auf unter 20 min begrenzt. Denn wenn Personen länger eine Überwachungsaufgabe durchführen, kann es passieren, dass sich ihre Vigilanz reduziert [Tei74]. Dem wollten wir vorbeugen, um Situationen, bei denen der Beobachter ein Ereignis übersehen hat, zu unterscheiden von unserem Untersuchungsgegenstand von Situationen, in denen eine Ereignis-Lokalisierung ohne manuelle Intervention geschieht. Um die Aufmerksamkeit der Versuchspersonen festzustellen, mussten sie die Anzahl der Eintritte des Zielbuchstabens 'e' während der Durchführung der Testaufgaben zählen.

Testaufgabe 1 dauerte 8:30 min und umfasste 72 Buchstaben, davon 24-mal den Zielbuchstaben 'e'. Testaufgabe 2 dauerte 9:30 min und umfasste 63 Buchstaben, davon 23-mal den Zielbuchstaben 'e'.[1] Die Instruktion der Versuchspersonen war bei beiden Testaufgaben, das Zielobjekt kontinuierlich zu beobachten und die 'e' zu zählen. Um Zählfehler gering zu halten, wurde die Anzahl alle 90 Sekunden für den letzten Beobachtungszeitraum über eine automatisch aufpoppende Dialogbox abgefragt.

[1] Das Verhältnis von Zielreiz und Standardreiz im Oddball-Paradigma beträgt normalerweise 1:9 (vgl. S.85), bei uns beträgt es 1:3. Diese Verschiebung des Zielreizes wurde vorgenommen, um ausreichend viele Zielreize als Datenbasis für die Analyse zur Verfügung zu haben. Unter den Bedingungen, dass die Testaufgaben insgesamt unter 20 min dauern sollten, jeder Buchstabe 2 Sekunden angezeigt werden sollte und zwischen je zwei Buchstaben 2 Sekunden liegen sollten, ergab sich dieses Verhältnis von Zielreiz und Standardreizen.

5.2.2.2 Versuchsaufbau, Versuchspersonen und Versuchsablauf

Die Versuchssoftware zur Präsentation der Testaufgaben wurde als JAVA-Anwendung unter Windows 7 implementiert. Die Animationen wurden mithilfe der auf der Programmiersprache JAVA aufbauenden Software „Processing“ implementiert. Der Versuchsaufbau ist wie in Abb. 4.3.

Zur Blickerfassung wurde der Tobii X60 ohne Kopfstabilisierung verwendet (vgl. Abschnitt 4.2). Zur Erfassung des EEG wurde ein actiCHamp Recorder wie in Abschnitt 4.4 beschrieben eingesetzt. Die Synchronisation zwischen Versuchsaufgabe, Blickerfassung und EEG erfolgte über eine Fotodiode, die am Monitor über einer Markierungsbox befestigt wurde. Die Markierungsbox änderte ihre Farbe zwischen schwarz (kein Buchstabe) und weiß (Buchstabe). Auf diese Weise wurde Synchronisierung der Datenströme auf Framelevel sichergestellt.

Die 10 Versuchspersonen (9 männlich, 1 weiblich; Alter zwischen 21 und 42, Altersdurchschnitt 25,9 Jahre) verfügten alle über normale oder auf normal korrigierte Sicht, fünf trugen eine Brille. Alle waren Studenten oder Kollegen, also keine Videoauswerteexperten. Drei hatten bereits einmal zuvor einen Remote-Eyetracker zur Blickinteraktion genutzt, keiner hatte Erfahrung mit EEG. Eine weitere Versuchsperson wurde von der Datenanalyse ausgeschlossen, weil einige EEG-Kanäle invalide gemessen hatten.

Der Versuchsablauf war wie folgt. Nach der Applikation der EEG-Kappe führten die Versuchspersonen eine Standard-9-Punkte-Kalibrierung durch. Es folgte eine kurze Einführung in das Experiment mit einer Instruktion zur Aufgabendurchführung: „Bitte fixieren Sie das Zielobjekt und zählen Sie, wie oft der Buchstabe ’e’ erscheint“. Danach folgte ein kurzes Training beider Testaufgaben (Dauer je 1 min mit 4-mal Distraktor-Buchstaben und 3-mal Zielbuchstabe). Danach absolvierten die Versuchspersonen eigenständig erst Testaufgabe 1 und dann Testaufgabe 2.

5.2.2.3 Forschungsfragen, Metriken und Ergebnisse

Dieses Experiment adressierte folgende Forschungsfrage:

- Frage 1: Mit welcher Trefferquote und räumlichen Genauigkeit sowie nach welcher Zeitdauer ist die Ereignis-Lokalisation mit Blick+EEG möglich?

Im Rahmen der vorliegenden Arbeit wurde die Analyse der Blickdaten durchgeführt, die zum Gesamtsystem die *räumliche Lokalisation der Ereignisse* beiträgt.

Um die räumliche Lokalisation zu bewerkstelligen, musste die Fixationsposition auf dem Zielobjekt bestimmt werden. Dazu wurden zunächst die Fixationen aus dem Blickdatenstrom mithilfe des I-DT-Algorithmus extrahiert ($d_{max} = 1°, t_{min} = 100\ ms$) [Sal00], vgl. Abschnitt 2.4.1, S.62.

Für eine erfolgreiche räumliche Ereignis-Lokalisation müssen die Fixationen zwei Kriterien erfüllen. Zum Ersten ist eine eindeutige Zuordnung der Fixationsposition zum Zielobjekt nur dann möglich, wenn die Fixationsdistanz zum Objektmittelpunkt maximal 1,58° beträgt (halber Abstand zwischen benachbarten Objektmittelpunkten im statischen Fall bei Testaufgabe 1). Zum Zweiten muss die räumliche Lokalisierung innerhalb des Zeitraums, den die zeitliche EEG-Lokalisierung benötigt (s.u.), erfolgen; d.h. Fixationen müssen 500 ms nach Beginn der Buchstabenanzeige abgeschlossen sein. Da als minimale Fixationsdauer 100 ms gefordert sind, muss der Fixationsstart (Fixation Onset Time *FOT*) in ein Zeitintervall zwischen 0 und 400 ms nach Beginn der Buchstabenanzeige fallen.

Um die Ergebnisse der beiden Testaufgaben vergleichbar zu machen, wurden die Festlegungen der Distanz < 1,58° sowie die FOT < 400 ms auch für Testaufgabe 2 genutzt. Ähnlich wie die Machbarkeitsuntersuchung aus Abschnitt 5.2.1 werden also Zielobjekte einer effektiven (selektierbaren) Größe von ca. 3° betrachtet.

Tabelle 5.10 zeigt die Ergebnisse für beide Testaufgaben. Die Trefferquote ist für das statische Zielobjekt deutlich besser als für das bewegte, sowohl für die

Lokalisation von Distraktor-Buchstaben als auch des Zielbuchstaben 'e'. Die Distanzen sowie die FOT sind für die Distraktor-Buchstaben und den Zielbuchstaben 'e' sehr ähnlich.

In den Beobachtungszeiten zwischen den Buchstabenanzeigen entspannten die Versuchspersonen ihre Aufmerksamkeit etwas. Denn jetzt betrug die Distanz 1,13 ± 0,56° beim statischen Zielobjekt und 1,59 ± 1,13° beim bewegten Zielobjekt. Somit lag während der gesamten Beobachtungszeit das Zielobjekt innerhalb eines Bereichs, der von der parafovealen Sicht von 5° erfasst wird [Irw92].

Tabelle 5.10: Räumliche Ereignis-Lokalisation für Testaufgabe 1 und 2: **Trefferquote Tq** in Prozent, **räumliche Genauigkeit Gen** (Fixationsdistanz zum Zielobjektmittelpunkt) in ° Sehwinkel und **FOT** in ms als Mittelwert (± 1 Standardabweichung).

	Testaufgabe 1 (statisch)			Testaufgabe 2 (bewegt)		
	Tq	Gen	FOT	Tq	Gen	FOT
Distraktoren {h,n,p,c}	92,6	0,82 (0,32)	203 (129)	81,9	0,78 (0,40)	217 (132)
Zielbuchst. 'e'	94,2	0,81 (0,32)	195 (127)	84,3	0,82 (0,38)	221 (132)

Die Ergebnisse der *zeitlichen Ereignislokalisierung* sind im Detail in [Hil14b] beschrieben. Dabei sind zwei Klassifikationsaufgaben zu lösen; die erste unterscheidet zwischen *Buchstabe* und *Kein Buchstabe*, die zweite zwischen *Zielbuchstabe e* und *Distraktor-Buchstabe*.

Dafür werden die EEG-Daten entsprechend dem visuellen Stimulus segmentiert und gelabelt. Segmente für *Buchstabe* sind 500 ms lang und beginnen mit der Anzeige eines Buchstaben. Segmente für *Kein Buchstabe* beginnen 1 Sekunde vor einem Ereignis, sodass sie weder mit der Buchstabenanzeige davor noch mit der danach überlappen.

Für die Analyse werden alle 28 Elektroden genutzt. Als Klassifikationsmodell wurde eine Support Vector Machine mit Radial Basis Function kernels verwendet [Put13, Hil14b].

Die Ergebnisse der zeitlichen Lokalisierung im EEG betrugen für die Klassifikation *Buchstabe* versus *Kein Buchstabe* 83,0% für das statische Zielobjekt in Testaufgabe 1 und 80,7% für das bewegte Zielobjekt in Testaufgabe 2. Die Klassifikation *Zielbuchstabe e* versus *Distraktor-Buchstabe* gelang mit 63,6% für das statische Zielobjekt, und mit 62,1% für das bewegte Zielobjekt.

Die *Zählergebnisse*, die die Versuchspersonen angaben, waren sehr genau. Bei 430 Buchstaben, verpassten sie nur 3 bei Testaufgabe 1 und 6 bei Testaufgabe 2. Offenkundig fand während der Aufgabendurchführung keine Verschlechterung der Vigilanz statt.

Bezüglich der **Forschungsfrage** ergeben sich damit folgende Antworten:

Die Ergebnisse zeigen, dass auch für abstrakte Paradigmata, die einfache Überwachungsaufgaben simulieren, die räumlich-zeitliche Ereignis-Lokalisierung mit Blick+EEG möglich ist.

Für die Lokalisierung von *Buchstabe* versus *Kein Buchstabe* liegt die zeitliche Akkuratheit für den Fall des statischen Zielobjekts bei 83,0%, die räumliche bei 92,6%. Unter Annahme der Unabhängigkeit beider Ergebnisse ergibt sich eine kombinierte räumlich-zeitliche Lokalisierung mit 76,9% (83,0% ∗ 92,6%). Für den Fall des bewegten Zielobjekts liegt die zeitliche Akkuratheit bei 80,7%, die räumliche bei 81,9%, die kombinierte räumlich-zeitliche Lokalisierung liegt also bei 66,1%.

Die Lokalisierung von *Distraktor-Buchstabe* versus *Zielbuchstabe e* gelingt mit 60,0% (63,3% ∗ 94,2%) für das statische Zielobjekt und mit 52,4% (62,1% ∗ 84,3%) für das bewegte Zielobjekt.

Mit Blick+EEG ist eine zeitliche Lokalisierung eines Ereignisses mit einer Latenz von 500 ms nach Eintrittszeitpunkt des Ereignisses möglich. Die räumliche Komponente liefert die FOT mit ca. 200 ± 130 ms für das statische Zielobjekt, und mit ca. 220 ± 130 ms für das bewegte Zielobjekt.

5.2.2.4 Fazit

Die räumlich-zeitliche Lokalisierung von *Buchstabe* versus *Kein Buchstabe* ist eine ähnliche Aufgabe wie die Klassifikation von *trial* versus *Nicht-trial* in Abschnitt 5.2.1. Die hier erzielten Ergebnisse von 76,9% (statisches Zielobjekt) und 66,1% (bewegtes Zielobjekt) stimmen gut mit den dortigen Ergebnissen von 78,5% bzw. 64,5% überein.

Die schwierigere räumlich-zeitliche Lokalisierung von *Distraktor-Buchstabe* versus *Zielbuchstabe e* gelingt mit 60,0% für das statische Zielobjekt und 52,4% für das bewegte Zielobjekt.

Die räumlich-zeitliche Lokalisierung gelingt innerhalb von 500 ms, sowohl für das statische als auch für das bewegte Zielobjekt. Dies ist vergleichbar mit dem Ergebnis der ersten Blick+EEG-Studie (Abschnitt 5.2.1).

5.3 Fazit

Zum Zwecke der Identifikation geeigneter Blickinteraktionstechniken für die Bewegtobjektselektion wurden die vier im Konzept (Abschnitt 3.1) als vielversprechend vorgeschlagenen Interaktionstechniken in Nutzerstudien experimentell untersucht. Tabelle 5.11 gibt eine Übersicht über die Versuchsbedingungen der insgesamt sechs Untersuchungen.

Da die einzelnen Untersuchungen mit einer eher kleinen Anzahl Versuchspersonen (zwischen 18 und 4) durchgeführt wurden, ist ihre statistische Aussagekraft limitiert und die Ergebnisse sind demzufolge als vorläufig zu betrachten.

Tabelle 5.11: Versuchsdesigns der sechs Untersuchungen.

Studie	Versuchsdesign										
	VP	Trials	Zielobjekte		Interaktionstechniken						
			Größe	Geschw.	**M**	**BT**	**BFT**	**EEG**	**MAGIC pointing**		
	N	m	°	°/s					**lib**	**kon**	**Button**
[Hil13a] S.132	18	91	4,01	2,88; 3,05; 3,20; 3,37	×	×			×		
[Hil14a] S.144	12		0,71		×					×	×
[Hil16b] S.158	12	240-720	4,75	4,8; 8,3; 11,9; 15,5; 19	×	×	×				
Längssch. S.173	4	240	4,75	4,8; 8,3; 11,9; 15,5	×	×					
		80	4,75	1,86							
		684	4,75	2,38							
		6 min	4,75 1,43	1,40 bis 4,66							
[Put13] S.196	11	64	2,38	-				×			
				2,88; 3,37							
[Hil14b] S.208	10	72	3,16	-				×			
		63		1,66; 2,38							

Blick+Taste wurde in drei Untersuchungen mit insgesamt 34 Versuchspersonen evaluiert (Abschnitte 5.1.1, 5.1.3 und 5.1.4). Blick+Taste zeigte sich als geeignete Alternative für die Mauseingabe, denn sie zeigte bei ähnlicher Effektivität (Selektionsfehlerquote) eine deutlich bessere Effizienz in Form einer erheblich kürzeren Selektionszeit.

Abb. 5.33 zeigt die Ergebnisse der Selektionsfehlerquote als Funktion der Objektgeschwindigkeit (Objektgröße 4-5°); für die Längsschnittstudie sind jeweils die Ergebnisse des Baseline-Tests (LS-<1/2/3>-B) und des finalen Tests (LS-<1,2,3>-F) nebeneinander dargestellt, um den Lernerfolg zu veranschaulichen. Man sieht, dass mit zunehmender Objekt-Geschwindigkeit die Selektionsfehlerquote steigt; dies gilt gleichermaßen für die Ergebnisse des finalen Tests der Längsschnittstudie sowie für alle übrigen Ergebnisse ohne Langzeit-Training. Im Vergleich zur Mauseingabe sind die Selektionsfehlerquoten für Blick+Taste stets vergleichbar gut, in wenigen Fällen sogar signifikant besser. Signifikant bessere Ergebnisse für die Effektivität in Form einer geringeren Anzahl Kollisionen zeigten sich zudem bei der Längsschnittstudie für Testaufgabe 4 (Air Traffic Control) für beide Objektgrößen.

Abb. 5.34 zeigt die Ergebnisse der Selektionszeit als Funktion der Objektgeschwindigkeit. In fast allen Fällen ist sie für Blick+Taste signifikant kürzer als für die Mauseingabe[1]. Die Ergebnisse für Blick+Taste sind für unterschiedliche Geschwindigkeiten ähnlich und liegen im Mittel um 500 ms. Bei [Hil13a] liegt die Selektionszeit etwa 1,5-mal so hoch. Bei dieser Aufgabe konnten Zielobjekte überall auf dem 17-Zoll-Monitor auftreten, sodass die Distanzen, die für eine Selektion zu überbrücken waren, deutlich größer waren als bei den anderen Aufgaben. Dies schlug sich noch deutlicher in den Ergebnissen der Mauseingabe nieder, wo die Selektionszeiten im Vergleich zu den anderen Aufgaben mehr als doppelt so lang waren.

[1] Nicht dargestellt sind Versuchsaufgabe 2 und 4 der Längsschnittstudie. Bei 2 war der Selektionsaufgabe eine Suchaufgabe vorangestellt, sodass die Selektionszeiten deutlich länger waren; Blick+Taste war im Mittel schneller als Mauseingabe, jedoch nicht signifikant. Bei 4 wurde keine Selektionszeit gemessen.

Die kürzesten Selektionszeiten wurden dann erzielt, wenn die Sichtbarkeit der Zielobjekte sehr kurz war. Dies ist vor allem bei Versuchsaufgabe 3 der Längsschnittstudie (LS-3-B bzw. LS-3-F in der Abbildung) zu sehen, wo deutlich unter 400 ms erzielt wurden; die Objektgeschwindigkeit ist hier mit 2,38°/s vergleichsweise langsam, aber es mussten viele Objekte in kurzer Zeit selektiert werden. Dass kürzere Sichtbarkeit die Versuchspersonen veranlasste, sich noch mehr um schnelle Selektion zu bemühen, ist noch deutlicher an den Ergebnissen der Mauseingabe zu sehen.

Die Längsschnittstudie mit 4 Versuchspersonen (Abschnitt 5.1.4) zeigte eine Stabilisierung der Lernkurve für fast alle Versuchsaufgabenbedingungen nach 1 bis 2 Monaten, bei den schwierigsten (1: Circle Select 15.5°/s sowie 3: Catch 'em All 18 Zielobjekte) nach 3 Monaten, wobei hier geringfügige Verbesserungen über die gesamte Trainingszeit von 6 Monaten zu beobachten waren. Die hohe Zufriedenstellung mit Blick+Taste zeigte sich in den sehr guten subjektiven Bewertungen, die für die überwiegende Anzahl Merkmale besser ausfiel als für die Mauseingabe. Für das wichtige Merkmal der Ermüdung der Augen reduzierte sich die subjektiv empfundene Ermüdung nach 6 Monaten Training signifikant (Verbesserung von 3,0 ± 0,0 auf 4,8 ± 0,5 auf einer 7-Punkte-Skala, vgl. Tabelle 5.7). In der initialen Studie (Abschnitt 5.1.1) votierten 16 der 18 Versuchspersonen für Blick+Taste als bevorzugte Interaktionstechnik; diese Frage wurde in den anderen beiden Studien (Abschnitte 5.1.3 und 5.1.4) nicht gestellt.

Die Tatsache, dass die guten Ergebnisse für Blick+Taste überwiegend (30 von 34 Versuchspersonen) bei Querschnittstudien (15 Versuchspersonen mit geringer Erfahrung mit Blick+Taste, 15 ohne jede Erfahrung mit blickbasierter Interaktion) erzielt wurde, lässt darauf schließen, dass die Anlernzeit für Blick+Taste kurz ist. Die Längsschnittstudie zeigte, dass Training die Leistung noch verbessern kann.

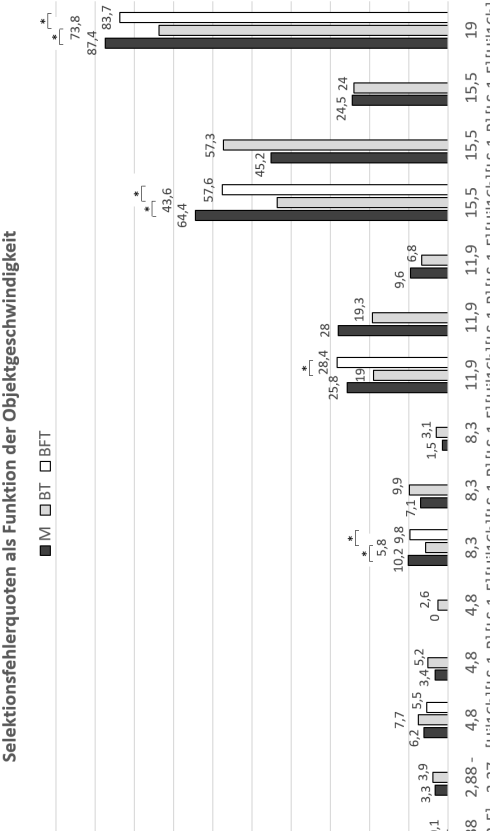


Abbildung 5.33: Selektionsfehlerquoten (Mittelwerte) als Funktion der Objektgeschwindigkeit. Signifikante Unterschiede sind markiert mit *. Die Ergebnisse der Längsschnittstudie sind unterstrichen. Die Ergebnisse aus Abschnitt 5.1.3 wurden von Selektionstrefferquoten in Selektionsfehlerquoten umgerechnet; zudem sind die Ergebnisse für die lineare und wellenförmige Bewegung gemittelt aufgeführt (vgl. Abb. 5.18).

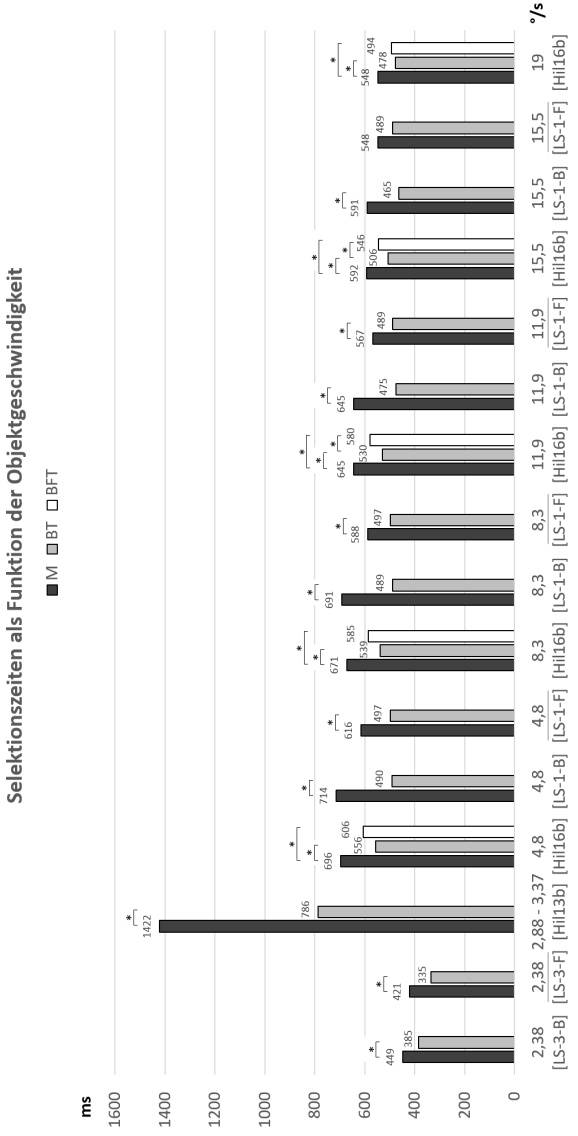


Abbildung 5.34: Selektionszeiten (Mittelwerte) als Funktion der Objektgeschwindigkeit. Signifikante Unterschiede sind markiert mit *. Die Ergebnisse der Längsschnittstudie sind unterstrichen. Die Ergebnisse aus Abschnitt 5.1.3 sind gemittelt über nah/entfernt startende Objekte bzw. lineare und wellenförmige Bewegung aufgeführt (vgl. Abb. 5.20).

Blick+Fußtaste wurde mit insgesamt 12 Versuchspersonen evaluiert (Abschnitt 5.1.3). Wie für Blick+Taste und Mauseingabe steigt auch für Blick+Fußtaste die Selektionsfehlerquote mit zunehmender Objektgeschwindigkeit (Abb. 5.33). Im Vergleich zur Mauseingabe ist sie stets vergleichbar gut; im Vergleich zu Blick+Taste ist sie für Geschwindigkeiten ab 8,3°/s signifikant schlechter. Die Selektionszeit ist stets signifikant kürzer als die der Mauseingabe; sie ist geringfügig länger im Vergleich zu Blick+Taste (signifikant für 8,3°/s und 15,5°/s) und beträgt im Mittel ca. 540 ms. Auch für Blick+Fußtaste war die Anlernzeit kurz, da die Leistungen von Versuchspersonen ohne jede Erfahrung mit dieser Interaktionstechnik erbracht wurden. Das erste Fazit für Blick+Fußtaste fällt also positiv aus, wenngleich die Versuchspersonen mit Blick+Taste noch bessere Leitung erbrachten.

Um auch Blick+Fußtaste als Alternative zur Mauseingabe bei der Bewegtobjektselektion empfehlen zu können, sind weitere Untersuchungen mit mehr Versuchspersonen und anhand einer größeren Variationsbreite von Versuchsaufgaben erforderlich. Außerdem muss eine subjektive Bewertung von Blick+Fußtaste erfolgen, bei der insbesondere der Aspekt der Ermüdung des Fußes betrachtet werden muss; denn dies könnte mit zunehmender Interaktionsdauer möglicherweise ein Störfaktor sein. Sollte dies nicht der Fall sein, kann Blick+Fußtaste eine Alternative zur Mauseingabe in Fällen sein, in denen manuelle Interaktion nicht möglich ist.

MAGIC pointing wurde in drei Varianten untersucht, Tabelle 5.12 stellt die Ergebnisse zusammen. Im Gegensatz zu Blick+Taste und Blick+Fußtaste wird bei MAGIC pointing das feine Zeigen mit der Mauseingabe bewerkstelligt; die räumliche Zeigegenauigkeit liegt daher grundsätzlich (bei unbewegten Selektionsobjekten) in der Größenordnung einzelner Pixel.

Keine der MAGIC pointing-Varianten erzielte eine bessere Effektivität oder Effizienz als die Mauseingabe. *MAGIC pointing liberal* (proaktiv, Mauszeiger permanent an der Blickposition visualisiert), das mit 18 Versuchspersonen an 4° großen Zielobjekten evaluiert wurde, erzielte sowohl eine signifikant höhere Selektionsfehlerquote als auch eine signifikant längere Selektionszeit; die Interaktionstechnik wurde von keiner Versuchsperson als die bevorzugte genannt.

Tabelle 5.12: Versuchsdesigns und Ergebnisse für die Untersuchungen zur Bewegtobjektselektion an abstrakten visuellen Stimuli mit MAGIC pointing. Versuchspersonen (VP) mit Anzahl *N*, Trials mit Anzahl m, Zielobjekte (ZO) mit den Eigenschaften Größe und Geschwindigkeit. Selektionsfehlerquote in Prozent und Selektionszeit in ms, jeweils als Mittelwerte; signifikant bessere Ergebnisse sind **fett** gedruckt.

Studie	Versuchsdesign				Ergebnisse							
	VP	Trials	ZO-Grö.	ZO-Gesch.	Selektionsfehlerquote				Selektionszeit			
	N	m	°	°/s	%				ms			
					M	**MAGIC pointing**			**M**	**MAGIC pointing**		
						lib	kon	Button		lib	kon	Button
[Hil13a] S.132	18	91	4,01	2,88-3,37	**3,3**	8,3	-	-	**1422**	1681	-	-
[Hil14a] S.144	12		0,71		**1,1**	-	7,5	**2,4**	**1744**	-	2516	**2081**

MAGIC pointing konservativ (Mauszeiger springt bei Mausbewegung an die Blickposition), das mit 12 Versuchspersonen an 0,71° großen Zielobjekten evaluiert wurde, erzielte ebenfalls sowohl eine signifikant höhere Selektionsfehlerquote als auch eine signifikant längere Selektionszeit; immerhin 25% der Versuchspersonen nannten sie als ihre bevorzugte Interaktionstechnik.

MAGIC Button wurde im Rahmen der vorliegenden Arbeit als neue Variante eingeführt. Hier steuert der Benutzer durch rechten Mausklick, dass der Mauszeiger an die Blickposition springt. Sie wurde ebenfalls mit 12 Versuchspersonen an 0,71° großen Zielobjekten evaluiert. Hier fällt die Bilanz im Vergleich zur Mauseingabe etwas besser aus. Die Selektionsfehlerquote ist vergleichbar gut wie die der Mauseingabe und signifikant besser als die für MAGIC pointing konservativ. Auch die Selektionszeit ist signifikant besser als die für MAGIC pointing konservativ; der Unterschied zur Mauseingabe ist nicht signifikant. Die Zufriedenstellung mit MAGIC Button ist vergleichbar mit der Mauseingabe. MACIC Button wurde zudem von 8 der 12 Versuchspersonen (66%) bevorzugt.

Blick+EEG wurde in zwei Untersuchungen mit insgesamt 21 Versuchspersonen evaluiert (Abschnitte 5.2.1 und 5.2.2). Im Gegensatz zu den anderen betrachteten Interaktionstechniken, bei denen der Benutzer aktiv und bewusst interagiert, ist Blick+EEG eine passive Eingabemethode, die keinerlei manuelle oder bewusste Aktion des Benutzers erfordert. Blick+EEG als implizite Interaktionstechnik für Selektionsoperationen zu nutzen, war zum Zeitpunkt der Untersuchungen Neuland; Blick+EEG zur Objektselektion existiert auch heute noch nicht als Eingabemethode.

Die Untersuchungen im Rahmen der vorliegenden Arbeit dienten also als erster Schritt, um die prinzipielle Machbarkeit von Blick+EEG als Selektionsalternative festzustellen. Dazu wurden Blick- und EEG-Daten während abstrakter Testaufgaben aufgezeichnet. Eine Offline-Analyse bestimmte dann, mit welcher Qualität die räumlich-zeitliche Lokalisation von Ereignissen (manifestiert durch Objekte mit Zielobjektverhalten) gelingt. Wie bei den Untersuchungen der expliziten Interaktionstechniken wurden dafür Trefferquoten

und Selektionszeiten bestimmt. Aufgrund des Pilotcharakters der Untersuchungen wurden neben bewegten Zielobjekten auch statische Zielobjekte betrachtet.

Die Analysen ergaben, dass die Selektionszeit (für statische wie bewegte Zielobjekte) wie bei Blick+Taste und Blick+Fußtaste bei ca. 500 ms liegt. Mit Blick+EEG wird der Benutzer jedoch von jeglicher motorischen Aktion der Extremitäten sowie bewusster kognitiver Aktion entlastet.

Rechnet man die erzielten Trefferquoten in Fehlerquoten um, so erhält man für bewegte Zielobjekte der Größe 3° und Geschwindigkeiten zwischen 1,66°/s und 3,37°/s je nach Komplexität der Erkennung eines Objektes als Zielobjekt auf Basis der P300 zwischen 33,9% und 47,6%. Dies ist erheblich schlechter als die Selektionsfehlerquoten mit Blick+Taste (oder Mauseingabe) bei vergleichbaren Geschwindigkeiten. Dabei ist jedoch zugunsten von Blick+EEG zu erwähnen, dass die Objektgrößen in den Untersuchungen mit expliziten Interaktionstechniken mit > 4° etwas größer waren.

Für die statischen Zielobjekte betrugen die Fehlerquoten je nach Komplexität der Erkennung eines Objektes als Zielobjekt auf Basis der P300 zwischen 21,5% und 40%.

Die Anlernzeit bei Blick+EEG war gering, denn die Versuchspersonen mussten im Training nur die Testaufgaben einüben und keine bestimmte Interaktionstechnik. Ein großer Nachteil von Blick+EEG ist aktuell, dass die erzielte Qualität der EEG-Messung nur mit sehr hochwertigen und hochpreisigen EEG-Geräten funktioniert, die das Tragen einer EEG-Kappe erfordern. Dies bedeutet einerseits hohe zeitliche Vorbereitungskosten bei der Anbringung der EEG-Kappe sowie große Belastung beim Tragen.

Um eine Alternative zur Mauseingabe für die Bewegtobjektselektion zu sein, müssen verschiedene Aspekte bei Blick+EEG verbessert werden. Zum einen betrifft dies die Selektionsfehlerquote, die durch Verbesserung der Ereigniserkennung erheblich verringt werden muss. Zum anderen betrifft dies die Hardware zur EEG-Messung, die schnellere Rüstzeiten und besseren Tragekomfort verlangt.

6 Blickbasierte Interaktion bei der Videobildauswertung

Dieses Kapitel betrachtet Möglichkeiten für die Nutzung blickbasierter Interaktion bei der Videobildauswertung. Es gliedert sich in drei Unterabschnitte.

Abschnitt 6.1 komplettiert die Untersuchungen im Rahmen von Beitrag 1 der vorliegenden Arbeit (vgl. Abschnitte 1.3.1 und 3.1). In Kapitel 5 hatte sich Blick+Taste als leistungsfähige Alternative zur Mauseingabe erwiesen. Jetzt wird überprüft, ob sich dies für die Bewegtobjektselektion in Überflugvideos bestätigen lässt.

Abschnitt 6.2 beschreibt die Untersuchungen im Rahmen von Beitrag 2 der vorliegenden Arbeit (vgl. Abschnitte 1.3.2 und 3.2). Die blickbasierte Interaktion bei automatischen Bildanalyseverfahren wird für zwei automatische Bildanalyseverfahren zur Aktivitätsdetektion betrachtet, einmal für ein Verfahren zur Bewegungsdetektion, einmal für ein Verfahren zum Einzelobjekttracking. Auch diese Untersuchungen erfordern Bewegtobjektselektion und vergleichen die Benutzerleistung mit Blick+Taste und Mauseingabe.

Abschnitt 6.4 beschreibt die Untersuchungen im Rahmen von Beitrag 3 der vorliegenden Arbeit (vgl. Abschnitte 1.3.3 und 3.3), die blickbasierte Klassifikation der Benutzertätigkeit bei der Bildfolgenanalyse.

Die Versuchssysteme der Untersuchungen in Abschnitt 6.1 und Abschnitt 6.2 basieren alle auf dem ABUL-System (vgl. Abschnitt 1.2, S. 4). Da das ABUL-System über weit mehr Funktionsumfang verfügt als für die Durchführung unserer Experimente erforderlich, wurde von den Kollegen des ABUL-Teams für die unterschiedlichen Experimente jeweils ein ABUL-Derivat mit eingeschränktem, auf das jeweilige Experiment zugeschnittenem Funktionsumfang

zur Verfügung gestellt. Die Änderungsdetektionsverfahren sind jeweils als Subsysteme enthalten und in den Abschnitten unten beschrieben. Zudem wurde in die Versuchssysteme der Tobii X60-Eyetracker für die Blickeingabe integriert.

6.1 Bewegtobjektmarkierung

In Ergänzung der Untersuchungen aus Kapitel 5 mit abstrakten visuellen Stimuli wurden zwei Untersuchungen zur Bewegtobjektselektion an Überflug-Videodatenmaterial durchgeführt, eine Pilotstudie mit Nicht-Videoexperten (Abschnitt 6.1.1) und eine Expertenstudie (Abschnitt 6.1.2). Beide betrachten Blick+Taste und Mauseingabe im Vergleich.

Als Ausprägung von Bewegtobjektselektion wird die Bewegtobjektmarkierung betrachtet, eine sehr häufige Interaktionsaufgabe bei der Bildfolgenanalyse. Der Videoanalyseexperte markiert dabei relevante Objekte mit einem Rahmen, damit andere Stellen auf diese Information zurückgreifen können (z. B. zur Dokumentation, Kommunikation im Team, Entscheidungsfindung oder zur Archivierung [Brü12]).

Nach Stand der Technik klickt der Benutzer bei der ABUL-Videoauswertestation mit der linken Maustaste auf das Objekt, wobei die Mausklick-Position die Rahmenmitte definiert. Die Rahmengröße wird vorab so konfiguriert, dass die auftragsgemäß zu erwartenden Zielobjekte komplett darin Platz finden. Bei Zielobjekten unterschiedlicher Größe orientiert sich die Rahmengröße am größten Objekt.

6.1.1 Pilotstudie

Dieses Experiment und seine Ergebnisse wurden bei der SPIE 2013 veröffentlicht [Hil13b].

Es werden zwei Interaktionstechniken verglichen: Blick+Taste und Mauseingabe.

Abbildung 6.1: Visueller Stimulus der Testaufgabe.

Die Selektion mit *Blick+Taste* erfolgt, indem der Benutzer das Objekt anblickt und währenddessen die ENTER-Taste des NumPad drückt.

Die *Mauseingabe* erfolgt als traditionelle Zeige-Klick-Interaktion mit Einzel-Klick der linken Maustaste.

6.1.1.1 Testaufgabe

Als visueller Stimulus für die Testaufgabe wurde eine Videosequenz von 4:30 Minuten Dauer genutzt. Sie entstammt Full Motion Video-Datenmaterial, das von einem UAV (Unmanned Aerial Vehicle) aufgezeichnet wurde. Abb. 6.1 zeigt ein beispielhaftes Einzelbild daraus. Die Videoauflösung beträgt 720 x

576 Pixel. Angezeigt wird das Video mit einer Größe von 27,5 cm Breite und und 22 cm Höhe, sodass ein Bildpixel auf dem Monitor 0,38 mm groß ist.

Die Testaufgabe enthält 32 Zielobjekte (parkende blaue PKW und Panzer, ein Kirchturmdach, ein graues Gebäudedach sowie wenige fahrende PKW und laufende Personen). Die Zielobjektgröße reicht von 5 x 5 Bildpixeln (5 Bildpixel entsprechen 0,17° bzw. 0,19 cm) bis zu 35 x 15 Bildpixeln (35 Bildpixel entsprechen 1,18° bzw. 1,33 cm).

Die Videobildrate beträgt 25 Hz. Das Videomaterial weist permanente Szenenbewegungen von Bild zu Bild auf, ähnlich denen in Abb. 5.3. Die Szenenbewegungen sind überwiegend nur einige Bildpixel bzw. wenige Millimeter groß. Dies resultiert in Objektgeschwindigkeiten im Bereich derer aus Abb. 5.3; für das zugehörige 3-minutige Überflugvideomaterial lagen die Objektgeschwindigkeiten[1] im Mittel ± 1 Standardabweichung bei 90 ± 58 Bildpixel/s (3,09 ± 1,99°/s bzw. 3,42 ± 2,20 cm/s).

Vereinzelt verschiebt sich die Szene aufgrund starker Sensorbewegung ruckartig mit bis zu 40 Bildpixeln zwischen zwei Bildern, was einer Geschwindigkeit von 1.000 Bildpixel/s (34,3°/s bzw. 38,0 cm/s) entspricht.

Für jede einzelne der 32 Selektionsaufgaben wurde der Zeitpunkt der Bewegtobjektmarkierung verbal vom Versuchsleiter getriggert. Dabei benannte dieser zunächst das zu selektierende Zielobjekt, z. B. „Blauer PKW!“. Nach 1 bis 3 Sekunden erging das Kommando „Jetzt!“ und die Versuchsperson musste unmittelbar selektieren.

6.1.1.2 Versuchsaufbau, Versuchspersonen und Versuchsablauf

Das Versuchssystem ist ein Derivat des ABUL-Systems. Angezeigt werden die Trainingsaufgaben und die Testaufgabe auf einem 24-Zoll-Monitor (Auflösung 1920 x 1200 Pixel).

[1] Berechnung mittels über 1 Sekunde akkumulierten Distanzen der Objektmittelpunkte von Bild zu Bild bei 25 Hz Bildrate.

Verwendete Eingabegeräte sind der Tobii X60 und die Comfort-Mouse 2000 for Business 3 (1000 dpi) von Microsoft. Die Blickrohdaten wurden nicht gefiltert, um ohne zusätzliche Latenz mit Blick+Taste interagieren zu können. Der manuelle Tastendruck bei Blick+Taste erfolgte mit der ENTER-Taste des NumPad einer gewöhnlichen Computertastatur.

Der Versuchsaufbau ist wie in Abb. 4.3.

Die Größe des Markierungsrahmens im ABUL-Versuchssystem wurde auf 60 Bildpixel gesetzt (Abb. 6.2). Dies entspricht 2,28 cm auf dem Monitor bzw. 2,06° Sehwinkel, wenn der Benutzer mit 65 cm Abstand der Augen vom Bildschirm entfernt sitzt. Die Größe wurde so gewählt, dass auch die größten Zielobjekte mit etwas Spielraum im Rahmen Platz finden. Außerdem sind ca. 2° die Größe, ab der statische Objekte als robust blickbasiert selektierbar gelten (vgl. z. B. [Ver08]). Dass dies eine angemessene Größe ist, lässt sich auch aus unseren Ergebnissen der Selektionsgenauigkeit aus Kapitel 5 begründen. Für ähnliche Geschwindigkeiten betrug in Abschnitt 5.2.2 für 1,67°/s bzw. 2,38°/s die Selektionsgenauigkeit 0,82 ± 0,38°; in Abschnitt 5.1.3 betrug sie 0,95 ± 0,14° für 4,8°/s. Berücksichtigt man, dass diese Distanz in jede Richtung vom Zielobjektmittelpunkt vorkommen kann, so ergibt sich durch Verdopplung der Selektionsgenauigkeit (Mittelwert + 1 Standardabweichung) ein Bereich von (0,82° + 0,38°) ∗ 2 = 2,40° bzw. (0,95° + 0,14°) ∗ 2 = 2,18°, in den die Selektionspositionen fielen.

An der Untersuchung nahmen 14 Versuchspersonen teil. Bei einer davon schlug die Kalibrierung fehl. Im Rahmen der Nutzerstudie wurden also Daten von 13 Versuchspersonen erfasst (12 männlich, 1 weiblich; Alter zwischen 21 und 34 Jahren, Altersdurchschnitt ± 1 Standardabweichung 27,6 ± 4,1 Jahre. Alle verfügten über normale oder auf normal korrigierte Sicht, vier trugen eine Brille. Alle waren Kollegen oder Studenten, also keine Videoanalyseexperten. Sieben hatten schon einmal einen Eyetracker genutzt, verfügten also über geringe Erfahrung mit der Nutzung von Blick+Taste.

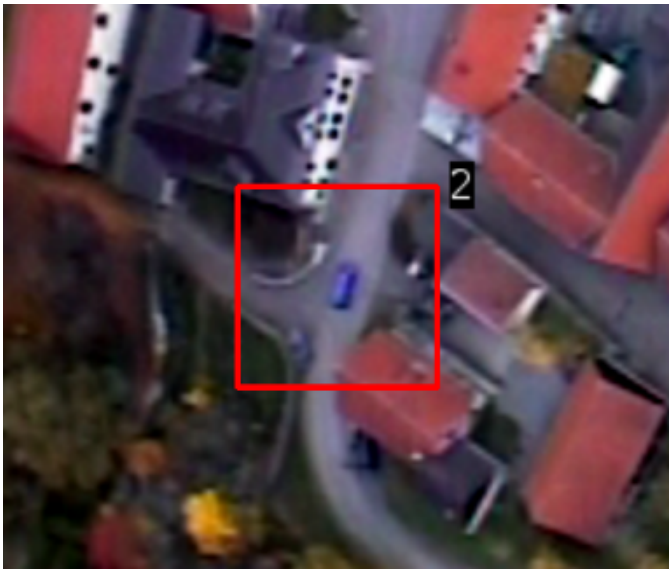


Abbildung 6.2: Bildausschnitt der Testaufgabe mit markiertem Zielobjekt (Blauer PKW).

Das Versuchsdesign nutzte ein Within-Subjects-Design, bei dem jede Versuchsperson die Testaufgabe je einmal mit Blick+Taste und mit Mauseingabe durchführte. Um Ermüdungs- und Lerneffekte zu kontrollieren, wurden die Versuchspersonen in 2 Versuchsgruppen eingeteilt, wobei jede Versuchsgruppe eine der 2 möglichen Technik-Reihenfolgen durchführte.

Der Versuchsablauf war wie folgt. Als Erstes wurde in einer allgemeinen Einführung die zu bewerkstelligende Aufgabe der Bewegtobjektmarkierung erläutert. Dann führten die Versuchspersonen mit jeder der beiden Interaktionstechniken folgenden Ablauf durch: Trainingsaufgabe 1, Trainingsaufgabe 2, Testaufgabe, ISO DIN EN 9241-9-Fragebogen [DIN00]. Die Versuchspersonen wurden instruiert, so schnell und so genau wie möglich zu selektieren. Zu Beginn der Blick+Taste-Sitzung wurde eine Standard-9-Punkt-Kalibrierung des Tobii X60 durchgeführt.

Da alle Versuchspersonen keine Videoauswerteexperten waren, wurde vergleichsweise viel Zeit auf das Training verwendet. Die Gesamtinteraktionszeit mit jeder Interaktionstechnik betrug dadurch 12 Minuten.

Trainingsaufgabe 1 bestand aus einer 3-minütigen Überflug-Videosequenz mit weniger Sensorbewegung als bei der Testaufgaben-Videosequenz. Bei Trainingsaufgabe 1 sollten die Versuchspersonen als Erstes nach Belieben Objekte markieren, um sich mit der Bewegtobjektselektion mit Mauseingabe bzw. mit Blick+Taste vertraut zu machen. Als Zweites sollten die Versuchspersonen auf verbalen Zuruf des Versuchsleiters Objekte markieren, so wie dies auch für die Testaufgabe erfolgen würde.

Trainingsaufgabe 2 nutzte den visuellen Stimulus der Testaufgabe. Hier wurden die Versuchspersonen aufgefordert, Objekte derselben Kategorie zu markieren, die auch in der Testaufgabe vorkommen würden. Durch Trainingsaufgabe 2 sollte sichergestellt werden, dass die Versuchspersonen die Objekttypen bei der Testaufgabe schnell genug erkennen würden, um sie auch schnell selektieren zu können.

6.1.1.3 Forschungsfragen, Metriken und Ergebnisse

Dieses Experiment adressierte folgende Forschungsfragen:

- Frage 1: Wie effektiv sind die Versuchspersonen mit Blick+Taste im Vergleich zur Mauseingabe?
- Frage 2: Wie ist die subjektive Bewertung der Interaktionstechniken?

Hierfür wurden folgende Metriken verwendet:

- Selektionsfehlerquote (Verpasste Zielobjekte): Sie berechnet sich als Anzahl fehlerhafter Selektionen / Gesamtaufgabenumfang (= 32 Objektselektionen). Ein Objekt gilt als erfolgreich selektiert, wenn der Selektionsrahmen eindeutig diesem Objekt zuzuordnen ist und der Rahmen das Objekt beinhaltet oder zumindest berührt. Im letzteren Fall vergrößert sich der selektionssensitive Bereich um die

Objektgröße; für das kleinste Objekt von 0,17° Seitenlänge beträgt er dann 3,15°, für das größte von 1,18° x 0,51° 4,19°.

- Selektionsgenauigkeit: Sie berechnet sich als Abstand (euklidische Distanz) zwischen der Selektionsposition und dem Zielobjektmittelpunkt.

Die Selektionsfehlerquote war für beide Interaktionstechniken ähnlich mit 13,6 ± 7,8 % für die Mauseingabe und 14,2 ± 7,6 % für Blick+Taste. Die beste bzw. schlechteste Selektionsfehlerquote einzelner Versuchspersonen betrug für beide Interaktionstechniken 0% bzw. 25%.

Die Selektionsgenauigkeit war für beide Interaktionstechniken ebenfalls ähnlich. Sie betrug für die Mauseingabe 20 ± 5 Bildpixel (0,69 ± 0,17° bzw. 0,73 ± 0,18 cm auf dem Monitor) und für Blick+Taste 17 ± 5 Bildpixel (0,57 ± 0,17° bzw. 0,62 ± 0,18 cm auf dem Monitor).

Abb. 6.3 zeigt die Ergebnisse der subjektiven Bewertung. Blick+Taste erzielt im Mittel gute bis sehr gute Bewertungen bei der großen Mehrheit der Merkmale und wird fast immer besser oder gleich gut bewertet wie die Mauseingabe[1]. Die subjektive Bewertung der Genauigkeit bestätigt die objektiv gemessenen Ergebnisse der Selektionsgenauigkeit.

Das einzige Merkmal, für das Blick+Taste eine deutlich schlechtere Bewertung als die Mauseingabe erzielt, ist die Ermüdung der Augen[2]. Eine Versuchsperson gab nach dem Experiment an, trotz des Trainings unsicher bei der Eyetracker-Nutzung gewesen zu sein. Überlegungen wie „Wie weit darf ich mich bewegen?“ oder „Sind meine Augen weit genug offen?“ führten dazu, dass sie sich bei Blick+Taste kaum zu bewegen traute und die Interaktion als anstrengend und ermüdend empfand.

Gefragt nach ihrer bevorzugten Interaktionstechnik votierten 2 Versuchspersonen für die Mauseingabe, 11 für Blick+Taste.

[1] Zweiseitige *t*-Tests bei abhängigen Stichproben, $\alpha = 0{,}05$: BT signifikant besser als M mit $p<0{,}001$ für Benutzungsgeschwindigkeit; mit $p<0{,}01$ für Nutzung insgesamt; mit $p<0{,}05$ für Allgemeine Zufriedenheit und Ermüdung Handgelenk.

[2] Zweiseitiger *t*-Test bei abhängigen Stichproben, $\alpha = 0{,}05$: M signifikant besser als BT mit $p<0{,}05$.

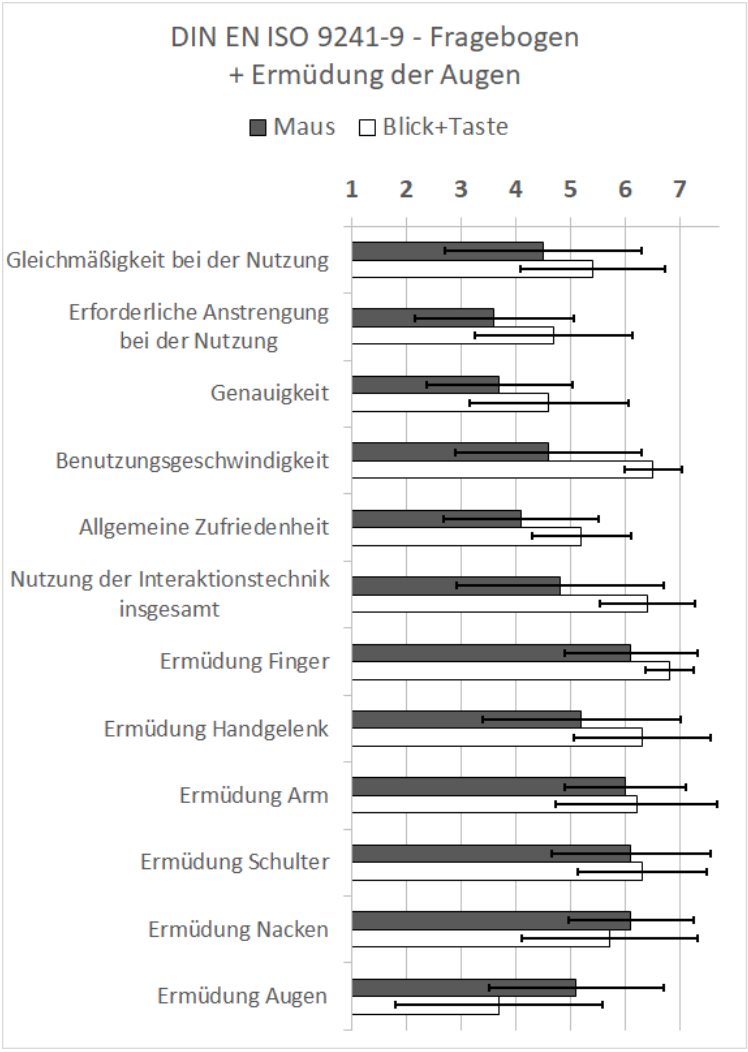


Abbildung 6.3: Subjektive Bewertung.

Bezüglich der **Forschungsfragen** ergeben sich folgende Antworten:

Frage 1: Die Versuchspersonen sind ähnlich effektiv mit Blick+Taste und Mauseingabe.

Frage 2: Bis auf die Ermüdung der Augen erzielt Blick+Taste bessere oder gleich gute subjektive Bewertungen wie die Mauseingabe. Außerdem bevorzugen 82% der Versuchspersonen Blick+Taste.

6.1.1.4 Fazit

Zum einen erzielt Blick+Taste eine ähnliche Effektivität für die Bewegtobjektmarkierung wie die Mauseingabe. Zum anderen war die subjektive Bewertung für fast alle Merkmale gleich gut oder besser und wurde von 82% der Versuchspersonen favorisiert. Allerdings war die subjektive Bewertung der Ermüdung der Augen für Blick+Taste deutlich schlechter als für die Mauseingabe.

6.1.2 Expertenstudie

Dieses Experiment und seine Ergebnisse wurden bei der ETRA 2016 veröffentlicht [Hil16a].

Es baut auf dem Experiment des vorigen Abschnitts auf. Es nutzt denselben Versuchsaufbau und vergleicht ebenfalls Blick+Taste und Mauseingabe für die Bewegtobjektmarkierung in einer Full-Motion-Video-Sequenz.

Die Selektion mit der *Blick+Taste*-Interaktion erfolgt, indem der Benutzer das Objekt anblickt und währenddessen die ENTER-Taste des NumPad drückt. Die *Mauseingabe* erfolgt als traditionelle Zeige-Klick-Interaktion mit Einzel-Klick der linken Maustaste.

Versuchsdesign und Ergebnisanalyse variieren jedoch in folgenden Aspekten:

- Die Versuchspersonen sind Videobildauswerter und somit Experten in Bildfolgenanalyse. Sie sind zudem zahlreicher, sodass ein Vergleich zwischen Brillenträgern und Nicht-Brillenträgern möglich ist.

Abbildung 6.4: Visueller Stimulus der Testaufgabe mit grünem Punkt als visuellem Hinweisreiz auf das Selektionsobjekt.

- Die Testaufgabe umfasst einen größeren Aufgabenumfang an Bewegtobjektmarkierungen.
- Als Trigger für den Start jeder Objektmarkierungsaufgabe wird ein grüner Punkt auf dem Selektionsobjekt visualisiert (Abb. 6.4). Dies ermöglicht zusätzlich die Bestimmung der Selektionszeit.
- Die Selektionsgenauigkeit wird nicht nur für die aufgezeichneten Blickdaten berechnet, sondern auch an einem nachträglich mit einem Blickfilteralgorithmus verarbeiteten Datensatz. Ziel ist herauszufinden, welchen Gewinn Blickfilterung bringt.

6.1.2.1 Testaufgabe

Für die Testaufgabe wurde die Videosequenz aus Abschnitt 6.1.1 auf 7:30 Minuten Dauer verlängert. Insgesamt enthält die Testaufgabe m = 65 Zielobjekte: PKWs, LKWs, Personen und Gebäude. Die Zielobjektgröße reicht von 9 x 9 Bildpixeln bis zu 120 x 100 Bildpixeln. 9 Bildpixel entsprechen 0,34 cm oder 0,30°, 120 Bildpixel entsprechen 4,5 cm oder 4,12°.

27 Objekte sind 1° Sehwinkel oder kleiner, d.h. kleiner als die Unsicherheitsregion, mit der der Eyetracker die Blickrichtung schätzt (vgl. Abschnitt 4.2, Tabelle 4.1). 38 Objekte sind größer als 1° Sehwinkel. Abb. 6.5 zeigt die Objektgrößen der vorkommenden Objekte.

Einige der Objekte (Gebäude und parkende Fahrzeuge) bewegen sich nur mit der Sensorbewegung, andere bewegen sich zusätzlich eigenständig. Auch die verlängerte Videosequenz weist durchgehend permanente Szenenbewegungen von Bild zu Bild auf, die überwiegend nur sehr klein sind, vereinzelt aber ruckartig und stark mit bis zu 40 Bildpixeln (entspricht 1000 Bildpixel/s bei 25 Hz Bildrate) zwischen zwei Bildern variieren (vgl. Abb. 5.3).

Jede Objektmarkierung wird durch einen grünen Kreis (Durchmesser 10 Bildpixel; RGB 0, 255, 0) mit schwarzem Mittelpunkt getriggert. Dieser Hinweisreiz wird 600 ms lang auf dem Objektzentrum visualisiert (vgl. den Einfluss der Darbietungszeit auf die Wahrnehmung eines Reizes auf S. 27). Um saliente Sichtbarkeit auf jedem Objekt sicherzustellen, wurden Größe, Farbe und Anzeigedauer vorab empirisch an zwei Kollegen unserer Abteilung evaluiert.

Der Vorteil des visuellen Triggers durch Hervorhebung im Vergleich zur verbalen Instruktion durch den Versuchsleiter (vgl. Abschnitt 6.1.1) ist, dass (1) der Versuchsleiter nicht im Versuchsraum anwesend sein muss und dass (2) der Beginn jeder einzelnen Selektionsaufgabe zeitlich genau markiert ist. Nachteilig im Sinne hohen Praxisbezugs ist, dass jetzt der Suchvorgang nach dem Zielobjekt wegfällt. Stattdessen durchmustern die Versuchspersonen die Szene in Erwartung des grünen Triggerpunkts und erkennen erst danach, was das Zielobjekt ist. Man könnte den visuellen Trigger jedoch auch so interpretieren, dass er den Zeitpunkt angibt, zu dem der Beobachter das

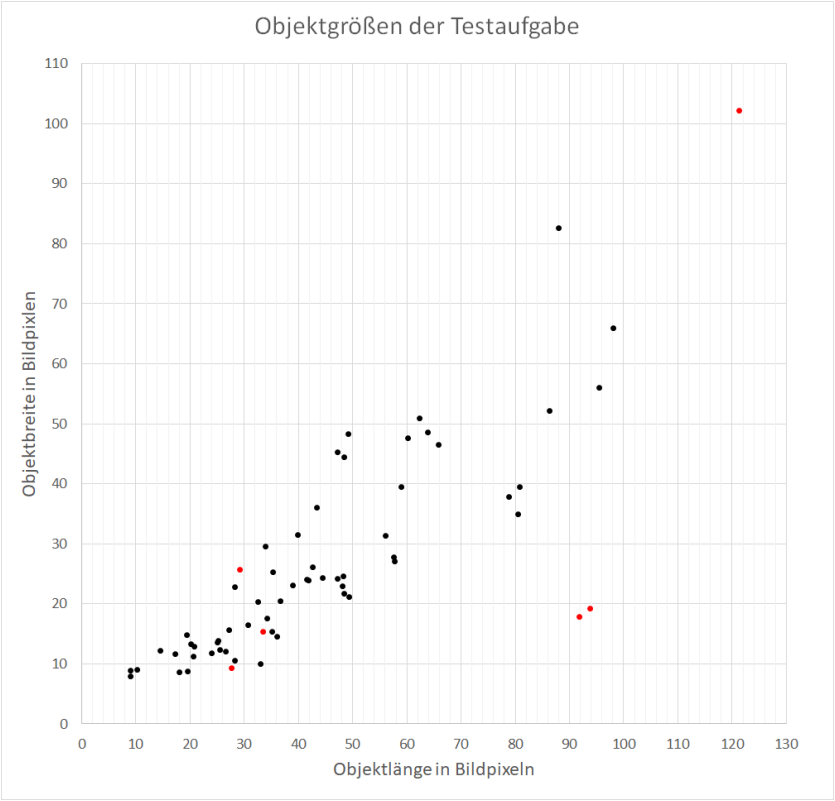


Abbildung 6.5: Objektgrößen der Testaufgabe. Jeder Punkt repräsentiert ein Objekt. Rote Punkte sind die sechs Objekte, die nur sehr kurz sichtbar sind (vgl. Ergebnisse in Tabelle 6.1).

Zielobjekt entdeckt hat, das er daraufhin schnellstmöglich markieren will (vgl. a. die Bemerkungen zu Hervorhebung in Abschnitt 5.1 auf S. 131).

Jede der 65 Einzelaufgaben (engl. *trial*) zur Bewegtobjektmarkierung startet mit der Anzeige des Triggerpunktes auf dem jeweiligen Selektionsobjekt. Jede der Einzelaufgaben endet bestenfalls mit der Markierung des Selektionsobjekts durch die Versuchsperson. Eine Einzelaufgabe wird als fehlgeschlagen

(verpasstes Objekt) gezählt, wenn der nächste Trigger erscheint, bevor die Versuchsperson die Selektion abgeschlossen hat. Zwischen je zwei Triggern liegt stets eine Zeitspanne von mindestens 7 Sekunden Dauer.

6.1.2.2 Versuchsaufbau, Versuchspersonen und Versuchsablauf

Der Versuchsaufbau war derselbe wie beim Experiment aus Abschnitt 6.1.1. Das Versuchssystem ist ein Derivat des ABUL-Systems. Angezeigt werden die Trainingsaufgaben und die Testaufgabe auf einem 24-Zoll-Monitor (Auflösung 1920 x 1200 Pixel).

Verwendete Eingabegeräte sind der Tobii X60 und die Comfort-Mouse 2000 for Business 3 (1000 dpi) von Microsoft. Die Blickrohdaten wurden nicht gefiltert, um ohne zusätzliche Latenz mit Blick+Taste interagieren zu können. Der manuelle Tastendruck bei Blick+Taste erfolgte mit der ENTER-Taste des NumPad einer gewöhnlichen Computertastatur.

Der Versuchsaufbau ist wie in Abb. 4.3. Die Größe des Markierungsrahmens beträgt wie in der Pilotstudie (Abb. 6.2) 60 Bildpixel (2,06° bzw. 2,28 cm auf dem Monitor).

An der Untersuchung nahmen 26 Videoanalyseexperten der Bundeswehr als Versuchspersonen teil (25 männlich, 1 weiblich). Sechs waren unter 30 Jahre alt, 13 zwischen 30 und 45 Jahren und sieben über 45 Jahre. Alle verfügten über normale oder auf normal korrigierte Sicht, zehn trugen eine Brille, eine Versuchsperson trug Kontaktlinsen. Alle waren langjährig erfahrene Benutzer von Desktop-Computern und Computermaus. Alle waren ohne jede Erfahrung mit Eyetrackern oder Blickinteraktion jeglicher Art.

Das Versuchsdesign nutzte ein vollständiges, ausbalanciertes Within-Subjects-Design, bei dem jede Versuchsperson die Testaufgabe je einmal mit jeder der beiden Techniken durchführte. Um Ermüdungs- und Lerneffekte zu kontrollieren, wurden die Versuchspersonen in 2 Versuchsgruppen mit je 13 Versuchspersonen eingeteilt, wobei jede Versuchsgruppe eine der 2 möglichen Technik-Reihenfolgen durchführte.

Der Versuchsablauf war wie folgt. Als Erstes wurde in einer allgemeinen Einführung die zu bewerkstelligende Aufgabe der Bewegtobjektmarkierung erläutert.

Als Zweites folgte ein Trainingsblock für Bewegtobjektselektion, für den die 3-minütige Testaufgabe der initialen Nutzerstudie (Abschnitt 5.1.1) als Trainingsaufgabe genutzt wurde. Zunächst führte jede Versuchsperson eine Standard-9-Punkt-Kalibrierung des Tobii X60, gefolgt von einer Kalibrierungsvalidierung durch. Dann folgte das Training der Blick+Taste-Interaktion, gefolgt vom Training der Mauseingabe.

Als Drittes bzw. Viertes folgten die Testblöcke für die beiden zu testenden Interaktionstechniken. Im Falle der Blick+Taste wurde zunächst die 9-Punkt-Eyetracker-Kalibrierung wiederholt. Dann folgte ein weiteres, kurzes Training mit der Aufgabe, wie sie auch in der Testaufgabe zu bewerkstelligen war. Die dabei verwendete Videosequenz war 2 Minuten lang und entstammte demselben Full Motion Video-Datenmaterial, aus dem auch die Testaufgabe entnommen wurde. Die Versuchspersonen wurden instruiert, die Bewegtobjektmarkierung so schnell und so genau wie möglich durchzuführen. Dann führten die Versuchspersonen die 7:30-minütige Testaufgabe durch. Abschließend bewerteten die Versuchspersonen die Interaktionstechnik subjektiv mithilfe des Fragebogens zur Einzelbewertung zur Erfassung der Zufriedenstellung aus der DIN EN ISO 9241-411 [DIN14], der um das Merkmal „Ermüdung der Augen“ ergänzt wurde (vgl. [Zha07] und Abschnitte 5.1.2, 5.1.4 und 6.1.1 bzw. [Hil13b, Hil14a]).

Versuchsort war für 4 Versuchspersonen das Blicklabor des Fraunhofer IOSB, die Datenerhebung mit den anderen 22 Versuchspersonen wurden an drei verschiedenen Bundeswehrstandorten durchgeführt.

6.1.2.3 Forschungsfragen, Metriken und Ergebnisse

Dieses Experiment adressierte folgende Forschungsfragen:

- Frage 1: Welche Leistung erzielen die Versuchspersonen mit Blick+Taste im Vergleich zur Mauseingabe?

- Frage 2: Gibt es Leistungsunterschiede bei Blick+Taste zwischen Brillenträgern und Nicht-Brillenträgern?
- Frage 3: Gibt es Leistungsunterschiede für sehr kurz sichtbare Objekte (Sichtbarkeit < 3 Sekunden)?
- Frage 4: Gibt es Leistungsunterschiede bei kleinen Objekten (Objektgröße ≤ 1°) verglichen mit großen Objekten (Objektgröße > 1°)?
- Frage 5: Lässt sich die Selektionsgenauigkeit durch die Applikation eines echtzeitfähigen Blickfilter-Algorithmus verbessern?
- Frage 6: Wie ist die subjektive Bewertung der Interaktionstechniken?
- Frage 7: Wie ist insbesondere die subjektive Bewertung des Merkmals „Ermüdung der Augen"? Fällt die Bewertung der Videoauswerteexperten anders aus als die der Laien aus der Pilotstudie (Abschnitt 6.1.1, Abb. 6.3)?

Hierfür wurden folgende Metriken verwendet:

- Selektionsfehlerquote (Verpasste Zielobjekte): Sie berechnet sich als Anzahl fehlerhafter Selektionen / Gesamtaufgabenumfang (= 65 Objektselektionen). Ein Objekt gilt als erfolgreich selektiert, wenn der Selektionsrahmen eindeutig diesem Objekt zuzuordnen ist und der Rahmen das Objekt beinhaltet oder zumindest berührt. Im letzteren Fall vergrößert sich der selektionssensitive Bereich um die Objektgröße; für das kleinste Objekt von 0,30° Seitenlänge beträgt er dann 3,33°, für das größte von 4,12° x 3,43° beträgt er 8,27°.
- Selektionsgenauigkeit: Sie berechnet sich als Abstand (euklidische Distanz) zwischen der Selektionsposition und der Objektposition. Die Objektposition umfasst dabei die gesamte Objektfläche. Der Abstand berechnet sich daher als die kürzeste Distanz zwischen der Selektionsposition und der Objektgrenze; jede Selektionsposition, die auf einem Bildpixel liegt, das zum Objekt gehört, gilt als perfekter Treffer mit Abstand Null (vgl. das Beispiel in Abb. 6.2).

- Selektionszeit: Sie berechnet sich als Zeitdauer zwischen Erscheinen des grünen Selektionstriggers und dem Selektionszeitpunkt.
- Zufriedenstellung: Die subjektive Bewertung erfolgt mithilfe des Fragebogens zur Einzelbewertung, vgl. [DIN14] auf einer 7-Punkte-Skala (7: beste Bewertung; 1: schlechteste Bewertung).

Um Forschungsfrage 5 nach der Verbesserungsmöglichkeit der Selektionsgenauigkeit mithilfe eines Blickfilteralgorithmus zu beantworten, wurden die aufgezeichneten Blickdaten der Versuchspersonen mit dem echtzeitfähigen *RT-SDFS* von Kumar u. a. [Kum08] (vgl. Abschnitt 2.4.1) nachträglich verarbeitet. Der Schwellenwert des Verfahrens, der Sakkadenblickdaten von Fixationsblickdaten trennt, wurde auf $S = 20\ Pixel$ gesetzt. Für diese geglätteten Blickdaten wurde dann ebenfalls die Selektionsgenauigkeit berechnet.

Tabelle 6.1 zeigt die Ergebnisse der *Selektionsfehlerquote*. Betrachtet man alle Versuchspersonen und alle Objekte (Zeile 1), erzielt Blick+Taste mit 12,0 ± 5,5 % eine bessere Selektionsfehlerquote als die Mauseingabe mit 14,3 ± 5,3 % (signifikanter Unterschied bei Zweistichproben t-Test bei abhängigen Stichproben, α = 0,05, mit $t(25)$=2,34; p<0,05). Bei einem Gesamtumfang von 1690 Einzelselektionen (65 Objekte ∗ 26 Versuchspersonen) entspricht dies 8 ± 3,6 verpassten Objekten mit BT, und 9 ± 3,4 verpassten Objekten mit M.

Getrennte Betrachtung von Brillenträgern und Nicht-Brillenträgern (Zeile 2 bzw. 3) zeigt, dass die Nicht-Brillenträger (Zeile 2) mit Blick+Taste eine bessere Leistung erzielen ($t(14)$=3,53; p<0,01); bei den Brillenträgern (Zeile 3) ist die Leistung mit beiden Interaktionstechniken gleich ($t(10)$=0,08; p=0,932). Außerdem ist die Leistung der Nichtbrillenträger für beide Interaktionstechniken besser als die der Brillenträger.

Die Selektionsfehlerquote für die 6 Objekte, die nur sehr kurz nach dem Triggerzeitpunkt (< 3 Sekunden) sichtbar (und damit selektierbar) sind, ist im Mittel etwas geringer mit Blick+Taste (Zeile 4); der Unterschied ist jedoch nicht signifikant ($t(25)$=2,03; p=0,052). Getrennte Betrachtung von Brillenträgern und Nicht-Brillenträgern zeigt, dass die Selektionsfehlerquote mit der Mauseingabe für beide Gruppen ähnlich ist. Die Nicht-Brillenträger erzielen jedoch

Tabelle 6.1: Selektionsfehlerquote (verpasste Objekte) als Mittelwerte (± 1 Standardabweichung) in Prozent, signifikant bessere Ergebnisse zwischen Mauseingabe und Blick+Taste sind **fett** (Zweistichproben *t*-Test bei abhängigen Stichproben, $\alpha = 0{,}05$).

Objekte	VP	M	BT
Alle (m=65)	Alle	14,3 (5,3)	**12,0 (5,5)**
	Ohne Br.	12,8 (4,1)	**8,9 (3,9)**
	Brille	16,2 (6,2)	16,1 (4,9)
Kurz Sichtbare (< 3 Sekunden, m=6)	Alle	55,8 (21,1)	46,2 (27,2)
	Ohne Br.	54,4 (20,0)	**38,9 (25,1)**
	Brille	57,6 (22,8)	56,1 (26,1)

mit Blick+Taste ein besseres Ergebnis ($t(14)=2{,}29$; $p<0{,}05$). Das Ergebnis der Brillenträger ist mit beiden Interaktionstechniken ähnlich.

Tabelle 6.2 zeigt die Ergebnisse für die *Selektionsgenauigkeit*. Die ersten beiden Ergebnis-Spalten zeigen die Ergebnisse der von den Versuchspersonen erzielten Leistung. Die dritte Ergebnis-Spalte (ganz rechts) zeigt die Ergebnisse bei nachträglicher Verarbeitung der Blickrohdaten mit dem Blickfilterverfahren *RT-SDFS*.

Die Ergebnisunterschiede zwischen den Techniken sind gering. Die besten Ergebnisse erzielte die Mauseingabe, gefolgt von Blick+Taste mit *RT-SDFS* und Blick+Taste. Nicht-Brillenträger erzielten bessere Ergebnisse als Brillenträger. Die Ergebnisse für Objekte > 1° sind besser als für Objekte ≤ 1°.

Die Selektionsgenauigkeit der Mauseingabe beträgt im Mittel bei Betrachtung aller Objekte und aller Versuchspersonen 0,55° Sehwinkel (Zeile 1), was 6,2 mm auf dem Monitor entspricht. Nicht-Brillenträger (Zeile 2) erzielen im Mittel 0,51° (5,8 mm), Brillenträger (Zeile 3) 0,6° (6,8 mm). Das beste Ergebnis erzielt die Gruppe der Nicht-Brillenträger für die Objekte ≤ 1° (Zeile 5) mit im Mittel 0,47° (5,3 mm).

Tabelle 6.2: Selektionsgenauigkeit in ° Sehwinkel als Mittelwerte (± 1 Standardabweichung) für Mauseingabe (M), Blick+Taste (BT) und für Blick+Taste mit nachträglicher Blickdatenfilterung (BT mit *RT-SDFS*).

Objekte	VP	M	BT	BT mit *RT-SDFS*
Alle	Alle	0,55 (0,84)	0,83 (0,83)	0,74 (0,78)
	Ohne Br.	0,51 (0,84)	0,77 (0,82)	0,69 (0,77)
	Brille	0,60 (0,86)	0,93 (0,84)	0,81 (0,69)
≤1°, m=27	Alle	0,53 (0,67)	0,89 (0,78)	0,78 (0,73)
	Ohne Br.	0,47 (0,58)	0,84 (0,80)	0,73 (0,73)
	Brille	0,60 (0,77)	0,96 (0,74)	0,86 (0,69)
> 1°, m=38	Alle	0,58 (0,97)	0,80 (0,82)	0,70 (0,70)
	Ohne Br.	0,57 (1,03)	0,73 (0,79)	0,66 (0,77)
	Brille	0,59 (0,88)	0,90 (0,86)	0,77 (0,69)

Die Selektionsgenauigkeit mit Blick+Taste beträgt im Mittel bei Betrachtung aller Objekte und aller Versuchspersonen 0,83° Sehwinkel (entspricht 9,4 mm auf dem Monitor). Auch hier erzielen die Nicht-Brillenträger ein besseres Ergebnis als die Brillenträger, mit 0,77° (8,7 mm) versus 0,93° (10,6 mm). Das beste Ergebnis erzielt die Gruppe der Nicht-Brillenträger für die Objekte > 1° (Zeile 8) mit im Mittel 0,73° (8,3 mm), das schlechteste die Gruppe der Brillenträger für die Objekte ≤ 1° (Zeile 6) mit im Mittel 0,96° (10,9 mm).

Im Vergleich zur ungefilterten Blick+Taste verbessert die Nachverarbeitung der aufgezeichneten Blickrohdaten mit dem *RT-SDFS*-Algorithmus die Selektionsgenauigkeit für alle Analyseergebnisse. Für alle Objekte und Versuchspersonen (Zeile 1) werden 0,74° (8,4 mm) erzielt, was einer Verbesserung um 1 mm entspricht. Für die Nicht-Brillenträger verbessert sich die Selektionsgenauigkeit um 0,9 mm, für die Brillenträger um 1,3 mm. Am nächsten an die Selektionsgenauigkeit der Mauseingabe kommt die Gruppe der Nicht-Brillenträger für Objekte > 1° mit 0,66° (7,5 mm), die Mauseingabe liegt hier

Tabelle 6.3: Selektionszeit in ms als Mittelwert (± 1 Standardabweichung) für Mauseingabe (M) und Blick+Taste (BT). Signifikant bessere Ergebnisse zwischen Mauseingabe und Blick+Taste sind **fett** (Zweistichproben *t*-Test bei abhängigen Stichproben, $\alpha = 0{,}05$).

Objekte	VP	M	BT
Alle	Alle	1315 (466)	**775 (319)**
	Ohne Br.	1305 (477)	**771 (334)**
	Brille	1329 (450)	**781 (296)**

bei 0,57° (6,5 mm). Für die Objekte ≤ 1° beträgt die Differenz zur Mauseingabe 0,25° (3 mm), für die Objekte > 1° beträgt sie 0,12° (1,4 mm).

Tabelle 6.3 zeigt die Ergebnisse für die *Selektionszeit*. Die Versuchspersonen waren mit Blick+Taste substanziell schneller als mit Mauseingabe (Alle: $t(25)=10{,}66$; $p<0{,}001$; Ohne Brille: $t(14)=7{,}22$; $p<0{,}001$; Mit Brille: $t(10)=8{,}89$; $p<0{,}001$). Vergleicht man die Mittelwerte, ist die Selektionszeit mit Blick+Taste 41% kürzer. Bei der Selektionszeit zeigte sich kein Unterschied zwischen Versuchspersonen mit und ohne Brille.

Abb. 6.6 zeigt die Ergebnisse für die Zufriedenstellung. Blick+Taste wird für alle Merkmale gleich gut oder besser bewertet als die Mauseingabe.

Zweistichproben *t*-Tests bei abhängigen Stichproben, $\alpha = 0{,}05$, zeigten signifikant bessere Ergebnisse für Gleichmäßigkeit der Nutzung ($t(25)=-2{,}9$; $p<0{,}01$), Benutzungsgeschwindigkeit ($t(25)=-4{,}54$; $p<0{,}001$), Allgemeine Zufriedenheit ($t(25)=-3{,}88$; $p<0{,}001$), Nutzung insgesamt ($t(25)=-3{,}05$; $p<0{,}01$), Ermüdung der Finger ($t(25)=-3{,}24$; $p<0{,}01$), Ermüdung des Handgelenks ($t(25)=-4{,}09$; $p<0{,}001$), Ermüdung des Arms ($t(25)=-3{,}60$; $p<0{,}01$) und Ermüdung der Schulter ($t(25)=-3{,}09$; $p<0{,}01$).

Gleich gut sind die Bewertungen für Anstrengung ($t(25)=-1{,}13$; $p=0{,}267$), Genauigkeit ($t(25)=0{,}089$; $p=0{,}929$), Ermüdung des Nackens ($t(25)=-0{,}54$; $p=0{,}294$) und Ermüdung der Augen ($t(25)=0{,}75$; $p=0{,}456$).

Gefragt nach ihrer bevorzugten Interaktionstechnik votierten 3 Versuchspersonen für die Mauseingabe, 18 für Blick+Taste, 5 hatten keine Präferenz.

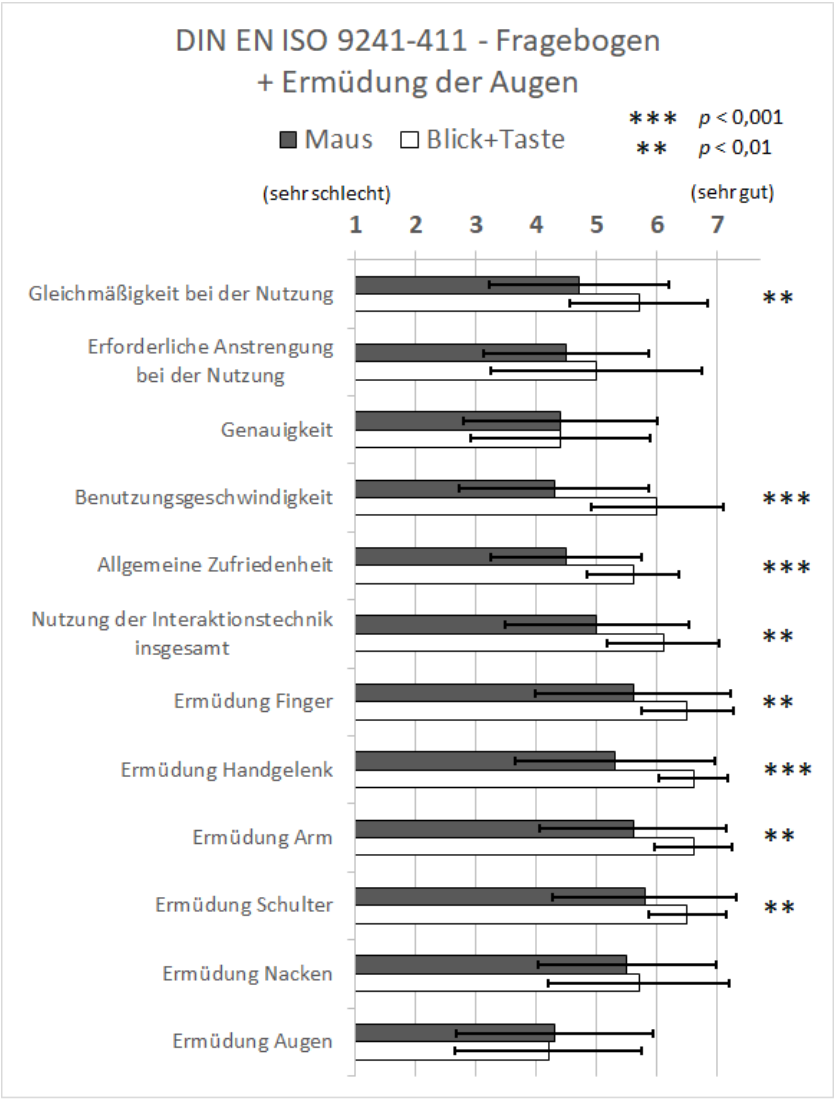


Abbildung 6.6: Subjektive Bewertung der Zufriedenstellung.

Bezüglich der **Forschungsfragen** ergeben sich folgende Antworten.

Frage 1: Die Versuchspersonen konnten mit Blick+Taste erheblich schneller (41%) markieren als mit der Mauseingabe.

Auch die Selektionsfehlerquote war mit Blick+Taste geringfügig besser. Die Selektionsgenauigkeit war hingegen geringfügig schlechter (ca. 3 mm), was sich aber in der gestellten Testaufgabe nicht in höherer Selektionsfehlerquote niederschlug. Mutmaßlich lag dies daran, dass eine Markierung als erfolgreich gezählt wurde, wenn der Rahmen dem Objekt eindeutig zuzuordnen war und mindestens ein Drittel der Objektfläche innerhalb des Rahmens lag. Die Rahmengröße von 2° war demzufolge eine angemessene Wahl für die gestellte Aufgabe.

Frage 2: Nicht-Brillenträger erzielten gegenüber Brillenträgern eine geringere Selektionsfehlerquote mit beiden Interaktionstechniken. Der Unterschied war deutlicher bei Blick+Taste und hier insbesondere bei Objekten, die nur sehr kurz in der Szene sichtbar waren. Die Selektionsgenauigkeit unterschied sich kaum. Es ist bekannt, dass Eyetracking bei Brillenträgern oft etwas schlechter funktioniert als bei Nicht-Brillenträgern (vgl. Abschnitt 2.2, S. 43). Zwar resultierte dies hier in kaum schlechterer Selektionsgenauigkeit, fiel aber bei kurz sichtbaren Objekten ins Gewicht. Die Selektionszeit war gleich für beide Gruppen.

Frage 3: Bei sehr kurz sichtbaren Objekten (Sichtbarkeit < 3 Sekunden) sind die Fehlerquoten sehr hoch. Blick+Taste schneidet hier etwas besser ab als die Mauseingabe, wobei dies vor allem für die Gruppe der Nicht-Brillenträger galt.

Frage 4: Die Selektionsgenauigkeit ist geringfügig besser für Objekte > 1° als für Objekte ≤ 1°.

Frage 5: Durch Blickdatenfilterung mit dem Algorithmus von Kumar u. a. [Kum08] konnte die Selektionsgenauigkeit für Blick+Taste um 1 mm (auf 2 mm) verbessert werden.

Frage 6: Die subjektive Bewertung der Interaktionstechniken ist für Blick+Taste für die überwiegende Anzahl der Merkmale besser und für die übrigen gleich wie die der Mauseingabe.

Frage 7: Das Merkmal Ermüdung der Augen wurde von den Videoauswerteexperten für beide Interaktionstechniken gleich gut bewertet.

6.1.2.4 Fazit

Die Ergebnisse dieser Untersuchung mit Videoanalyseexperten bestätigen die der Pilotstudie mit Nicht-Experten aus Abschnitt 6.1.1 bezüglich der Selektionsfehlerquote. Diese ist in beiden Fällen für Blick+Taste und Mauseingabe ähnlich gut und liegt zudem in beiden Studien bei ähnlichen Werten (Blick+Taste mit 12,0 ± 5,5 % bzw. 14,2 ± 7,6 %, Mauseingabe mit 14,3 ± 5,3 % bzw. 13,6 ± 7,8 %).

Die Selektionsgenauigkeit ist hingegen in der Videoexpertenstudie für die Mauseingabe etwas besser. Wie in der Pilotstudie sind die Unterschiede jedoch zu gering, um die Selektionsfehlerquote bei den gegebenen Zielobjektgrößen und gegebener Größe des Markierungsrahmens von 2° negativ zu beeinflussen.

Die Selektionszeit war signifikant besser mit Blick+Taste. Mit 41% kürzerer Selektionszeit fällt der Unterschied in der Videoexpertenstudie noch wesentlich deutlicher aus als bei den Untersuchungen an abstrakten Testaufgaben aus Abschnitt 5.1.

Auch die hohe Zufriedenstellung der Versuchspersonen mit Blick+Taste wurde bestätigt. Im Vergleich mit den Nicht-Experten aus Abschnitt 6.1.1 (Abb. 6.3), war die Bewertung an drei Stellen jedoch anders. Erstens empfanden die Videoexperten weniger Anstrengung mit beiden Interaktionstechniken. Zweitens bewerteten sie die Genauigkeit für beide Interaktionstechniken gleich, die Nicht-Experten hatten die Genauigkeit für Blick+Taste besser bewertet.

Drittens bewerteten die Videoexperten die Ermüdung der Augen praktisch identisch mit 4,3 ± 1,6 für Blick+Taste und mit 4,1 ± 1,5 für Mauseingabe. Die

Nicht-Experten bewerteten Blick+Taste mit 3,7 ± 1,9 als deutlich ermüdender als die Mauseingabe mit 5,1 ± 1,6 (Abb. 6.3). Möglicherweise ist die geringere Ermüdung der Videoexperten auf ihre Vertrautheit mit der Aufgabenstellung zurückzuführen. Auch die Versuchspersonen der Längsschnittstudie hatten nach ihrer 6-monatigen Trainingszeit, in der sie mit der Aufgabe der Bewegtobjektselektion vertraut geworden waren, mit 4,8 ± 0,5 eine ähnliche Ermüdung angegeben wie die Videoexperten. Diese Bewertung hatte sich signifikant verbessert im Vergleich zur Situation vor der Trainingsperiode, wo sie bei 3,0 ± 0,0 lag.

6.2 Blickbasierte Interaktion bei automatischen Bildanalyseverfahren

Dieser Abschnitt umfasst die Ergebnisse der Nutzerstudien zu Beitrag 2, der blickbasierte Interaktion bei automatischen Bildanalyseverfahren zur Aktivitätsdetektion betrachtet (vgl. Abschnitt 1.3.2 bzw. Abschnitt 3.2 für die Konzeption).

Abschnitt 6.2.1 bringt eine Untersuchung zur Bewegtobjektmarkierung bei Assistenz eines automatischen Verfahrens zur *Bewegungsdetektion*. Abschnitt 6.2.2 und Abschnitt 6.2.3 bringen eine Pilotstudie bzw. eine Expertenstudie zur Initialisierung eines automatischen Verfahrens zum *Einzelobjekttracking*.

Die Evaluationen aus Kapitel 5 und Abschnitt 6.1 zeigten, dass Blick+Taste mit Filterung der Blickrohdaten durch den *RT-SDFS* [Kum08] unter den evaluierten blickbasierten Interaktionstechniken am besten abschnitt. Daher wurde diese Interaktionstechnik auch für die blickbasierte Interaktion bei automatischen Bildanalyseverfahren genutzt.

6.2.1 Bewegtobjektmarkierung bei Assistenz durch automatische Bewegungsdetektion

Dieses Experiment und seine Ergebnisse wurden bei der SPIE 2017 veröffentlicht [Hil17].

Die Aufgabenstellung bei dieser Untersuchung ist ähnlich wie bei der Bewegtobjektmarkierung im vorigen Abschnitt 6.1. Der Benutzer hat wie dort die Aufgabe, Objekte im Video zu markieren.

Verglichen werden wieder die Interaktionstechniken Blick+Taste und Mauseingabe. Zusätzlich wird der Einfluss der Verfügbarkeit eines automatischen Verfahrens zur Bewegungsdetektion untersucht. Insgesamt werden also drei Interaktionstechniken betrachtet: Blick+Taste, Mauseingabe sowie Blick+Taste mit automatischer Bewegungsdetektion.

Die Selektion mit der *Blick+Taste*-Interaktion erfolgt, indem der Benutzer das Objekt anblickt und währenddessen die ENTER-Taste des NumPad drückt. Bei *Blick+Taste mit automatischer Bewegungsdetektion* ist zusätzlich das Bewegungsdetektionsverfahren des Versuchssystems aktiviert.

Die *Mauseingabe* erfolgt als traditionelle Zeige-Klick-Interaktion mit Einzel-Klick der linken Maustaste.

Vor der Vorstellung der Testaufgaben wird im folgenden Abschnitt zunächst das automatische Bewegungsdetektionsverfahren kurz beschrieben.

6.2.1.1 Automatisches Verfahren zur Bewegungsdetektion

Automatische Bewegungsdetektion detektiert Bildregionen, die sich unabhängig vom Szenenhintergrund bewegen. Der Algorithmus, der in ABUL integriert ist, wurde von Teutsch u. a. [Teu12] vorgeschlagen und wird im Folgenden mit IMD (für engl. Independent Motion Detection) abgekürzt. IMD basiert auf der Detektion und dem Tracking von markanten Punkt-Merkmalen von Bild zu Bild [Shi94].

Abbildung 6.7: Szene mit IMD-Verfahrensergebnissen als rote Hervorhebungen durch das Versuchssystem.

Im ersten Verarbeitungsschritt werden die Punktmerkmale zur Bild-zu-Bild-Registrierung genutzt. Hierbei wird eine globale Bild-Transformation (Homografie) zwischen aufeinanderfolgenden Bildern geschätzt, um die Einzelbilder zur Deckung zu bringen [Har04]. Die Bild-zu-Bild-Registrierung sorgt dafür, dass die dominierende Bildbewegung, die von der Bewegung der Sensorplattform herrührt, eliminiert wird.

Im zweiten Verarbeitungsschritt werden an den getrackten, markanten Punkten relative Bildgeschwindigkeiten bestimmt. Die Bewegungsdetektion erfolgt an Punkten mit signifikanten relativen Geschwindigkeiten und separiert Punkte bewegter Objekte von Punkten auf dem (bewegungskompensiert statischen) Hintergrund.

Im dritten Verarbeitungsschritt wird für die unabhängig bewegten Punkte eine halbtransparente Visualisierung bestimmt, die dem Benutzer angezeigt wird. Abb. 6.7 zeigt ein beispielhaftes Einzelbild aus dem Pool der Testaufgaben mit visualisierten Verfahrensergebnissen.

6.2.1.2 Testaufgaben

Für diese Untersuchung wurden vier Testaufgabentypen A bis D gestaltet. Sie unterscheiden sich in Kategorie und Anzahl der Zielobjekte, die zu markieren sind, sowie in der Dauer der Sichtbarkeit der Zielobjekte in der Szene. Tabelle 6.4 beschreibt ihre Charakteristika. Abb. 6.8 zeigt ein Beispiel für die größte Objektgröße von 2,5 cm (2,26°), Abb. 6.9 ein Beispiel für die kleinste Objektgröße von 0,1 cm (0,09°).

Als Datenmaterial für die Testaufgaben wurde eine längere Überflugvideosequenz genutzt, aus der kürzere Videoclips ausgeschnitten wurden. Die Geschwindigkeiten, mit denen sich die Objekte bewegen, liegen im Bereich zwischen 60 Pixel/s (2,06°/s bzw. 2,28 cm/s) und 380 Pixel/s (13,05°/s bzw. 14,44 cm/s).

Eine einzelne Aufgabe läuft so ab, dass die Versuchspersonen zunächst die Aufgabenkategorie (A, B, C oder D) genannt bekommen, aus der die Zielobjekte stammen. Dann startet der erste Videoclip der Testaufgabengruppe und die Versuchspersonen müssen alle Zielobjekte mit einem Rahmen markieren (vgl. Aufgabenstellung in Abschnitt 6.1). Je zwei Videoclips werden voneinander getrennt durch die Anzeige eines einheitlich grauen Bildes als Szene für 4 Sekunden. Die Rahmengröße für die Markierung beträgt 60 Bildpixel (2,06° bzw. 22,8 mm auf dem Monitor).

Tabelle 6.4: Die vier Testaufgabentypen A, B, C, D und ihre Charakteristika.

Testaufgaben-Typ	Aufgabenumfang (Sichtbarkeitsdauer)	Zielobjekte Kategorie	Anzahl	Größe
A	12 Clips (21-70s) insges. 9 min	Motorisierte Landfahrzeuge	67	0,5-2,5 cm
B	5 Clips (5-19s) insges. 1:30 min	Motorisierte Landfahrzeuge (Sichtbarkeit < 3s, Geschwindigkeit bis 380 Pixel/s)	7	1,0-2,5 cm
C	6 Clips (24-53s) insges. 4 min	Einzelne Arten motorisierter Landfahrzeuge	27	0,5-2,5 cm
D	11 Clips (25-68s) insges. 7 min	Alle Arten menschlicher Bewegung	99	0,1-2,0 cm
Gesamt:	21:30 min		200	

Abbildung 6.8: Beispielhafte Markierung eines Objektes der größten Objektgröße von 2,5 cm Länge; Situation ohne Assistenz durch automatische Bewegungsdetektion.

Abbildung 6.9: Beispielhafte Markierung eines Objektes der kleinsten Objektgröße von 0,1 cm Länge; Situation mit Assistenz durch automatische Bewegungsdetektion.

6.2.1.3 Versuchsaufbau, Versuchspersonen und Versuchsablauf

Das Versuchssystem ist ein Derivat des ABUL-Systems. Angezeigt werden die Trainings- und Testaufgaben auf einem 24-Zoll-Monitor (Auflösung 1920 x 1200 Pixel). Die Größe des Markierungsrahmens beträgt 60 Pixel (2,06° bzw. 22,8 mm auf dem Monitor). Der Versuchsaufbau ist wie in Abb. 4.3.

Zur Blickerfassung wird der Tobii X60 verwendet (vgl. Abschnitt 4.2). Die Blickrohdaten werden gefiltert mit dem *RT-SDFS* [Kum08] (s. a. Abschnitt 2.4.1). Der Schwellenwert zur Trennung von Fixationen und Sakkaden wurde wie empfohlen auf $S = 20\ Pixel$ gesetzt. Da dieser Algorithmus eine zusätzliche Latenz von 1 Blickdatensample bewirkt, beträgt die Latenz, mit der das Blicksignal geliefert wird, jetzt 33 ms (X60) + 17 ms (Blickfilterung) = 50 ms.

Der manuelle Tastendruck bei Blick+Taste erfolgte mit der ENTER-Taste des NumPad einer gewöhnlichen Computertastatur. Die Mauseingabe erfolgte mit einer Comfort-Mouse 2000 for Business 3 (1000dpi) von Microsoft.

Die 10 Versuchspersonen (9 männlich, 1 weiblich, Durchschnittsalter 29 Jahre) verfügten alle über normale oder auf normal korrigierte Sicht (1 mit Brille, 1 mit Kontaktlinsen). Alle waren Studenten oder Kollegen und demzufolge ohne Erfahrung in Videobildauswertung. Alle waren langjährig erfahrene Benutzer von Desktop-Computern und Computermaus. Drei hatten einmal zuvor Erfahrung mit Eyetrackern oder Blickinteraktion gesammelt.

Der Versuchsablauf war wie folgt. Da die Testaufgaben sehr umfangreich waren, wurden nicht alle Kombinationen aus Interaktionstechnik (BT, M) und Verfügbarkeit der automatischen Verfahrensergebnisse (Ja, Nein) betrachtet, sondern nur drei davon: BT mit Verfahrensergebnissen, BT und M. Auf diese Weise konnte einerseits der Unterschied zwischen Blick+Taste und Mauseingabe wie in den Experimenten zur Bewegtobjektmarkierung (Abschnitt 6.1) festgestellt werden und andererseits auch der Unterschied zwischen Blick+Taste mit bzw. ohne Verfahrensergebnisse. Zudem absolvierten die Versuchspersonen die drei Interaktionstechniken in getrennten Sitzungen an drei unterschiedlichen Tagen. Je fünf Versuchspersonen starteten mit Mauseingabe gefolgt von Blick+Taste bzw. mit Blick+Taste gefolgt von Mauseingabe; Blick+Taste mit automatischer Bewegungsdetektion war stets die Interaktionstechnik in der dritten Sitzung. Die Reihenfolge der Testaufgaben war stets A - B - C - D.

Jede Sitzung dauerte etwa 45 Minuten. Nach einer kurzen Einführung mit Erläuterung der Aufgabe „Bewegtobjektmarkierung“ absolvierten die

Versuchspersonen bei den blickbasierten Bedingungen eine Standard-9-Punkte-Eyetrackerkalibierung. Darauf folgte ein kurzes Training anhand von Trainingsaufgaben, die aus demselben Videodatenmaterial ausgeschnitten wurden wie die Testaufgaben. Dann folgten die Testaufgaben, wobei nach jeder Testaufgabe unmittelbar der NASA-TLX-Fragebogen ausgefüllt wurde. Abschließend erfolgte die Bewertung der Zufriedenstellung mit dem Fragebogen zur Einzelbewertung aus der DIN EN ISO 9241-411 [DIN14], der um das Merkmal Ermüdung der Augen ergänzt wurde (vgl. [Zha07] und Abschnitte 5.1.2, 5.1.4, 6.1.1 und 6.1.2 bzw. [Hil13b, Hil14a, Hil16a]).

6.2.1.4 Forschungsfragen, Metriken und Ergebnisse

Dieses Experiment adressierte folgende Forschungsfragen:

- Frage 1: Welche Selektionstrefferquote erzielen die Versuchspersonen mit Blick+Taste im Vergleich zur Mauseingabe (beide ohne automatisches Bewegungsdetektionsverfahren)?
- Frage 2: Welche Selektionstrefferquote erzielen die Versuchspersonen mit Blick+Taste mit bzw. ohne Verfügbarkeit automatischer Verfahrensergebnisse?
- Frage 3: Wie unterscheidet sich die subjektiv empfundene Belastung zwischen den drei Interaktionstechniken?
- Frage 4: Wie ist die subjektive Bewertung der Zufriedenstellung für Blick+Taste? Wie ist insbesondere die subjektive Bewertung des Merkmals Ermüdung der Augen?

Hierfür wurden folgende Metriken verwendet:

- Selektionstrefferquote: Sie berechnet sich als Anzahl korrekter Selektionen / Anzahl Zielobjekte der Testaufgabe <A-D>. Ein Objekt gilt als erfolgreich selektiert, wenn der Selektionsrahmen eindeutig diesem Objekt zuzuordnen ist und der Rahmen das Objekt beinhaltet oder zumindest berührt. Im letzteren Fall vergrößert sich der selektionssensitive Bereich um die Objektgröße; für das kleinste

Objekt von 0,09° Seitenlänge beträgt er dann 3,04°, für das größte von 2,26° x 0,52° 5,23°.

- NASA-Task Load Index: Die subjektive Bewertung der Belastung erfolgte mithilfe des Fragebogens *NASA-Task Load Index*, vgl. [Har88, Har06] bzw. S. 93.
- Zufriedenstellung: Die subjektive Bewertung erfolgte mithilfe des Fragebogens zur Einzelbewertung [DIN14] auf einer 7-Punkte-Skala (7: beste Bewertung; 1: schlechteste Bewertung).

Abb. 6.10 zeigt die Ergebnisse für die Selektionstrefferquoten. Da die Ergebnisse sich für die vier Testaufgabentypen A bis D zum Teil stark unterscheiden, sind sie separat aufgeführt. Es zeigte sich, dass für alle Testaufgabentypen Blick+Taste mit automatischer Bewegungsdetektion besser abschneidet als die beiden anderen Interaktionsbedingungen. Die Selektionstrefferquoten für Blick+Taste und Mauseingabe sind für alle Aufgabentypen ähnlich gut.

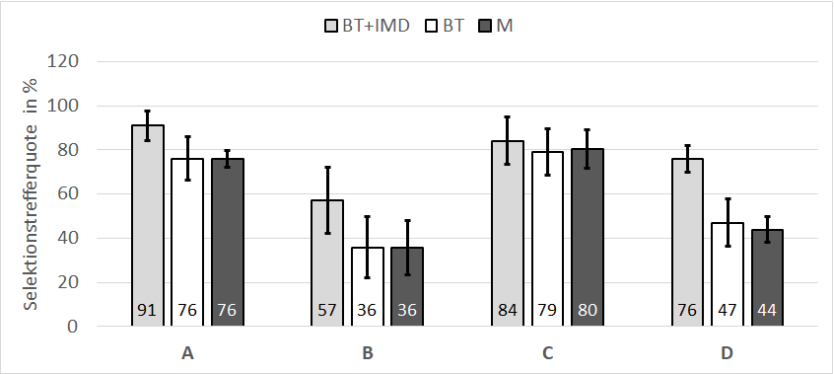


Abbildung 6.10: Ergebnisse der Selektionstrefferquote in Prozent als Mittelwerte ± 1 Standardabweichung, separat für die vier Testaufgaben A bis D.

Die besten Ergebnisse wurden für die Testaufgaben A und C erzielt. Dies ist auf die Größe der Zielobjekte zurückführbar (0,5 bis 2,5 cm auf dem Monitor)

sowie auf die Sichtbarkeitsdauer > 3 s und die moderaten Objektgeschwindigkeiten von ca. 60-100 Pixel/s (2,06-3,43°/s).

Testaufgabe B nutzt ebenfalls große Zielobjekte (1,0 bis 2,5 cm auf dem Monitor), die aber alle < 3 s sichtbar sind und zudem die höchsten Geschwindigkeiten von bis zu 380 Pixel/s (13,05°/s) aufweisen.

Die Schwierigkeit bei Testaufgabe D sind die teilweise sehr kleinen Objektgrößen: Motorradfahrer und Radfahrer haben eine Länge von 2 mm, Fußgänger umfassen nur 1 mm². Fußgänger umfassen auf dem Monitor also oft nur wenige Pixel und bewegen sich zudem langsam im Vergleich zur Hintergrundbewegung, weshalb sie schwierig zu detektieren sind.

Die Verbesserung bei Verfügbarkeit automatischer Bewegungsdetektion ist deutlich und beträgt im Mittel 16% für Testaufgabe A, 37% für Testaufgabe B und 38% für Testaufgabe D[1]. Die nur geringfügige Verbesserung für Testaufgabe C (im Mittel 6%[2]) mag zumindest teilweise durch die Aufgabenstellung bzw. nachträgliche Erklärung durch die Versuchspersonen erklärbar sein. Im Gegensatz zu den anderen Testaufgaben waren hier nicht „alle motorisierten Fahrzeuge" zu markieren, sondern bestimmte Typen wie „LKWs", „Busse" oder „PKWs". Die Versuchspersonen gaben an, dass sie zwar ein motorisiertes Fahrzeug detektiert hatten, es aber nicht als zur gefragten Klasse zugehörig erkannt und demzufolge nicht selektiert hatten. Videoauswerteexperten hätten aufgrund ihrer Erfahrung mit Bildfolgen hier vermutlich ein besseres Ergebnis erzielt.

Abb. 6.11 zeigt die Ergebnisse der Bewertung der subjektiv empfundenen Belastung mit dem Fragebogen NASA-Task Load Index, separat für die vier Testaufgabentypen. Eine Bewertung mit dem Wert 100 repräsentiert sehr hohe Belastung, eine Bewertung mit dem Wert 0 repräsentiert keine Belastung. Die Ergebnisse zeigen deutlich geringer empfundene Belastung mit den beiden Blick+Taste-Interaktionsbedingungen im Vergleich zur Mauseingabe.

[1] Zweiseitiger *t*-Test bei abhängigen Stichproben, $\alpha = 0{,}05$, zwischen BT und BT+IMD mit signifikantem Unterschied bei Testaufgabe A $t(9)=-5{,}79$; $p<0.{,}01$; bei Testaufgabe B $t(9)=-3{,}30$; $p<0{,}01$; bei Testaufgabe D $t(9)=-14{,}49$; $p<0{,}001$.

[2] *t*-Test bei Testaufgabe C nicht signifikant mit $t(9)=-1{,}87$; $p=0{,}094$.

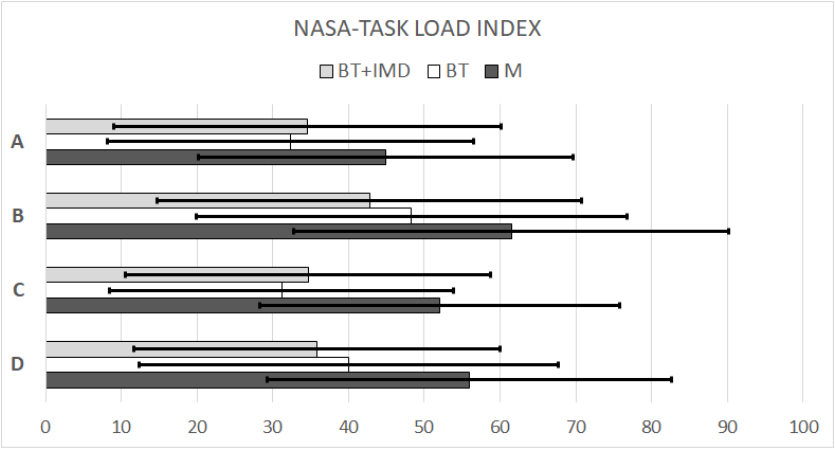


Abbildung 6.11: Ergebnisse der subjektiv empfundenen Belastung als Mittelwerte ± 1 Standardabweichung, separat für die vier Testaufgaben A bis D.

Die Bewertungen für die beiden Blick+Taste-Bedingungen sind etwa gleich gut. Es fällt jedoch auf, dass für die schwierigeren Testaufgaben B (kurze Sichtbarkeit) und D (teilweise sehr kleine Zielobjekte) – die schon geringere Selektionstrefferquoten erzielt hatten) – die Bedingung mit Assistenz durch automatische Bewegungsdetektion etwas besser bewertet wird. Es leuchtet ein, dass in diesen Situationen die Assistenz die empfundene Belastung reduziert. Für die einfacheren Testaufgaben A und C, wo die Zielobjekte einfacher zu detektieren sind, mag die permanente Hervorhebung der automatisch detektieren Bewegung als zusätzliche visuelle Belastung empfunden worden sein. Dennoch hat auch bei den Testaufgaben A und C die automatische Assistenz die Effektivität deutlich verbessert.

Abb. 6.12 zeigt die Ergebnisse der Bewertung der Zufriedenstellung. Die Mauseingabe erhält für alle Merkmale die schlechteste Bewertung[1]. Während bei den bisherigen Untersuchungen mit Nicht-Experten stets die Mauseingabe etwas besser bezüglich der Ermüdung der Augen bewertet wurde (vgl. Abb. 6.3 bzw. Tabelle 5.7), wird sie jetzt etwas schlechter bewertet[2].

Die Bewertungen der beiden blickbasierten Interaktionsbedingungen sind vergleichbar gut. Blick+Taste erzielt etwas bessere Ergebnisse für Allgemeine Zufriedenheit ($t(9)=2{,}68$; $p<0{,}05$) und Nutzung der Interaktionstechnik insgesamt ($t(9)=2{,}75$; $p<0{,}05$).

Bezüglich der **Forschungsfragen** ergeben sich folgende Antworten.

Frage 1: Blick+Taste und Mauseingabe erzielen vergleichbar gute Selektionstrefferquoten.

Frage 2: Die Verfügbarkeit einer Assistenz durch automatische Bewegungsdetektion erhöht die Selektionstrefferquote erheblich.

Frage 3: Die subjektiv empfundene Belastung ist für die Mauseingabe erheblich höher als für die blickbasierten Interaktionsbedingungen.

Frage 4: Die subjektive Bewertung der Zufriedenstellung für Blick+Taste und für Blick+Taste mit automatischer Bewegungsdetektion ist sehr gut und übertrifft die Mauseingabe für fast alle Merkmale deutlich, unter anderem wird die Ermüdung der Augen als geringer empfunden.

[1] Zweiseitiger t-Test bei abhängigen Stichproben, $\alpha = 0{,}05$, zwischen BT und M mit signifikantem Unterschied mit $p<0{,}001$ für Geschwindigkeit, Allgemeine Zufriedenheit und Nutzung insgesamt; mit $p<0{,}01$ für Handgelenk und Arm; mit $p<0{,}05$ für Erforderliche Anstrengung und Schulter.

[2] Zweiseitiger t-Test bei abhängigen Stichproben, $\alpha = 0{,}05$, liefert für den Vergleich Mauseingabe mit Blick+Taste+IDM $t(9)=-2{,}37$; $p<0{,}05$, mit Blick+Taste $t(9)=-2{,}12$; $p=0{,}062$.

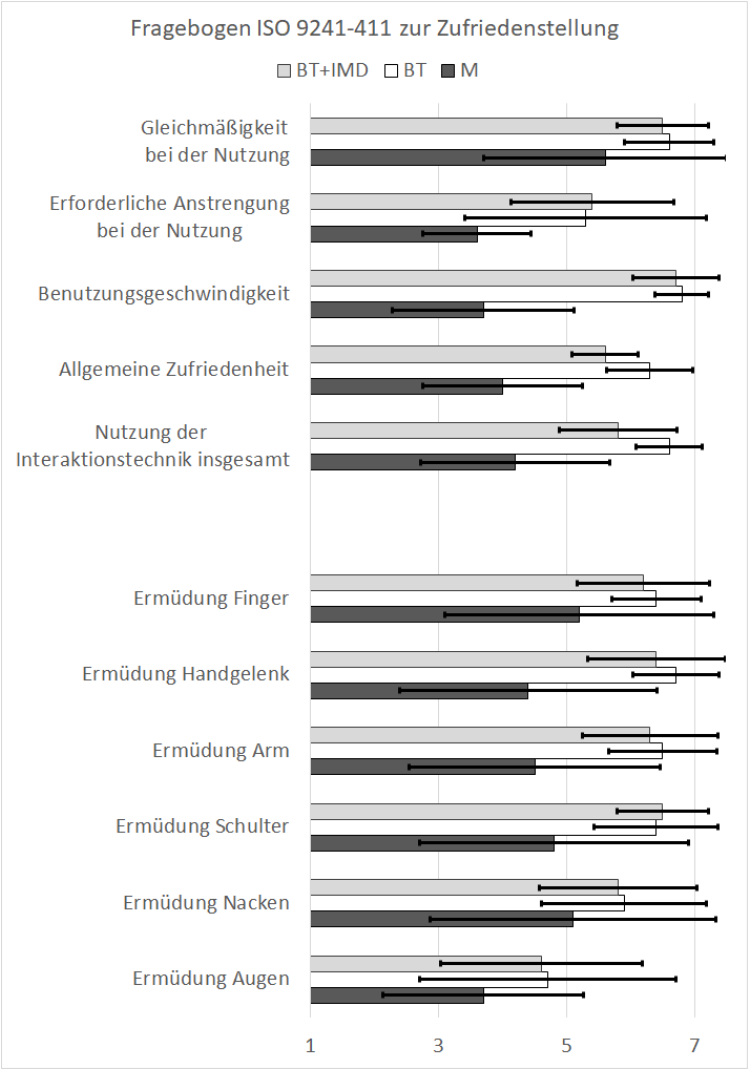


Abbildung 6.12: Ergebnisse der subjektiven Bewertung der Zufriedenstellung als Mittelwerte ± 1 Standardabweichung, separat für die vier Testaufgaben A bis D. 7: beste Bewertung; 1: schlechteste Bewertung.

6.2.1.5 Fazit

Die Ergebnisse dieser Untersuchung bestätigen die der Bewegtobjektmarkierung aus Abschnitt 6.1 (und auch die der Untersuchungen aus Abschnitt 5.1) bezüglich der vergleichbar guten Effektivität von Blick+Taste und Mauseingabe.

Die Assistenz durch ein automatisches Verfahren zur Bewegungsdetektion, das dem menschlichen Beobachter detektierte Bewegung als halbtransparente Hervorhebung visualisiert, verbessert die Effektivität erheblich. Das Ergebnis dieser Informationsfusion aus automatischer Objektdetektion und menschlicher Objektdetektion lieferte je nach Testaufgabentyp eine Erhöhung der Selektionstrefferquote um bis zu 38%.

Die Ergebnisse der subjektiv empfundenen Belastung mit dem Fragebogen NASA-TLX zeigten für die Mauseingabe eine deutlich höhere Belastung als für Blick+Taste und Blick+Taste+IMD, die ähnlich bewertet wurden.

Ein ähnliches Ergebnis zeigt auch die subjektive Bewertung der Zufriedenstellung. Die beiden blickbasierten Interaktionstechniken erzielen ähnlich gute Bewertungen und deutlich bessere als die Mauseingabe für fast alle Merkmale. Insbesondere die Ermüdung der Augen wird als deutlich höher für die Mauseingabe bewertet. Dies war bei den Untersuchungen zur Bewegtobjektmarkierung aus Abschnitt 6.1.1 und Abschnitt 6.1.2 nicht der Fall. Eine Erklärungsmöglichkeit ist, dass jetzt der Aufgabenumfang mit einer Gesamtdauer der Testaufgaben von 21:30 min und insgesamt 200 Zielobjekten am größten war. Dadurch schlug sich möglicherweise die geringere Interaktionskomplexität von Blick+Taste in geringerer subjektiv empfundener Belastung (NASA-TLX) sowie geringerer Ermüdung der Augen nieder.

6.2.2 Initialisierung eines automatischen Verfahrens zum Einzelobjekttracking: Pilotstudie

Dieses Experiment und seine Ergebnisse wurden bei der SPIE 2015 veröffentlicht [Hil15].

Auch bei dieser Untersuchung kommt wieder die Bewegtobjektselektion als Interaktionsaufgabe vor. Im Unterschied zu allen vorangegangenen Untersuchungen bewirkt die Bewegtobjektselektion jetzt keine direkte Markierung eines Zielobjekts, sondern die Initialisierung eines Einzelobjekttrackers für ein Zielobjekt. Der Markierungsrahmen für das Zielobjekt wird dann durch das automatische Einzelobjekttracking-Verfahren realisiert.

Verglichen werden drei Interaktionsbedingungen: *Mauseingabe*, *Blick+Taste* und Blick+Taste mit kontinuierlicher Visualisierung der Blickposition, im Folgenden mit *Blick+Taste+BlickVis* abgekürzt.

Die kontinuierliche Visualisierung gilt zwar in der Fachliteratur als problematisch (vgl. Abschnitt 2.4.2.3 bzw. unser Ergebnis für MAGIC pointing liberal aus Abschnitt 5.1.1.4). Eine Visualisierung wurde jedoch seitens der Videoanalyseexperten wiederholt als Möglichkeit angefragt, da die Mauseingabe die Selektionsposition über den Mauszeiger zeigt. Daher wurde Blickvisualisierung in dieser Pilotstudie für die etwas komplexere Aufgabe der Einzelobjekttracker-Initialisierung mitgetestet.

Auch in dieser Untersuchung erfolgt die *Mauseingabe* als traditionelle Zeige-Klick-Interaktion mit Einzel-Klick der linken Maustaste.

Die Selektion mit der *Blick+Taste*-Interaktion erfolgt, indem der Benutzer das Objekt anblickt und währenddessen die ENTER-Taste des NumPad drückt.

Die Selektion mit *Blick+Taste+BlickVis* erfolgt wie mit Blick+Taste, nur dass währenddessen die Blickposition permanent visualisiert wird. Abb. 6.13 zeigt die realisierten Varianten. Beide sind halbtransparent grün visualisiert.

Die Ring-Visualisierung interpretiert die Blickposition als Bereich analog zum fovealen Bereich scharfen Sehens, der etwa 2° beträgt (vgl. S. 26). Der Durchmesser wurde daher mit 60 Bildpixeln (entsprechend 2,06° bei 65 cm Augenabstand vom Monitor) gewählt; der Mittelpunkt repräsentiert die vom Eyetracker gelieferte Blickposition auf dem Monitor.

Die Punkt-Visualisierung interpretiert die Blickposition als Blickpunkt und visualisiert die vom Eyetracker gelieferte Blickposition auf dem Monitor; der Durchmesser beträgt 6 Bildpixel (0,2° bei 65 cm Augenabstand vom Monitor).

Um die Gesamtdauer der Datenerfassung nicht zu lange werden zu lassen, wurde nur eine der Blickvisualisierungen getestet, und zwar die Ring-Visualisierung. Diese Entscheidung fußt auf einem kurzen Test der beiden Visualisierungsvarianten mit 5 Versuchspersonen, die alle die Ringvisualisierung bevorzugten.

Vor der Vorstellung der Testaufgaben wird im folgenden Abschnitt zunächst das automatische Verfahren zum Einzelobjekttracking kurz beschrieben.

(a) Visualisierung der Blickposition als Ring.

(b) Visualisierung der Blickposition als Punkt.

Abbildung 6.13: Realisierte Systemoptionen für die Visualisierung der Blickposition.

6.2.2.1 Einzelobjekttracking-Verfahren

Das verwendete Einzelobjekttracking-Verfahren kombiniert Module für die Bild-zu-Bild-Registrierung, Bewegungsdetektion (Independent Motion Detection) und Objekttracking. Es wurde von Teutsch u. a. [Teu12] vorgestellt und ist als Subsystem in ABUL integriert. Bild-zu-Bild-Registrierung und Bewegungsdetektion basieren auf dem Tracking von Punkt-Merkmalen zwischen je zwei Videoframes [Teu12, Shi94]. Damit der Tracking-Algorithmus ein Objekt detektiert und trackt, muss der Benutzer den Anstoß dazu geben, indem er das Objekt selektiert, das getrackt werden soll. Durch die Selektion wird dem Tracking-Algorithmus die Bildposition übermittelt, an der sich das Objekt befindet.

Aus Benutzersicht befindet sich der Tracking-Algorithmus in einem von drei Zuständen (Abb. 6.14): WAIT, INIT oder TRACK. Im Zustand WAIT wartet der Tracking-Algorithmus auf die Eingabe der Bildposition durch die Selektionsoperation des Benutzers.

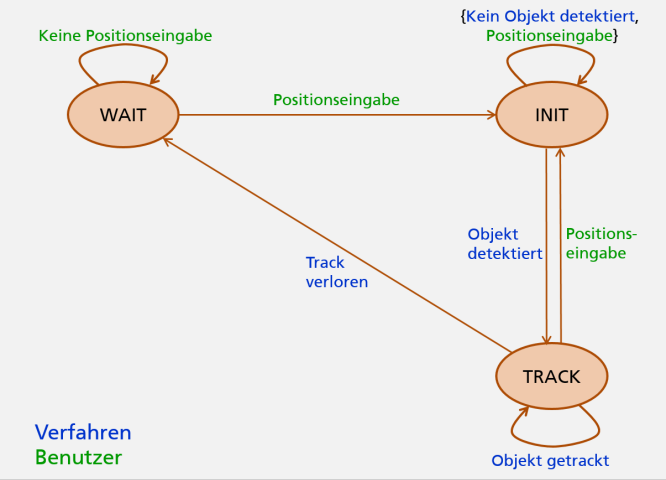


Abbildung 6.14: Zustände des Trackingverfahrens aus Benutzersicht.

Durch Positionseingabe wechselt der Tracking-Algorithmus in den Zustand INIT und startet das Bewegungsdetektionsmodul für einen quadratischen Bildbereich (Region of Interest *ROI*), dessen Zentrum die Positionseingabe bildet. Abb. 6.15 zeigt die visuelle Rückmeldung, die der Algorithmus dem Benutzer in diesem Zustand liefert: grün die ROI, gelb die vom Benutzer eingegebene Selektionsposition.

Der Tracking-Algorithmus versucht nun, innerhalb der ROI Objekte zu detektieren und zu segmentieren, die sich unabhängig vom Hintergrund bewegen. Dazu wird die dominierende Bildbewegung des Hintergrunds, die von der Bewegung des Sensorträgers herrührt, mithilfe von Bild-zu-Bild-Registrierung eliminiert. Um einer Drift der initialen ROI entgegenzuwirken, wird Bewegungskompensation durchgeführt. Dazu wird die ROI von Bild zu Bild extrapoliert mithilfe der globalen Bildtransformation, die für die Bild-zu-Bild-Registrierung geschätzt wird. Die Anzeige der ROI bewegt sich im Bild mit und erscheint wie im Bildhintergrund an die Selektionsposition angeheftet.

Sobald der Tracking-Algorithmus ein Objekt nahe dem Zentrum der ROI detektiert, wechselt er in den Zustand TRACK und startet das Tracking-Modul für das Objekt. Detektiert der Tracking-Algorithmus in der ROI mehrere bewegte Objekte, wird das Objekt mit der kürzesten Distanz zur Selektionsposition als Trackingobjekt ausgewählt. Über einen Schwellenwert lässt sich die maximal erlaubte Distanz zwischen Objektmittelpunkt und ROI-Mittelpunkt konfigurieren. Abb. 6.16 zeigt die visuelle Rückmeldung des Algorithmus im Zustand TRACK: in Grün die ROI, in Magenta eingerahmt mit der Boundingbox das Trackingobjekt.

Der Übergang aus dem Zustand TRACK in den Zustand WAIT geschieht, wenn das Trackingmodul das Objekt nicht länger tracken kann. In diesem Fall muss der Benutzer erneut selektieren, um den Tracking-Algorithmus wieder in den Zustand INIT zu bringen und die Bewegtobjektdetektion für eine neue ROI zu starten. Der Übergang aus dem Zustand TRACK in den Zustand INIT geschieht außerdem, wenn der Benutzer während des Trackings erneut eine Bildposition selektiert. In diesem Fall bricht der aktuelle Trackingprozess ab und ein neuer Trackingprozess mit einer neuen ROI wird gestartet.

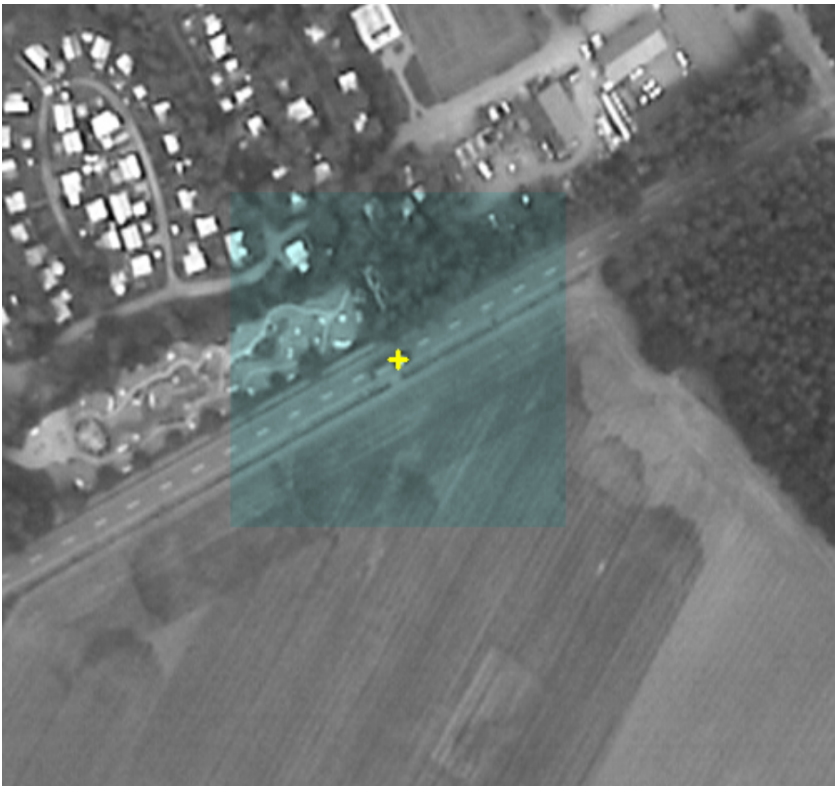


Abbildung 6.15: Visuelle Rückmeldung im Zustand INIT.

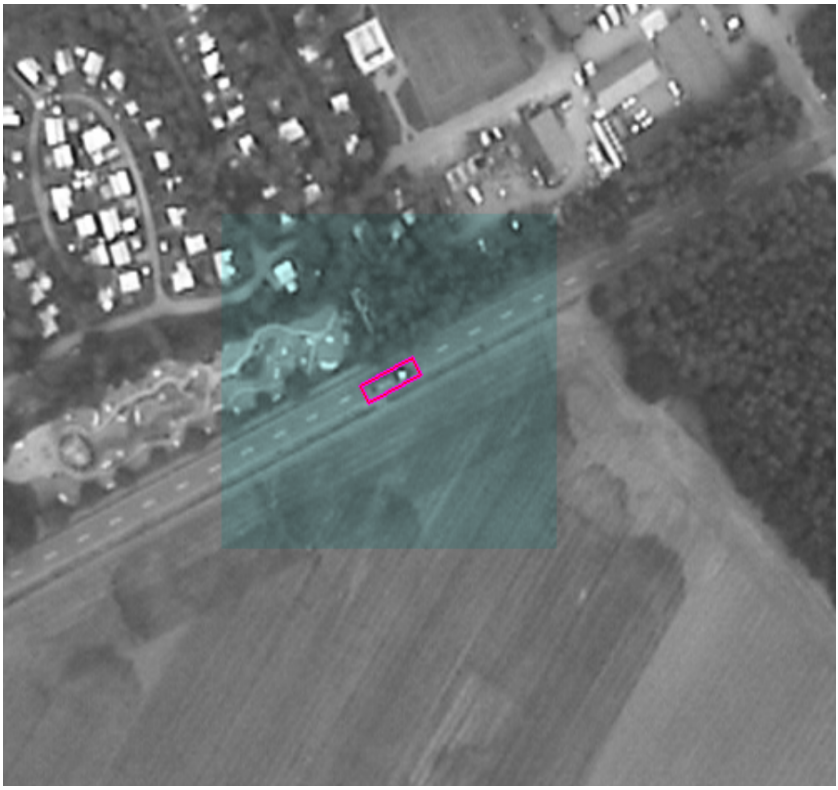


Abbildung 6.16: Visuelle Rückmeldung im Zustand TRACK.

6.2.2.2 Testaufgaben

Für die Nutzerstudie wurden 12 Testaufgaben gestaltet. Das Datenmaterial, aus dem die 12 zugehörigen Videoclips ausgeschnitten wurden, stammte aus zwei Full Motion Video-Sequenzen eines Überflugs von 60 bzw. 120 s Dauer.

Das Zielobjekt wurde nicht durch Hervorhebung markiert, sondern es wurde ein LKW als Zielobjekt festgelegt (vgl. Abb. 6.16), der in allen 12 Testaufgaben

zu selektieren war. Die Länge des LKW schwankte je nach Szene zwischen 22 und 30 Bildpixeln[1] (0,76°, 0,84 cm bzw. 1,03°, 1,14 cm).

Die Geschwindigkeit des Zielobjekts betrug im Mittel ± 1 Standardabweichung für die 12 Testaufgaben 101 ± 29 Bildpixel/s (3,47 ± 1,00°/s bzw. 3,83 ± 1,10 cm/s auf dem Monitor). Alle Testaufgaben enthielten vereinzelte starke Szenenschwankungen. Die resultierenden Höchstgeschwindigkeiten lagen zwischen 216 und 806 Bildpixel/s (7,42°/s, 8,21 cm/s bzw. 27,67°/s, 30,62 cm/s). Die Höchstgeschwindigkeit betrug im Mittel ± 1 Standardabweichung 512 ± 226 Bildpixel/s (17,57 ± 7,76°/s, 19,46 ± 8,58 cm/s).

Die Testaufgaben zeigten das Zielobjekt in unterschiedlichen Situationen. Sechs zeigten für die Realität der Videobildanalyse typische Situationen, in denen die Selektion eines Objektes schwierig ist: rasche Kameraschwenks, zeitweise Verdeckung des Zielobjekts durch Schatten oder Bäume oder räumlich nahe Distraktor-Objekte (PKWs, LKWs). Die Testaufgaben/Videoclips hatten eine Dauer zwischen 3 und 5 s, um einerseits genügend Zeit für die Durchführung einer Selektionsoperation zu lassen und andererseits doch etwas Zeitdruck zu erzeugen.

Die Videoclips wurden zu einer Sequenz zusammengefügt, wobei zwischen je zwei Videoclips für zwei Sekunden eine mittelgraue Szene angezeigt wurde, um die einzelnen Testaufgaben voneinander zu separieren. Die Videosequenz hatte eine Dauer von 90 Sekunden insgesamt.

Um sicherzustellen, dass die Versuchspersonen das Zielobjekt schnell in jeder Szene erkennen würden, wurden drei Trainingsaufgaben konzipiert. Trainingsaufgabe 1 und 2 nutzten die oben genannte Full Motion Video-Sequenz von 60 s Dauer, Trainingsaufgabe 3 diejenige von 120 s Dauer.

In Trainingsaufgabe 1 sollten die Versuchspersonen das Zielobjekt, auf das ein Tracker vorab initialisiert worden war, über die gesamte Zeitdauer hinweg beobachten; dies sollte mit der veränderlichen visuellen Erscheinung des Zielobjekts sowie des magentafarbenen Trackingrahmens vertraut machen.

[1] 1 Bildpixel umfasst auf dem Monitor 0,38 mm; die Pixelgröße des Monitors ist 0,27 mm, vgl. a. Abschnitt 6.1.1.1.

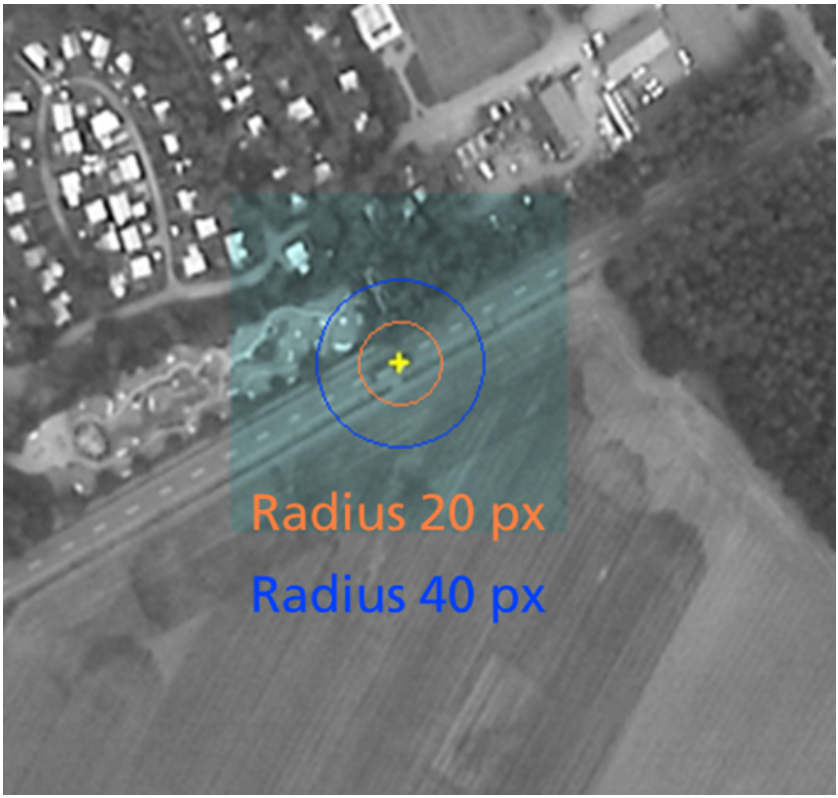


Abbildung 6.17: Die Selektionstoleranzen des automatischen Einzelobjekttracking-Verfahrens für Testbedingung A (20 Pixel Radius) und B (40 Pixel Radius) im Vergleich zur Größe des Zielobjekts.

In den Trainingsaufgaben 2 und 3 sollten die Versuchspersonen das Zielobjekt selektieren (ggf. mehrmals), sodass das Trackingverfahren erfolgreich initialisiert und der magentafarbene Rahmen kontinuierlich angezeigt würde.

Die Testaufgaben wurden mit zwei Konfigurationen der Selektionstoleranz des Einzelobjekttracking-Subsystems getestet. Wie oben beschrieben (Abb. 6.15 bzw. S. 268) sucht der Algorithmus auf dem kompletten grünen

Bereich nach bewegten Objekten. Der Suchbereich kann aber über einen Schwellenwert konfiguriert werden. In unserer Untersuchung wurde der Schwellenwert für die Selektionstoleranz in Testbedingung A auf einen Radius von 20 Bildpixeln gesetzt, in Testbedingung B auf einen Radius von 40 Bildpixeln (vgl. Abb. 6.17). Testbedingung A umfasst also einen Bereich mit Durchmesser 40 Bildpixel (1,37° bzw. 1,52 cm), Testbedingung B mit Durchmesser 80 Bildpixel (2,75° bzw. 3,04 cm). Testbedingung B toleriert also deutlich ungenauere Positionseingaben als Testbedingung A.

6.2.2.3 Versuchsaufbau, Versuchspersonen und Versuchsablauf

Das Versuchssystem ist ein Derivat des ABUL-Systems. Angezeigt werden die Trainings- und Testaufgaben auf einem 24-Zoll-Monitor (Auflösung 1920 x 1200 Pixel). Der Versuchsaufbau ist wie in Abb. 4.3.

Zur Blickerfassung wird der Tobii X60 verwendet (vgl. Abschnitt 4.2). Die Blickrohdaten werden gefiltert mit dem *RT-SDFS* [Kum08] (s. a. Abschnitt 2.4.1). Der Schwellenwert zur Trennung von Fixationen und Sakkaden wurde wie empfohlen auf $S = 20\ Pixel$ gesetzt. Da dieser Algorithmus eine zusätzliche Latenz von 1 Blickdatensample bewirkt, beträgt die Latenz, mit der das Blicksignal geliefert wird, jetzt 33 ms (X60) + 17 ms (Blickfilterung) = 50 ms.

Der manuelle Tastendruck bei Blick+Taste erfolgte mit der ENTER-Taste des NumPad einer gewöhnlichen Computertastatur. Die Mauseingabe erfolgte mit einer Comfort-Mouse 2000 for Business 3 (1000 dpi) von Microsoft.

Die 6 Versuchspersonen (alle männlich, Durchschnittsalter 26 Jahre) verfügten alle über normale oder auf normal korrigierte Sicht (1 mit Brille). Alle waren Studenten oder Kollegen und demzufolge ohne Erfahrung in Videobildauswertung. Alle waren langjährig erfahrene Benutzer von Desktop-Computern und Computermaus und ohne jegliche Erfahrung mit Eyetrackern oder Blickinteraktion.

Das Versuchsdesign nutzte ein vollständiges, ausbalanciertes Within-Subjects-Design, bei dem jede Versuchsperson die Testaufgabe je einmal mit

jeder der drei Interaktionsbedingungen durchführte. Um Ermüdungs- und Lerneffekte zu kontrollieren, wurden die Versuchspersonen in 6 Versuchsgruppen (mit je 1 Versuchsperson eingeteilt), wobei jede Versuchsgruppe eine der 6 möglichen Interaktionstechnik-Reihenfolgen durchführte.

Der Versuchsablauf war wie folgt. Nach einer kurzen Einführung mit Erläuterung der Aufgabe „Initialisierung eines Einzelobjekttrackers“ absolvierten die Versuchspersonen bei den blickbasierten Interaktionstechniken eine Standard-9-Punkte-Eyetrackerkalibierung. Darauf folgten zunächst die Trainingsaufgaben 1 (nur bei der ersten getesteten Interaktionsbedingung), 2 und 3; danach folgten die Testaufgaben, zunächst mit Testbedingung A (kleinere Selektionstoleranz), dann mit Testbedingung B. Für beide Testbedingungen A und B wurden die 12 Videoclips je 4-mal präsentiert. Die Instruktion war, die Tracker-Initialisierung so schnell und so genau wie möglich zu bewerkstelligen. Abschließend wurden die Versuchspersonen nach ihrer bevorzugten Interaktionstechnik gefragt.

6.2.2.4 Forschungsfragen, Metriken und Ergebnisse

Dieses Experiment adressierte folgende Forschungsfragen:

- Frage 1: Welche Trefferquote erzielen die Versuchspersonen bei der Tracker-Initialisierung? Wie schneiden die blickbasierten Interaktionstechniken im Vergleich zur Mauseingabe ab?
- Frage 2: Welche Selektionsgenauigkeit erzielen die Versuchspersonen?
- Frage 3: Welche Selektionszeit erzielen die Versuchspersonen?
- Frage 4: Wie wirkt sich die Blickvisualisierung aus? Ist sie nützlich und empfehlenswert?
- Frage 5: Welche Interaktionstechnik bevorzugen die Versuchspersonen?

Hierfür wurden folgende Metriken verwendet:

- Trefferquote: Sie berechnet sich als Anzahl erfolgreicher Tracker-Initialisierungen / Gesamtaufgabenumfang (= 4 ∗ 12 Objektselektionen). Eine Tracker-Initialisierung gilt als erfolgreich, wenn das Zielobjekt bis zum Ende des jeweiligen Videoclips den magentafarbenen Trackingrahmen aufweist.
- Selektionsgenauigkeit: Sie berechnet sich als euklidische Distanz zwischen Zielobjektmittelpunkt und Positionseingabe.
- Selektionszeit: Sie berechnet sich als Differenz des Selektionszeitpunkts und des Startzeitpunkts eines Videoclips.

Tabelle 6.5 zeigt die Ergebnisse, in der oberen Hälfte für Testbedingung A mit restriktiverer Selektionstoleranz, in der unteren Hälfte für Testbedingung B.

Die *Trefferquote* ist für alle Bedingungen (Interaktionstechniken und Testbedingungen) sehr hoch, wobei Blick+Taste+BlickVis geringfügig abfällt.

Die *Selektionsgenauigkeit* ist ebenfalls für alle Bedingungen sehr hoch und für alle Interaktionstechniken vergleichbar gut. Die im Mittel 14 Bildpixel von Blick+Taste (Mittelwert bei Testbedingung A) entsprechen 0.48° bzw. 5,3 mm auf dem Monitor, die 19 Bildpixel bei Testbedingung B entsprechen 0,65° bzw. 7,2 mm.

Die *Selektionszeit* ist für die restriktivere Testbedingung A erheblich länger als für die großzügigere Testbedingung B. Für Testbedingung B ist die Selektionszeit im Mittel für Mauseingabe um 23%, für Blick+Taste um 33%, für Blick+Taste+Blickvisualisierung um 35% kürzer.

Vergleicht man die Interaktionstechniken, so ist Blick+Taste für beide Testbedingungen die schnellste Interaktionstechnik, Blick+Taste+BlickVis die langsamste. Die Mauseingabe liegt für die restriktivere Testbedingung A gleichauf mit Blick+Taste, für Testbedingung B ist Blick+Taste jedoch im Mittel um 200 ms (12%) schneller als die Mauseingabe.

Tabelle 6.5: Ergebnisse für Trefferquote in Prozent, Selektionsgenauigkeit in Bildpixeln (BPx) und Selektionszeit in ms, jeweils als Mittelwert ± 1 Standardabweichung.

Testbedingung A (20 BPx)	Mauseingabe	Blick+Taste	Blick+Taste+BlickVis
Trefferquote	93,9 (5,0)	93,9 (6,0)	91,1 (6,3)
Sel.Genauigkeit	16 (3)	14 (2)	16 (3)
Selektionszeit	2032 (604)	2045 (490)	2380 (432)
Testbedingung B (40 BPx)			
Trefferquote	93,9 (5,7)	95,0 (5,0)	92,2 (5,3)
Sel.Genauigkeit	18 (2)	19 (4)	19 (2)
Selektionszeit	1566 (282)	1375 (198)	1554 (170)

Bezüglich der **Forschungsfragen** ergeben sich folgende Antworten.

Frage 1: Die Trefferquoten sind für alle Interaktionstechniken sehr hoch, Blick+Taste+BlickVis fällt geringfügig ab.

Frage 2: Die Selektionsgenauigkeit ist für alle Interaktionstechniken vergleichbar gut und variiert im Mittel zwischen 14 und 19 Bildpixeln (entspricht 0,48° bzw. 0,65° oder 5 mm bzw. 7 mm auf dem Monitor).

Frage 3: Die Selektionszeit ist mit 1375 ± 198 ms am kürzesten mit Blick+Taste für Testbedingung B mit der großzügigeren Selektionstoleranz von 40 Bildpixeln. Mauseingabe und Blick+Taste+BlickVis benötigen im Mittel 200 ms länger.

Frage 4: Es zeigte sich, dass die Blickvisualisierung nicht zu höherer Trefferquote oder besserer Selektionsgenauigkeit beitrug. Außerdem war ihre Selektionszeit für beide Testbedingungen am schlechtesten.

Frage 5: 66% (4 von 6) der Versuchspersonen bevorzugten Blick+Taste, 33% die Mauseingabe.

6.2.2.5 Fazit

Auch für die Aufgabe der Initialisierung eines Einzelobjekttrackers zeigte sich Blick+Taste als leistungsfähige Alternative zur Mauseingabe mit vergleichbar guter Effektivität bei besserer Effizienz (Selektionszeit im Mittel um 12% kürzer). Sie wurde zudem von 66% der Versuchspersonen favorisiert.

Die Blickvisualisierung brachte keine Verbesserung, sondern eher eine Verschlechterung der Ergebnisse. Blick+Taste+BlickVis wurde auch von keiner Versuchsperson als bevorzugte Interaktionstechnik genannt.

6.2.3 Initialisierung eines automatischen Verfahrens zum Einzelobjekttracking: Expertenstudie

Die Ergebnisse dieses Experiments wurden bei der AVSS 2019 veröffentlicht [Hil19].

Dieses Experiment umfasst die Expertenstudie zur Pilotstudie aus dem vorigen Abschnitt 6.2.2. Das Versuchsdesign variiert in folgenden Aspekten:

- Die Versuchspersonen sind Videobildauswerter und somit Experten in Bildfolgenanalyse.
- Es werden nur Blick+Taste und Mauseingabe getestet.
- Es wird nur Testbedingung B mit der großzügigeren Selektionstoleranz (Radius 40 Bildpixel) des Einzelobjekttracking-Verfahrens getestet.

6.2.3.1 Testaufgaben

Vgl. Beschreibung in Abschnitt 6.2.2.2.

6.2.3.2 Versuchsaufbau, Versuchspersonen und Versuchsablauf

Versuchssystem und Versuchsaufbau waren identisch wie bei der Untersuchung aus Abschnitt 6.2.2, s. Abschnitt 6.2.2.3. Versuchsort war jedoch nicht das Fraunhofer IOSB, sondern zwei Bundeswehr-Standorte.

18 Videoauswerteexperten (16 männlich, 2 weiblich; 5 unter 30 Jahre, 9 zwischen 30 und 41 Jahren, 4 über 41 Jahre) nahmen an der Nutzerstudie teil. Alle verfügten über normale oder auf normal korrigierte Sicht. Alle waren erfahrene Benutzer von Desktop-Computern und Computermaus, 16 hatten keine Erfahrung mit Eyetracking, 2 hatten die Erfahrung der Teilnahme an der Nutzerstudie aus Abschnitt 6.1.2.

Das Versuchsdesign nutzte ein vollständiges, ausbalanciertes Within-Subjects-Design, bei dem jede Versuchsperson die Testaufgabe je einmal mit Blick+Taste und einmal mit Mauseingabe durchführte. Um Ermüdungs- und Lerneffekte zu kontrollieren, wurden die Versuchspersonen in 2 Versuchsgruppen mit je 9 Versuchsperson eingeteilt, wobei jede Versuchsgruppe eine der 2 möglichen Interaktionstechnik-Reihenfolgen durchführte.

6.2.3.3 Forschungsfragen, Metriken und Ergebnisse

Dieses Experiment adressierte folgende Forschungsfragen:

- Frage 1: Welche Trefferquote erzielen die Versuchspersonen bei der Tracker-Initialisierung? Wie schneidet Blick+Taste im Vergleich zur Mauseingabe ab?
- Frage 2: Welche Selektionszeit erzielen die Versuchspersonen?
- Frage 3: Welche Interaktionstechnik bevorzugen die Versuchspersonen?

Die verwendeten Metriken sind dieselben wie in Abschnitt 6.2.2.4.

Tabelle 6.6 zeigt die Ergebnisse.

Tabelle 6.6: Ergebnisse der Expertenstudie für Trefferquote in Prozent und Selektionszeit in ms, jeweils als Mittelwert ± 1 Standardabweichung. Signifikant bessere Ergebnisse sind **fett** gedruckt.

Testbedingung B (40 Bildpixel)	Mauseingabe	Blick+Taste
Alle Testaufgaben (m=12)		
Trefferquote in %	97,1 (5,4)	97,1 (4,1)
Selektionszeit	2136 (375)	1982 (326)
Einfachere Testaufgaben (m=6)		
Trefferquote in %	99,1 (2,1)	99,1 (2,1)
Selektionszeit	1704 (279)	1671 (295)
Schwierigere Testaufgaben (m=6)		
Trefferquote in %	95,1 (6,8)	95,1 (4,6)
Selektionszeit	2594 (549)	**2274 (464)**

Bezüglich der **Forschungsfragen** ergeben sich folgende Antworten.

Frage 1: Betrachtet man alle Testaufgaben (Tabelle 6.6, oberes Drittel), so ist die Trefferquote sehr hoch und für Blick+Taste und Mauseingabe gleich gut. Bei einfacheren Testaufgaben (Tabelle 6.6, mittleres Drittel) liegt die Trefferquote noch etwas höher, bei schwierigeren Testaufgaben mit herausfordernden Selektionssituationen wie starke Szenenschwenks oder Verdeckungen (vgl. Abschnitt 6.2.2.2, S. 271) etwas niedriger.

Frage 2: Betrachtet man alle Testaufgaben (Tabelle 6.6, oberes Drittel), so ist die Selektionszeit mit 1982 ± 326 ms für Blick+Taste im Mittel ca. 150 ms (7%) kürzer als für Mauseingabe mit 2136 ± 375 ms; der Unterschied ist jedoch nicht signifikant (Zweiseitiger t-Test bei abhängigen Stichproben, $\alpha = 0{,}05$: $t(17)=1{,}57$; $p=0{,}067$). Betrachtet man nur die einfacheren Testaufgaben, so sind beide Interaktionstechniken gleich schnell. Für die schwierigeren Testaufgaben ist Blick+Taste signifikant schneller ($t(17)=1{,}97$; $p<0{,}05$) um 12%.

Es fällt auf, dass die Selektionszeiten der Experten länger waren als die der Nicht-Experten (vgl. Tabelle 6.5: M 1566 ± 282 ms, BT 1375 ± 198 ms). Dies liegt mutmaßlich daran, dass die Experten – obwohl auch sie die Instruktion erhalten hatten, so schnell und so genau wie möglich das Zielobjekt zu selektieren – den Schwerpunkt auf die Genauigkeit gelegt hatten, was auf Kosten der Schnelligkeit ging.

Frage 3: 56% (10 von 18) der Experten bevorzugten Blick+Taste, 33% die Mauseingabe, 11% möchten zwischen beiden wechseln können.

6.2.3.4 Fazit

Die Ergebnisse der Expertenstudie bestätigen die Ergebnisse der Pilotstudie. Blick+Taste ist bei der Aufgabe der Initialisierung eines Einzelobjekttracking-Verfahrens genauso effektiv wie die Mauseingabe. Blick+Taste ist zudem etwas schneller. Signifikant schneller waren die Videoexperten aber nur in schwierigen Selektionssituationen. Auch die Videoexperten nannten Blick+Taste in der Mehrheit (56%) als bevorzugte Interaktionstechnik, weitere 11% bevorzugten, zwischen Blick+Taste und Mauseingabe zu wechseln.

6.3 Übersicht über die Ergebnisse aus Abschnitt 6.1 und Abschnitt 6.2

Die Tabellen 6.7 und 6.8 fassen die Versuchsdesigns bzw. die wichtigsten Ergebnisse tabellarisch zusammen. Die Ergebnisse bestätigen die aus Kapitel 5: Blick+Taste ist eine effektive und effiziente Interaktionsalternative zur Mauseingabe für die Interaktion mit bewegten Objekten in Full Motion Video. Dies zeigte sich für die Interaktionsaufgaben Bewegtobjektmarkierung sowie Tracker-Initialisierung.

Bei einer Aufgabe zur Bewegtobjektmarkierung bewirkte die Verfügbarkeit von Verfahrensergebnissen eines automatischen Verfahrens zur Bewegtobjektdetektion eine signifikant niedrigere Selektionsfehlerquote. Diese durch

Informationsfusion aus Detektionsleistung des Verfahrens und Erkennungsleistung des Benutzers erzielte Ergebnis wurde untermauert von einer guten Bewertung der Zufriedenstellung: Die Kombination aus *Blick+Taste mit automatischer Bewegungsdetektion* wurde mit derselben hohen Zufriedenstellung bewertet wie Blick+Taste und erzielte im Vergleich zur Mauseingabe eine signifikant bessere Bewertung des Merkmals Ermüdung der Augen.

Von Videoanalyseexperten wie von Nicht-Experten wird Blick+Taste mehrheitlich als präferierte Interaktionstechnik genannt. Videoanalyseexperten ohne Erfahrung in blickbasierter Interaktion (N=26) bewerteten zudem den Aspekt der Ermüdung der Augen für Mauseingabe und Blick+Taste gleich gut.

Tabelle 6.7: Versuchsdesigns der fünf Untersuchungen aus Abschnitt 6.1 und Abschnitt 6.2.

Studie	Versuchsdesign							
	VP	Trials	Zielobjekte		Interaktionstechniken			
			Größe	Geschw.	**M**	**BT**	**BT+IMD**	**BT+BlickVis**
	N	m	°	°/s				
[Hil13b] S. 228	13	32	0,17-1,18[a]	3,09 (1,99)	×	×		
[Hil16a] S. 236	26	65	0,30-4,12[a]		×	×		
[Hil17] S. 251	10	67	0,45-2,26[a]	3,09 (1,99)				
		7	0,90-2,26[a]	2,06-13,05	×	×	×	
		27	0,45-2,26[a]	3,09 (1,99)				
		99	0,09-1,81[a]	3,09 (1,99)				
[Hil15] S. 265	6		0,76-1,03[b]		×	×		×
		48	0,76-1,03[c]	3,54 (1,02)				
[Hil19] S. 277	18		0,76-1,03[c]		×	×		

[a] Größe des Markierungsrahmens 2,06°
[b] Selektionstoleranz 1,37°
[c] Selektionstoleranz 2,75°

Tabelle 6.8: Zielobjekteigenschaften (Größe und Geschwindigkeit) mit Ergebnissen für die fünf Untersuchungen zur Bewegtobjektselektion an Full Motion Video-Bildfolgen aus Abschnitt 6.1 und Abschnitt 6.2. Trefferquote in Prozent und Selektionszeit in ms, jeweils als Mittelwert ± 1 Standardabweichung. Signifikant bessere Ergebnisse sind **fett** gedruckt.

Studie	Ergebnisse					
	Selektionsfehlerquote[a] %			Selektionszeit ms		
	M	**BT**	**BT+IMD bzw. BT+BlickVis**	**M**	**BT**	**BT+BlickVis**
[Hil13b] S. 228	13,6 (7,8)	14,2 (7,6)	-	-	-	-
[Hil16a] S. 236	14,3 (5,3)	**12,0 (5,5)**	-	1315 (466)	**775 (319)**	-
[Hil17] S. 251	24,0 (3,9)	24,0 (9,7)	**9,0 (6,7)**	-	-	-
	64,3 (12,1)	64,3 (13,9)	**42,9 (15,1)**	-	-	-
	19,6 (8,7)	21,1 (10,8)	15,9 (10,8)	-	-	-
	56,2 (5,8)	52,9 (10,9)	**24,0 (6,0)**	-	-	-
[Hil15] S. 265	6,1 (5,0)	6,1 (6,0)	8,9 (6,3)	2032 (604)	2045 (490)	2380 (432)
	6,1 (5,7)	5,0 (5,0)	7,8 (5,3)	1566 (282)	1375 (198)	1554 (170)
[Hil19] S. 277	2,9 (5,4)	2,9 (4,1)	-	2136 (375)	1982 (326)	-

[a] Ergebnisse von Selektionstrefferquoten [Hil17, Hil15, Hil19] wurden in Selektionsfehlerquoten umgerechnet.

6.4 Blickbasierte Klassifikation der Benutzertätigkeit bei der Bildfolgenanalyse

Die Ergebnisse dieses Experiments wurden bei der ETRA 2018 veröffentlicht [Hil18b]. Versuchsgestaltung, Implementierung des Versuchssystems, Versuchsdurchführung und erste Datenanalysen wurden im Rahmen einer Masterarbeit durchgeführt (vgl. [Küh17]).

Um die blickbasierte Klassifikation der Benutzertätigkeit bei der Bildfolgenanalyse untersuchen zu können, wurde eine Datenerhebung mit 30 Versuchspersonen durchgeführt, die die vier in Abschnitt 3.3 konzipierten Tätigkeiten *Erkunden, Beobachten, Suchen* und *Verfolgen* absolvieren.

6.4.1 Versuchsaufgaben

Im ersten Schritt wurde für jede der vier Tätigkeiten eine Versuchsaufgabe gestaltet. Als visueller Stimulus wurde eine 4-minütige Videosequenz genutzt, aufgenommen von einer ortsfesten PTZ-Kamera aus einer Höhe von 12 m in Schrägsicht mit einer unveränderlichen Kameraperspektive (vgl. Abb. 6.18). Die Dynamik im Video besteht primär aus Personen, die gehen oder radeln, sowie aus leichten Bewegungen der Blätter der Bäume.

Diese Videosequenz war der visuelle Stimulus für alle Versuchsaufgaben. So sollte sichergestellt werden, dass Unterschiede im Blickverhalten nicht von Unterschieden im Bildmaterial (mit unterschiedlichen Ereignissen) herrühren würden.

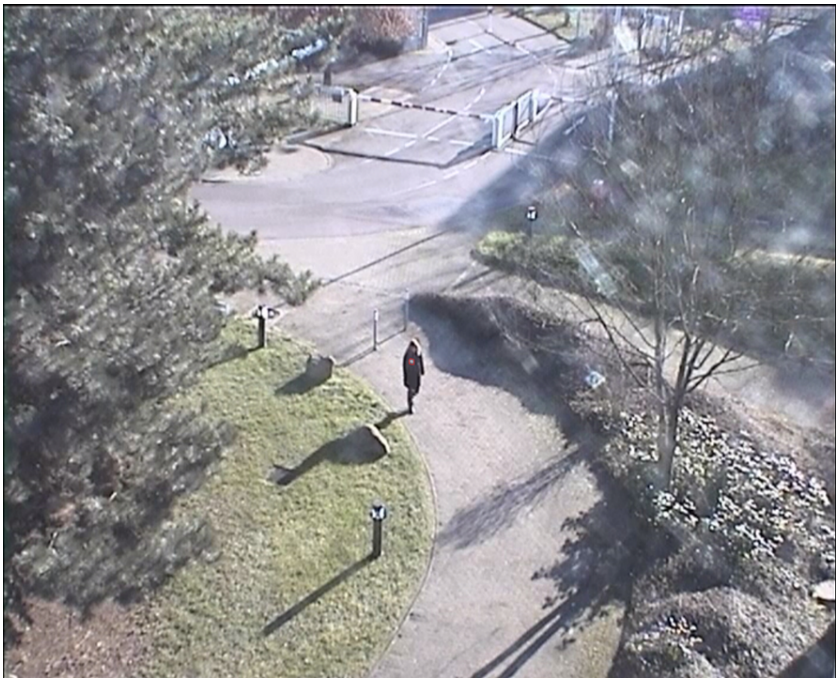


Abbildung 6.18: Szene aus der 4-minütigen Videosequenz des visuellen Stimulus.

Die Versuchsaufgaben waren wie folgt gestaltet:

- *Erkunden*:
  - Instruktion: Betrachten Sie die Szene mit dem Zweck, mit den Szeneninhalten vertraut zu werden.
  - Überprüfung: Um sicherzustellen, dass die Versuchspersonen die gesamte Szene erkunden würden, wurden unmittelbar nach Beendigung der Versuchsaufgabe Szeneninhalte abgefragt (dies wurde vor der Aufgabendurchführung angekündigt): Wie viele große Steine lagen im Gras? Welche Art Baum (Laubbaum oder Nadelbaum) war am linken

Bildrand zu sehen? Wie viele Personen erschienen? In welche Richtung fielen die Schatten?

- *Beobachten*:
  - Instruktion: Beobachten Sie die Szene sorgfältig mit dem Zweck, drei Regelverstöße zu entdecken: Nicht erlaubt sind Radeln, Betreten der Wiese auf der linken Szenenhälfte, Betreten des schmalen Weges in der rechten Szenenhälfte.
  - Überprüfung: Unmittelbar nach Beendigung der Versuchsaufgabe wurde die Gesamtanzahl Regelverstöße abgefragt.
- *Suchen*:
  - Instruktion: Suchen Sie nach Personen, die entweder ein weißes Schild an der Jacke oder am Taillenbund tragen oder eine Tasche/Rucksack tragen.
  - Überprüfung: Unmittelbar nach Beendigung der Versuchsaufgabe wurde die Gesamtzahl Ereignisse (Schilder und Taschen) abgefragt.
- *Verfolgen*:
  - Instruktion: Behalten Sie stets die größte Personengruppe im Blick, die gerade in der Szene zu sehen ist, und verfolgen Sie sie mit den Augen. Die kleinste Gruppe besteht aus einer einzelnen Person. Sind zwei gleich große Gruppen zugleich in der Szene zu sehen, so verfolgen Sie diejenige, die sich näher am unteren Rand der Szene befindet.
  - Überprüfung: Keine.

Die 4-minütige Videosequenz ist oft mit vielen Personen belebt; gelegentlich sind nur wenige Vordergrundereignisse zu sehen, bspw. nur eine einzige Person. Aus diesem Grund hatten wir die Befürchtung, dass bei der Tätigkeit *Erkunden* die Versuchspersonen in Situationen mit wenigen Ereignissen das

Erkunden einstellen würden, insbesondere je länger die Videosequenz fortschreitet, da es dann nichts Neues zu erkunden gibt. Aus diesem Grund wurden für *Erkunden* nur die ersten 30 s der Videosequenz präsentiert.

6.4.2 Datenerhebung

An der Datenerhebung nahmen 30 Versuchspersonen (10 weiblich, 20 männlich; Altersdurchschnitt 35 Jahre, 16 zwischen 18 und 30 Jahren, 11 zwischen 30 und 50 Jahren, 3 älter als 50 Jahre) teil. Alle hatten normale oder auf normal korrigierte Sicht, 1 trug Kontaktlinsen, 9 eine Brille. Keine Versuchsperson verfügte über Expertise in Videoanalyse.

Die Präsentation der Versuchsaufgaben erfolgte auf einem 24-Zoll-Monitor (Auflösung 1920 x 1200+Pixel). Die Videosequenz (Original aufgezeichnet mit einer Auflösung von 704 x 576 Pixel) wurde mittig auf dem Monitor in einer Größe von 1198 x 980 Pixel (27,9° x 22,9° Sehwinkel) präsentiert.

Zur Aufzeichnung der Blickdaten wurde der Tobii X60 ohne Kopfstabilisierung verwendet; der Versuchsaufbau war wie in Abb. 4.3.

Die Tätigkeiten wurden stets in der Reihenfolge *Erkunden, Beobachten, Suchen, Verfolgen* präsentiert, damit die Versuchspersonen jede Versuchsaufgabe so naiv wie möglich beginnen. Denn obwohl für jede Versuchsaufgabe andere Szeneninhalte relevant sind, ist es nicht möglich zu verhindern, dass die Vertrautheit der Versuchspersonen mit den Szeneninhalten mit jeder Aufgabe zunimmt. Die Reihenfolge wurde so festgelegt, dass die Versuchspersonen minimales Vorwissen für die jeweils nächste Versuchsaufgabe gewinnen würden. Zudem wurden alle Instruktionen erst unmittelbar vor der zugehörigen Versuchsaufgabe gegeben.

Jede Versuchsperson absolvierte alle Versuchsaufgaben in einer Sitzung. Zu Beginn wurde der Ablauf der Datenerhebung erläutert. Dann folgte eine Standard-Tobii-9-Punkte-Kalibrierung, an die sich eine Validierung der Kalibrierung anschloss, bei der die Versuchspersonen 9 weitere Punkte fixierten; einer war in der Bildschirmmitte platziert, die übrigen so, dass sie die Präsentationsfläche der Videosequenz begrenzten.

Kalibrierung sowie Validierung der Kalibrierung wurden so lange wiederholt, bis für jeden Validierungspunkt die räumliche Genauigkeit < 1,5° Sehwinkel betrug, was für alle Versuchspersonen erreicht wurde. Für 13 Versuchspersonen wurde ein Wert < 1,0° Sehwinkel erzielt (vgl. DeAngelus u. a. [DeA09], die in ihrem Experiment < 1° forderten). Da wir jedoch Kontaktlinsen und Brillen erlaubten und die Datenaufzeichnung zudem ohne Kopfstabilisierung durchführten, schwächten wir diese Forderung auf 1,5° ab. Darauf absolvierten die Versuchspersonen die vier Versuchsaufgaben.

6.4.3 Merkmalsextraktion und Klassifikation

Aus den aufgezeichneten Blickrohdaten wurden mithilfe des I-VT-Algorithmus [Sal00] (vgl. Abschnitt 2.4.1, S. 55) Fixationen extrahiert. Der Schwellenwert v_{max}, der Fixationspunkte und Sakkadenpunkte trennt, wurde für jede Versuchsperson individuell gewählt, da die Menge an Rauschen in den Blickrohdaten individuell unterschiedlich war. Jeder individuelle Schwellenwert wurde aus dem individuellen Ergebnis der Validierung der Kalibrierung berechnet, und zwar als durchschnittliche Varianz aller neun Validierungspunkte. Die Schwellenwerte für v_{max} reichten von 30°/s bis 130°/s.

Auf Basis der erhaltenen Fixationen wurden die aggregierten Fixations- und Sakkaden-basierten Merkmale für den Merkmalsvektor des Klassifikators bestimmt: Fixationen pro Sekunde (arith. Mittel), Fixationsdauer (arith. Mittel, Varianz), Fixationsdurchmesser (arith. Mittel), Sakkadenamplitude (arith. Mittel, Varianz), Sakkadengeschwindigkeit (arith. Mittel, Varianz), Fixationswinkel (arith. Mittel, Varianz).

Für die Klassifikation wurden als Klassifikationsverfahren Random Forest (RF), Lineare Diskriminanzanalyse (LDA) und Quadratische Diskriminanzanalyse (QDA) angewendet, wofür die Smile library[1] (Statistical Machine Intelligence and Learning Engine) genutzt wurde.

[1] Vgl. https://haifengl.github.io/smile/

Bei allen wurde die Hälfte aller Datensamples als Trainingsmenge genutzt, die andere Hälfte als Testmenge. Danach wurden die Rollen für eine 50:50-Kreuzvalidierung getauscht.

Als Anzahl Entscheidungsbäume für den Random Forest wurden 20, 50, 80, 100, 150, 200, 250 und 500 getestet. Die besten Ergebnisse wurden mit 100 erzielt. Jeder der Entscheidungsbäume enthielt so viele Datensamples wie verfügbar waren, zufällig gewählt aus den vier Aufgaben.

6.4.4 Ergebnisse

Die Klassifikation wurde in verschiedenen Varianten durchgeführt. Variante (A) klassifiziert alle ***vier Tätigkeiten***. Datenbasis sind hier die Blickrohdaten der ***ersten 30 Sekunden*** der 4-minütigen Videosequenz, da für *Erkunden* nur die ersten 30 s genutzt wurden (s.o.); auf diese Weise basiert die Klassifikation für alle Tätigkeiten auf gleich vielen Blickrohdaten. Variante (B) klassifiziert die ***drei Tätigkeiten*** *Beobachten, Suchen* und *Verfolgen* auf Basis der Blickrohdaten der ***gesamten 4-minütigen*** Videosequenz.

Das beste Klassifikationsergebnis für die Klassifikation aller ***vier Tätigkeiten*** (Variante A) wurde mit LDA erzielt. Die Korrektklassifikationsrate betrug 59,5% (95% KI=49,8-68,0%, Zufallswahrscheinlichkeit 25%). Abb. 6.19 zeigt die Konfusionsmatrix.

Man sieht, dass die überwiegende Anzahl Datensamples zur korrekten Tätigkeit zugeordnet wurde. Konfusion tritt vor allem zwischen den Klassen *Erkunden* und *Beobachten* sowie zwischen *Suchen* und *Verfolgen* auf.

Letzteres leuchtet ein, sobald man die 30 s Bilddatenmaterial bzgl. ihrer Bildinhalte analysiert. Tatsächlich sind hier nur 2 Personen zu verfolgen: eine von Sekunde 5 bis 15, eine von Sekunde 15 bis 32. Betrachtet man die Tätigkeit *Suchen*, so trägt tatsächlich die erste der 2 Personen ein Schild; um dieses Schild erkennen zu können, müssen die Versuchspersonen diese Person eine kurze Weile mit den Augen verfolgen.

Betrachtet man die Instruktionen für *Erkunden* und *Beobachten*, so fällt auf, dass innerhalb der 30 Sekunden keine Regelübertretung stattfindet. Es ist daher anzunehmen, dass die Versuchspersonen einfach die Bereiche, die auf Regelübertretungen zu überprüfen sind, visuell durchmustern. Diese Bereiche – die Grasfläche und der schmale Weg – umfassen aber eine recht große Fläche der Gesamtszene, sodass es einleuchtet, dass die Blickmuster ähnlich sind wie bei der Tätigkeit *Erkunden*, wo die gesamte Szene visuell zu durchmustern ist.

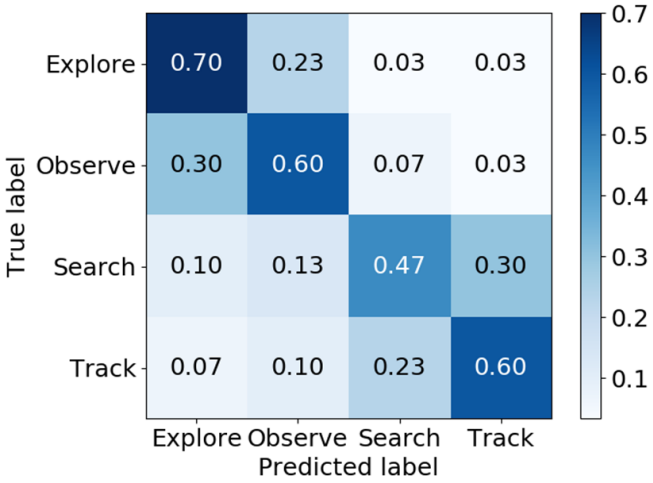


Abbildung 6.19: Konfusionsmatrix für die Klassifikation aller vier Tätigkeiten für das beste Klassifikationsergebnis von 59,5% (95% KI=49,8-68,0%, Zufallswahrscheinlichkeit 25%), erzielt mit LDA.

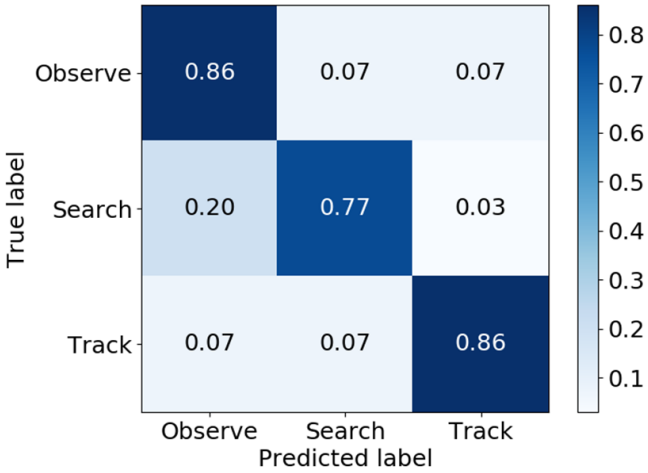


Abbildung 6.20: Konfusionsmatrix für die Klassifikation der drei Tätigkeiten *Beobachten, Suchen* und *Verfolgen* für das beste Klassifikationsergebnis von 83,7% (95% KI=74,0-90,4%, Zufallswahrscheinlichkeit 33%), erzielt mit Random Forest.

Das beste Klassifikationsergebnis für die Klassifikation der ***drei Tätigkeiten*** *Beobachten, Suchen* und *Verfolgen* (Variante B) wurde mit Random Forest erzielt. Die Korrektklassifikationsrate betrug 83,7% (95% KI=74,0-90,4%, Zufallswahrscheinlichkeit 33%). Abb. 6.20 zeigt die Konfusionsmatrix. Man sieht, dass die überwiegende Anzahl Datensamples zur korrekten Tätigkeit zugeordnet wurde.

Ergänzend zu diesen beiden Klassifikationen wurden weitere betrachtet, bei denen die Merkmale der Merkmalsvektoren auf Basis deutlich kürzerer Blickdatenausschnitte von 5 bzw. 3 Sekunden Länge bestimmt wurden. Denn wenn tatsächlich auf Basis der blickbasierten Tätigkeitsklassifikation die Benutzerintention geschätzt und eine passende Systemeingabe getätigt werden soll

(vgl. Abschnitt 1.3.3), muss die Klassifikation auf Basis deutlich kürzerer Blickdatenausschnitte funktionieren.

Die Ergebnisse waren folgendermaßen. Bei Klassifikation aller ***vier Tätigkeiten*** beträgt die beste erzielte Korrektklassifikationsrate auf der Basis von Blickdatensamples von ***5 s*** Dauer 45.3% (95% KI=41.6-49.0%, 25% Zufallswahrscheinlichkeit), erzielt mit LDA. Auf der Basis von Blickdatensamples von ***3 s*** Dauer beträgt die beste erzielte Korrektklassifikationsrate 38.0% (95% KI=35.2-40.8%, 25% Zufallswahrscheinlichkeit), ebenfalls erzielt mit LDA.

Bei Klassifikation der ***drei Tätigkeiten*** *Beobachten, Suchen* und *Verfolgen* beträgt die beste erzielte Korrektklassifikationsrate auf der Basis von Blickdatensamples von 5 s Dauer 52,3% (95% KI=50,8-53,8%, 33% Zufallswahrscheinlichkeit), erzielt mit LDA. Auf der Basis von Blickdatensamples von 3 s Dauer beträgt die beste erzielte Korrektklassifikationsrate 47,7% (95% KI=46,5-48,9%, 33% Zufallswahrscheinlichkeit), ebenfalls erzielt mit LDA.

Tabelle 6.9 zeigt die Details der Merkmale und der erzielten Ergebnisse. Die Klassifikation wurde für alle möglichen Merkmalsmengen durchgeführt. Die Tabelle zeigt nur die Ergebnisse der vollständigen Merkmalsmenge sowie für alle Mengen mit je einem ausgelassenen Merkmal, da mit kleineren Merkmalsmengen schlechtere Ergebnisse erzielt wurden.

Tabelle 6.9: Klassifikationsergebnisse als Korrektklassifikationsrate in Prozent, abhängig von der Dauer der Blickdatensamples. Die besten Ergebnisse sind ***fett kursiv*** gedruckt. Merkmale (abgekürzt) von links nach rechts: FixationsDauer, FixationsDurchMesser, Fixationen pro Sekunde, FixationsWinkel, SakkadenAmplitude, SakkadenGeschwindigkeit.

Dauer	Merkmale						Korrektklassifikationsraten					
							A: 4 Tätigkeiten (Klassen)			B: 3 Tätigkeiten (Klassen)		
	FD	FDM	FpS	FW	SA	SG	RF	LDA	QDA	RF	LDA	QDA
A: 30s ; B: 4min	x	x	x	x	x	x	46,8	56,5	44,0	81,3	70,0	71,3
		x	x	x	x	x	40,8	53,3	50,0	76,3	75,3	75,7
	x		x	x	x	x	48,5	56,5	49,3	***83,7***	70,0	74,7
	x	x		x	x	x	50,0	56,8	47,5	79,0	67,7	77,0
	x	x	x		x	x	41,8	50,0	42,5	67,7	63,7	65,7
	x	x	x	x		x	52,8	***59,3***	42,5	80,0	66,7	71,3
	x	x	x	x	x		52,0	55,0	45,8	81,3	68,0	75,7
5s	x	x	x	x	x	x	42,0	44,3	38,0	51,7	52,0	49,3
		x	x	x	x	x	41,8	42,5	38,8	51,0	50,7	46,7
	x		x	x	x	x	42,3	***45,3***	36,5	51,3	***52,3***	50,0
	x	x		x	x	x	44,3	43,0	38,3	52,0	51,7	47,7
	x	x	x		x	x	36,8	40,5	36,0	51,0	51,0	48,0
	x	x	x	x		x	38,8	44,3	37,0	51,0	***52,3***	49,0
	x	x	x	x	x		43,0	43,8	36,5	52,0	51,7	47,0
3s	x	x	x	x	x	x	36,8	36,3	35,8	47,0	47,3	41,3
		x	x	x	x	x	37,5	37,5	36,0	47,3	***47,7***	40,7
	x		x	x	x	x	37,8	37,5	36,5	46,3	47,3	41,7
	x	x		x	x	x	37,8	37,3	37,0	47,0	***47,7***	41,7
	x	x	x		x	x	36,3	37,3	34,5	46,7	47,3	41,3
	x	x	x	x		x	37,3	***38,0***	47,3	47,3	66,7	41,3
	x	x	x	x	x		36,5	37,0	35,5	47,0	***47,7***	45,3

Insgesamt lässt sich feststellen, dass die Klassifikation sowohl für vier und drei Tätigkeiten als auch mit unterschiedlicher Länge der Blickdatensamples (30 s bzw. 4 min, 5 s, 3 s) deutlich über Zufallswahrscheinlichkeit gelang. Die Korrektklassifikationsraten sind jedoch in allen Fällen nicht hoch genug, um darauf basierend Systemeingaben zu tätigen.

Verbesserungspotenzial liegt zum einen in einer Überarbeitung und Anpassung der Merkmale für den Merkmalsvektor. In unserer Untersuchung wurden oft genutzte und einfach bestimmbare Merkmale des Blickverhaltens genutzt. Eine Möglichkeit der Ergänzung sind Merkmale, die die zeitliche Struktur des Blickpfads besser erfassen; Kanan u. a. [Kan15] nutzt solche Merkmale z. B. für die blickbasierte Klassifikation, wenn Versuchspersonen Gesichter bewerten, Le Meur u. a. [Le 17] nutzt sie für die blickbasierte Klassifikation der Altersgruppe eines Beobachters aus seinem Blickverhalten während der Betrachtung von Einzelbildern.

Zum anderen besteht Verbesserungspotenzial bei der Wahl des Klassifikationsverfahrens. Auch diesbezüglich wurden in unserer Untersuchung, die sich als früher Schritt bei der Klassifikation von Tätigkeiten bei der Bildfolgenanalyse versteht, einfache, in der Fachliteratur oft genutzte Verfahren verwendet.

7 Zusammenfassung und Ausblick

Ziel der vorliegenden Arbeit ist, für die Anwendungsaufgabe der Bildfolgenanalyse eine leistungsfähige und belastungsarme Benutzungsschnittstelle zu gestalten, indem diese um blickbasierte Interaktion ergänzt wird. Es zeigte sich, dass Eyetracking zur Informationseingabe die Interaktion mit bewegten Objekten in dynamischen Szenen unterstützen kann. Die dazu geleisteten Forschungsarbeiten sind in den folgenden drei Abschnitten beschrieben.

Es ist festzuhalten, dass bei den meisten Untersuchungen die Anzahl Versuchspersonen vergleichsweise gering war, sodass die statistische Aussagekraft der Ergebnisse nur begrenzt ist und sie als vorläufig anzusehen sind.

7.1 Beitrag 1: Blickbasierte Selektion bewegter Objekte

Selektionsoperationen werden heute typischerweise mit manuellen Interaktionstechniken bewerkstelligt, an Desktop-Computern üblicherweise mit der Computermaus. Während dies für statische Objekte angemessen ist, ist es bei bewegten Objekten in dynamischen Szenen herausfordernd und fehlerträchtig. Höhere Geschwindigkeiten erschweren die Selektion [Jag80, Hof91, Has11]. Deshalb schlägt die Fachliteratur vor, bewegte Objekte während des Selektionsvorgangs zu vergrößern, z. B. auf ca. 2° bei langsameren Geschwindigkeiten von ca. 1,5-3,5°/s [Rag20] und bis zu ca. 14° bei sehr hohen Geschwindigkeiten von 19°/s [Has11]. Wenngleich diese Maßnahme die Selektionsfehlerquote reduziert, bleibt für den Benutzer dennoch eine hohe kognitive und manuelle Belastung.

Zielsetzung in Beitrag 1 ist daher, blickbasierte Interaktionstechniken zu identifizieren, mit denen ein Benutzer die Bewegtobjektselektion mit geringerer Belastung bei gleicher oder besserer Leistung im Vergleich zur Mauseingabe durchführen kann. Von besonderem Interesse ist, eine sehr schnelle Selektion zu gewährleisten. Denn bei der Anwendungsaufgabe der Bildfolgenanalyse bewegen sich Objekte nicht nur, sondern sind zudem oft zeitlich nur begrenzt in der Szene sichtbar.

In insgesamt 8 Nutzerstudien werden vier multimodale blickbasierte Interaktionstechniken einerseits an abstrakten, andererseits an realitätsnahen Testaufgaben mit Bewegtobjektselektion evaluiert: die expliziten Interaktionstechniken **Blick+Taste, Blick+Fußtaste, MAGIC pointing** sowie die implizite Interaktionstechnik **Blick+EEG**. Die Bewertung erfolgt mithilfe von Maßen der Effektivität (Selektionsfehlerquote, Selektionsgenauigkeit), der Effizienz (Selektionszeit) sowie der Zufriedenstellung der Versuchspersonen (Fragebogen der DIN CEN ISO/TS 9241-411 [DIN14], Frage nach der Präferenz).

Die abstrakten Versuchsaufgaben sind in Anlehnung an die Gegebenheiten der Anwendung Bildfolgenanalyse sowie der Fachliteratur gestaltet. Die (sichtbaren) Objektgrößen und Objektgeschwindigkeiten entsprechen den Gegebenheiten von Objekten in Überflugvideos; die Größen betragen zwischen 0,17° und 2,38°, die Geschwindigkeiten liegen zwischen 1,40°/s und 19°/s. Aus der Fachliteratur ist die Maßnahme der virtuellen Objektvergrößerung übernommen, indem für die Objekte selektionssensitive (unsichtbare) Größen zwischen 1,43° und 8,27° realisiert werden.

Blick+Taste (BT) nutzt die vom Eyetracker gelieferte Blickposition zum Zeigen und die ENTER-Taste zur Selektionsauslösung. Sie wird in 5 Nutzerstudien mit insgesamt N=73 Versuchspersonen (VP) (darunter 26 Videoanalyseexperten (VAExp)) evaluiert. Die Ergebnisse zeigen, dass BT eine Alternative zur Mauseingabe (M) ist. BT erzielt bei vergleichbar guter Selektionsfehlerquote und Zufriedenstellung eine erheblich kürzere Selektionszeit und wird von einer deutlichen Mehrheit der Versuchspersonen als präferierte Interaktionstechnik genannt.

Als ***Selektionsfehlerquote*** erzielt ***BT*** bei abstrakten Versuchsaufgaben (N=34; Objektgröße zwischen 0,60° [selektionssensitiv unsichtbar vergrößert auf 4,01°] und 2,38° [4,75°]; Objektgeschwindigkeit 1,40°/s bis 19°/s) geschwindigkeitsabhängig zwischen 0% und 73,8%. Die Ergebnisse von BT sind ***vergleichbar gut wie*** die der ***M***; in einer Studie mit N=12 ist BT für hohe Geschwindigkeiten (8,3°/s, 15,5°/s, 19°/s) ***signifikant besser*** (vgl. Abschnitt 5.1.3 bzw. Abb. 5.33).

Bei realitätsnahen Versuchsaufgaben mit niedrigeren Geschwindigkeiten ist ***BT vergleichbar gut*** wie ***M*** (N=39, davon 26 VAExp; 0,17° [3,15°] bis 4,12° x 3,43° [8,27°]; Objektgeschwindigkeit mit Mittelwert ± 1 Standardabweichung 3,09 ± 1,99°/s).

Die ***Selektionsgenauigkeit*** wurde in 3 der 5 Nutzerstudien (N=51, davon 26 VAExp) bestimmt. Sie liegt für ***BT*** geschwindigkeitsabhängig zwischen 0,57 ± 0,17° und 1,92 ± 0,17°. Sie ist bei abstrakten Versuchsaufgaben (N=12; 2,38° [4,75°]; 4,8°/s bis 19°/s) bis 11,9°/s für BT ***vergleichbar gut wie*** für ***M***, für 15,5° und 19° ***geringfügig schlechter*** für BT; der Unterschied beträgt im Mittel ca. 0,18° (ca. 2 mm auf dem Monitor), was bei der gegebenen Objektgröße keine Auswirkung auf die Selektionsfehlerquote hat.

Bei realitätsnahen Versuchsaufgaben zeigt sich ein ähnliches Bild. ***BT*** ist ***vergleichbar gut*** (N=13; 0,17° [3,15°] bis 1,18° [4,19°]; 3,09 ± 1,99°/s) oder ***geringfügig schlechter*** (N=26 VAExp; 0,30° [3,33°] bis 4,12° x 3,43° [8,27°]; 3,09 ± 1,99°/s). Der Unterschied beträgt im Mittel ca. 0,28° (ca. 3 mm), was bei der gegebenen Objektgröße keine Auswirkung auf die Selektionsfehlerquote hat. Nachträgliche Filterung der Blickrohdaten mit dem echtzeitfähigen Blickfilterverfahren von Kumar u. a. [Kum08] verringert den Unterschied auf ca. 2 mm. Bei Brillenträgern (N=11) sind Selektionsgenauigkeit und Selektionsfehlerquote etwas schlechter als bei Nicht-Brillenträgern (N=15).

Als ***Selektionszeit*** erzielt ***BT*** zwischen 335 und 786 ms. Die Unterschiede sind erklärbar durch die Art der Aufgabenstellung und den Abstand der Blickposition (als Zeigegerät) vom Objekt bei Beginn einer Selektionsoperation. BT ist ***signifikant*** und teilweise ***erheblich schneller*** als M, bei den abstrakten

Versuchsaufgaben um bis zu 46%, bei den realitätsnahen Versuchsaufgaben mit 26 VAExp um 41%.

Bezüglich der ***Zufriedenstellung*** erzielt ***BT*** für fast alle Merkmale stets gute bis sehr gute Bewertungen. BT schneidet zudem stets für ***fast alle Merkmale gleich gut oder besser*** ab als M.

Einzige Ausnahme ist das Merkmal ***Ermüdung der Augen bei ungeübten Benutzern***. Bei realitätsnahen Versuchsaufgaben bewerten die Nicht-Experten (unerfahren mit BT und in Bildfolgenanalyse) in einer Studie (N=13) BT ***signifikant schlechter*** als M, in einer anderen (N=10) ***signifikant besser***. Die ***VAExp*** (unerfahren nur mit BT) hingegen bewerten BT und M ***vergleichbar gut***. Die VP (N=4) der ***Längsschnittstudie*** bewerten vor Beginn ihrer Trainingsphase (unerfahren mit BT und in Bildfolgenanalyse) die Ermüdung der Augen für BT schlecht (3,0 ± 0,0 auf einer Skala von 1 bis 7; 7 beste Bewertung) und ***nach 6 Monaten Training signifikant besser*** (4,8 ± 0,5).

Als ***bevorzugte Interaktionstechnik*** (nachgefragt in 3 der 5 Studien, N=57) nennen 45 VP (79%) ***BT***, 7 VP nennen M, 5 VP sind unentschieden und würden zwischen beiden abwechseln wollen.

Die ***Längsschnittstudie*** (N=4; 4 abstrakte Versuchsaufgaben, wöchentliches Training von 2-mal 1 Stunde pro Woche über 6 Monate) zeigt neben der Verbesserung der subjektiven Bewertung der Ermüdung der Augen eine teils signifikante Verbesserung der Selektionsfehlerquote, die sich je nach Aufgabenschwierigkeit nach 1 bis 3 Monaten stabilisiert.

Blick+Fußtaste (BFT) nutzt die Blickposition zum Zeigen und eine Fußtaste zur Selektionsauslösung. Sie erzielt im Vergleich zur Mauseingabe eine vergleichbar gute Selektionsfehlerquote bei erheblich kürzerer Selektionszeit. Die Ergebnisse sind jedoch etwas schlechter verglichen mit BT. BFT könnte also eine Alternative zur M sein in Fällen, wo manuelle Interaktionen vollständig vermieden werden müssen.

Die Ergebnisse für BFT fußen auf einer Nutzerstudie mit N=12 Versuchspersonen (Objektgröße 2,38° [4,75°]; Objektgeschwindigkeiten 4,8°/s bis 19°/s).

Die ***Selektionsfehlerquote*** liegt geschwindigkeitsabhängig zwischen 5,5% und 83,7%. Sie ist ***vergleichbar gut wie M***, jedoch ***signifikant schlechter als BT*** für Geschwindigkeiten ab 8,3°/s. Die ***Selektionsgenauigkeit*** ist ***vergleichbar gut wie BT***, für 15,5° und 19° ***geringfügig schlechter als M*** (Unterschied ca. 0,27° oder 3 mm). Die ***Selektionszeit*** ist für alle Geschwindigkeiten ***signifikant kürzer als M*** und geringfügig länger als BT (ca. 50 ms geschwindigkeitsunabhängig).

MAGIC pointing (MP) wird für drei Varianten in 2 Nutzerstudien (N=30 VP) an abstrakten Versuchsaufgaben untersucht. Zwei Varianten sind die von Zhai u. a. [Zha99] für die Selektion statischer Objekte eingeführten ***MP liberal*** und ***MP konservativ***. Die dritte Variante – ***MAGIC Button*** – wird im Rahmen der vorliegenden Arbeit neu vorgeschlagen.

Im Gegensatz zu BT und BFT wird bei MP der Blick nur zum groben Zeigen auf ein Zielobjekt genutzt, um die Amplitude der manuellen Zeigerverschiebung zu verkleinern und dadurch die manuelle Belastung zu verringern; das feine Zeigen und die Selektionsauslösung übernimmt eine abschließende Mauseingabe. Entscheidend bei MP ist die Art und Weise, in der die Kontrolle des Mauszeigers zwischen Blickposition und manueller Mauskontrolle wechselt. Bei MP liberal ist der Mauszeiger permanent an der Blickposition visualisiert, außer der Benutzer bewegt die Maus. Bei MP konservativ springt der Mauszeiger erst dann an die Blickposition, wenn der Benutzer die Maus bewegt. Bei ***MAGIC Button*** springt der Mauszeiger an die Blickposition, wenn der Benutzer die rechte Maustaste klickt.

Für ***MP liberal*** und ***MP konservativ*** sind sowohl die Selektionsfehlerquote als auch die Selektionszeit ***signifikant schlechter als*** für die ***Mauseingabe*** und sie erzielen deutlich schlechtere subjektive Bewertungen der Zufriedenstellung.

MAGIC Button erzielt eine ***Selektionsfehlerquote*** von 2,4%, ***vergleichbar gut wie M*** mit 1,1% (Unterschied nicht signifikant) (N=12; 0,71°; 2,88°/s bis 3,37°/s) . Die ***Selektionszeit*** ist ***ebenfalls nicht signifikant schlechter***.

Für die ***Zufriedenstellung*** erzielt MAGIC Button gute bis sehr gute Bewertungen, für alle Merkmale ***vergleichbar gut wie M***, auch für die Ermüdung der Augen mit 4,8 ± 1,3 (Mauseingabe 5,0 ± 1,6).

Als ***bevorzugte Interaktionstechnik*** nennen 8 VP ***MAGIC Button***, 3 VP MAGIC konservativ und 1 VP die M.

MAGIC Button zeigt sich also als Alternative zur Mauseingabe mit vergleichbarer Qualität. In der Praxis könnten beide im Wechsel genutzt werden, um die manuelle Belastung etwas zu verteilen: Während MAGIC Button weniger Armbewegungen erfordert, belastet die Mauseingabe die Finger weniger.

Blick+EEG versucht anhand von Gehirnaktivität, Benutzerentscheidungen zu erkennen und für Selektionsoperationen zu nutzen. Im Gegensatz zu den betrachteten expliziten Interaktionstechniken, bei denen der Benutzer aktiv und bewusst interagiert, ist Blick+EEG eine implizite, passive Eingabemethode, die keinerlei manuelle oder bewusste Aktion des Benutzers erfordert. Blick+EEG hat dadurch das Potenzial, Objekte auch in Fällen zu markieren, wo der Benutzer zu spät reagiert, um bewusst selektieren zu können.

Der Ansatz für die Nutzung bei Selektionsoperationen ist, den Blick zum Zeigen zu verwenden und das EEG zur Selektionsauslösung. Im Gegensatz zu den expliziten Interaktionstechniken übernimmt das EEG zusätzlich den kognitiven Teil des Entscheidungsprozesses, ob ein Objekt Zielobjekt ist oder nicht. Bei den expliziten Interaktionstechniken hingegen nutzt der Mensch seine Kognition bewusst für die Entscheidung, auf ein Zielobjekt zu zeigen und dabei die Selektion auszulösen.

Blick+EEG für Selektionsoperationen zu nutzen, war zum Zeitpunkt der Untersuchungen Neuland und existiert auch heute noch nicht als Eingabemethode. Die zwei Nutzerstudien der vorliegenden Arbeit (N=21; 2,38° bis 3,16°; 0°/s, 1,66°/s bis 3,37°/s) dienen also als erster Schritt, um die prinzipielle Machbarkeit von Blick+EEG als Selektionsalternative festzustellen. Dazu werden Blick- und EEG-Daten während abstrakter Testaufgaben aufgezeichnet. Eine Offline-Analyse bestimmt dann, mit welcher Qualität die räumlich-zeitliche Lokalisation von Ereignissen (manifestiert durch Objekte mit Zielobjektverhalten) gelingt.

Im Rahmen der vorliegenden Arbeit wird die Analyse der *räumlichen Lokalisation* anhand der Blickdaten durchgeführt; die zeitliche EEG-Lokalisation auf Basis des ereigniskorrelierten Potenzials P300 wird von EEG-Experten der Universität Bremen durchgeführt. Aufgrund des Pilotcharakters der Untersuchungen werden neben bewegten Objekten auch statische Objekte betrachtet.

Die ***Selektionsfehlerquote*** beträgt für statische Objekte je nach Komplexität der Erkennung eines Objektes als Zielobjekt auf Basis der P300 zwischen 21,5% und 40%, für bewegte Objekte zwischen 33,9% und 47,6%. Dies ist erheblich schlechter als die Selektionsfehlerquoten mit BT oder M bei vergleichbaren Geschwindigkeiten; allerdings waren in deren Nutzerstudien die selektierbaren Objektgrößen zum Teil etwas größer.

Die ***Selektionszeit*** liegt für statische und bewegte Objekte bei ca. 500 ms. Dieser Wert entspricht dem von BT und BFT für vergleichbare Versuchsaufgaben. Mit Blick+EEG wird der Benutzer jedoch von jeglicher motorischer Aktion der Extremitäten sowie bewusster kognitiver Aktion entlastet.

Diese ersten Ergebnisse für Blick+EEG sind vielversprechend. Um eine Alternative zur Mauseingabe zu sein, muss die Selektionsfehlerquote deutlich verbessert werden. Für eine Nutzung in der Praxis muss zudem robuste EEG-Erfassung auch mit gebrauchstauglicher Hardware möglich sein, die schnellere Rüstzeiten und besseren Tragekomfort bietet als die in unseren Untersuchungen genutzte EEG-Kappe.

7.2 Beitrag 2: Blickbasierte Interaktion bei automatischen Bildanalyseverfahren

Die zentrale Rolle bei der Bildfolgenanalyse hat der menschliche Experte. Denn da der Mensch zusätzlich zu seinen visuellen Fähigkeiten über Erfahrungswissen bezüglich der Analyseaufgabe verfügt, trifft stets er die finale Entscheidung bezüglich der Relevanz von Objekten oder Ereignissen. Automatische Bildanalyseverfahren, wie das ABUL-System sie bietet, haben die

Rolle einer Assistenz und sind dank ihrer heutigen Zuverlässigkeit eine wertvolle Unterstützung [Hof13]. Automatische Bewegungsdetektion kann unterstützen, bewegte Objekte zu finden, automatisches Einzelobjekttracking, ein bewegtes Objekt nicht aus den Augen zu verlieren. Sie dienen dadurch der Aufmerksamkeitssteuerung des Benutzers und reduzieren dessen perzeptive und kognitive Belastung.

Da automatische Verfahrensergebnisse jedoch zusätzlichen visuellen Input produzieren, wird die visuelle Perzeption des Benutzers zusätzlich belastet. Außerdem entsteht bei der Nutzung automatischer Verfahren zusätzlicher Interaktionsaufwand.

Zielsetzung in Beitrag 2 ist daher, automatische Bildanalyseverfahren in Kombination mit blickbasierter Interaktion zu untersuchen, um festzustellen, ob dies zu einer insgesamt leistungsfähigeren und belastungsärmeren Benutzungsschnittstelle führt. Zu diesem Zweck wird die leistungsfähigste blickbasierte Interaktionstechnik aus Beitrag 1, Blick+Taste, im Vergleich zur Maus-eingabe als Stand der Technik evaluiert.

Für ein automatisches Verfahren zur **Bewegungsdetektion** werden Leistungsfähigkeit und Belastung des Benutzers für die Aufgabe der Bewegtobjektmarkierung *mit* und *ohne* automatische Verfahrensergebnisse untersucht. Die Bedingung *mit* automatischen Verfahrensergebnissen (Independent Motion Detection, IMD) beschreibt einen Fall von Informationsfusion aus automatischer Detektionsleistung und menschlicher Erkennungsleistung. Evaluiert werden ***M, BT*** und ***BT+IMD***.

In einer Nutzerstudie (N=10; 0,09° [3,04°] bis 2,26 x 0,52° [5,23°]; 2,06°/s bis 13,05°/s) erzielt ***BT+IMD*** in Abhängigkeit von Objektgröße und Objektgeschwindigkeit eine ***Selektionsfehlerquote*** zwischen 9% bis 42,9%. BT+IMD ist damit ***signifikant besser als BT und M***, deren Ergebnisse zwischen 19,6% und 64,3% liegen.

Die ***subjektiv empfundene Belastung*** wird mit dem NASA-TLX-Fragebogen für ***BT und BT+IMD deutlich geringer als*** für ***M*** bewertet.

Die Bewertung der ***Zufriedenstellung*** ist ebenfalls für ***BT und BT+IMD*** ähnlich gut; sie ist zudem für einige Merkmale erheblich besser als für M. BT und BT+IMD erzielen gute bis sehr gute Bewertungen für alle Merkmale, mit Ausnahme der Ermüdung der Augen mit einer mittelmäßigen Bewertung. Die Ergebnisse sind für zwei Drittel der erfragten Merkmale ***signifikant besser als M***.

Insgesamt zeigt sich die Interaktionstechnik ***BT+IMD***, die Informationsfusion zwischen Mensch und automatischem Verfahren nutzt, den anderen ***bzgl. der Effektivität klar überlegen***. Zudem erzielte sie zusammen mit BT die ***beste Bewertung bzgl. subjektiv empfundener Belastung und Zufriedenstellung***.

Für ein automatisches Verfahren zum **Einzelobjekttracking** wird die Leistungsfähigkeit des Benutzers für die Interaktionsaufgabe der Tracker-Initialisierung für ein bewegtes Objekt untersucht. Der Benutzer muss dabei durch Bewegtobjektselektion dem System die Position des Zielobjekts genügend präzise mitteilen, sodass das automatische Einzelobjekttracking-Verfahren in der Lage ist, das Zielobjekt robust zu tracken. Falls die Präzision nicht ausreicht, gelingt die Initialisierung entweder nicht oder das Verfahren trackt das Objekt nur kurz und bricht dann ab. In beiden Fällen muss dann erneut aufgesetzt werden, was den Interaktionsaufwand erhöht.

In einer Pilotstudie (N=6; 0,76° [1,37° bzw. 2,75°]; 3,54 ± 1,02°/s) werden ***M, BT*** und ***BT+BlickVis*** evaluiert. Letztere visualisiert permanent die Blickposition als halbtransparenten Ring von 2° Größe. Die Tracker-Initialisierung wird mit zwei verschiedenen Bedingungen für den selektionssensitiven Bereich rund um die Selektionsposition evaluiert (1,37° restriktiver; 2,75° großzügiger).

BT erzielt sowohl die ***beste Selektionsfehlerquote*** (5%) als auch die ***kürzeste Selektionszeit*** und wird von 4 der 6 VP als ***präferierte Technik*** genannt. M erzielt für den restriktiveren selektionssensitiven Bereich vergleichbar gute Ergebnisse, ist für den großzügigeren jedoch etwas schlechter als BT; 2 VP nennen M als präferierte Interaktionstechnik. BT+BlickVis erzielte für beide Bedingungen die schlechtesten Ergebnisse.

In einer ***Expertenstudie*** (N=18; 0,76° [2,75°]; 3,54 ± 1,02°/s) mit Videoanalyseexperten der Bundeswehr werden ***M*** und ***BT*** für den großzügigeren selektionssensitiven Bereich evaluiert. Die ***Selektionsfehlerquote*** ist mit ca. 3% für ***M und BT gleich gut***. Die ***Selektionszeit*** ist ***tendenziell kürzer mit BT***, ***signifikant*** jedoch nur für die Hälfte der Tracker-Initialisierungen ***in herausfordernden Selektionssituationen*** mit starken Szenenschwenks oder Objektverdeckungen. Als ***präferierte Interaktionstechnik*** nennen 10 VP ***BT***, 6 VP M und 2 VP möchten zwischen beiden wechseln können.

Die Ergebnisse von **Beitrag 2** zeigen, dass ***BT geeignet*** ist ***für die Interaktion mit automatischen Bildanalyseverfahren***.

Das Ergebnis aus **Beitrag 1** für ***BT als leistungsfähige, belastungsarme Alternative zur Mauseingabe*** wird bestätigt: Bewegtobjektselektion ist mit BT bei gleicher Effektivität (Selektionsquote) mit kürzerer Selektionszeit möglich.

Auf die Frage nach der ***präferierten Interaktionstechnik*** nennen von den insgesamt N=81 VP der Nutzerstudien aus Beitrag 1 und Beitrag 2 ***73% BT*** und 18% M, die anderen 9% möchten zwischen beiden wechseln. Von den insgesamt N=44 befragten ***Videoanalyseexperten*** nennen ***64% BT***, 20% M, die anderen 16% möchten zwischen beiden wechseln.

7.3 Beitrag 3: Blickbasierte Klassifikation der Benutzertätigkeit bei der Bildfolgenanalyse

Um im ABUL-System eine Systemfunktion, z. B. ein automatisches Bildanalyseverfahren zu nutzen, muss der Benutzer es durch Selektion der entsprechenden Schaltfläche aktivieren. Dazu muss er seine Aufmerksamkeit zumindest kurz von der zu analysierenden Bildfolge abwenden. Je nach Auswahl bewirkt danach die Selektion eines Objektes in der Bildfolge eine andere Systemreaktion, bspw. wird das Objekt markiert oder es wird ein Tracking-Verfahren initialisiert.

Die explizite Benutzeraktion zur Selektion der Schaltfläche wäre unnötig, wenn man die Intention des Benutzers schätzen könnte anhand der Tätigkeit, die der Benutzer aktuell durchführt (z. B. ein Objekt suchen, um es zu markieren, oder ein Objekt verfolgen, um einen Einzelobjekttracker aufzusetzen). Gelänge die Klassifikation der Benutzertätigkeit, könnte die Systemreaktion automatisch erfolgen.

Die Klassifikation komplexer Tätigkeiten auf der Basis des Blickverhaltens ist herausfordernd und wird heute vor allem in der Kognitionspsychologie erforscht. Dort versucht man, aus dem Blickverhalten beim Betrachten *statischer Szenen* verschiedene Aufgabenstellungen bzw. Tätigkeiten wie bspw. „Suchen" oder „Memorieren" zu klassifizieren.

Zielsetzung in Beitrag 3 ist ein erster Schritt, für ***dynamische Szenen***, wie sie bei der Bildfolgenanalyse vorkommen, die ***Benutzertätigkeit auf Basis des Blickverhaltens zu klassifizieren***.

Dafür werden zunächst vier typische Tätigkeiten der Bildfolgenanalyse – ***Erkunden, Beobachten, Suchen, Verfolgen*** – identifiziert und voneinander abgegrenzt. Für jede dieser Tätigkeiten wird dann eine Versuchsaufgabe gestaltet. Die dabei aufgezeichneten Blickbewegungen der Versuchspersonen werden dann algorithmisch zu einer Menge Fixationen aggregiert, aus denen wiederum Fixations- und Sakkadenparameter extrahiert werden. Verwendete Merkmale, die als Merkmalsvektor für die Klassifikation der Tätigkeiten dienen, sind: *Fixationen pro Sekunde, Fixationsdauer, Fixationsdurchmesser, Sakkadenamplitude, Sakkadengeschwindigkeit, Fixationswinkel.*

Verwendete Klassifikatoren sind Random Forest, LDA und QDA. Die Klassifikation erfolgt auf Basis unterschiedlich langer Segmente der Blickdatenprotokolle, zum einen auf der Gesamtlänge der Videosequenz (30 s bzw. 4 min), zum anderen auf kurzen Segmenten von 3 s bzw. 5 s Dauer.

Die Klassifikation für ***vier Tätigkeiten*** gelingt bei ***Segmentlänge von 30 s*** mit einer ***Korrektklassifikationsrate*** von ***59,5% (95% KI=49,8-68,0%, Zufallswahrscheinlichkeit 25%) mit LDA***. Konfusion tritt vor allem zwischen den Klassen *Erkunden* und *Beobachten* sowie zwischen *Suchen* und *Verfolgen* auf; möglicherweise ist dies zumindest teilweise durch die Charakteristika

der Videosequenz sowie die Versuchsaufgabenstellungen erklärbar. Die beste Korrektklassifikationsrate für die ***Segmentlänge von 5 s*** erzielt ***LDA mit 45.3% (95% KI=41.6-49.0%)***, für die ***Segmentlänge von 3 s*** mit ***38.0% (95% KI=35.2-40.8%)***.

Die Klassifikation der ***drei Tätigkeiten*** *Beobachten, Suchen, Verfolgen* gelingt bei ***Segmentlänge von 4 min*** mit einer Korrektklassifikationsrate von ***83,7% (95% KI=74,0-90,4%, Zufallswahrscheinlichkeit 33%)*** mit Random Forest. Die beste Korrektklassifikationsrate für die ***Segmentlänge von 5 s*** erzielt ***LDA mit 52,3% (95% KI=50,8-53,8%)***, für die ***Segmentlänge von 3 s*** mit ***47,7% (95% KI=46,5-48,9%)***.

In allen Fällen gelingt die Klassifikation deutlich über Zufallswahrscheinlichkeit, sie ist jedoch nicht gut genug, um darauf basierend Systemeingaben zu tätigen.

7.4 Ausblick

Die Ergebnisse der vorliegenden Arbeit geben zahlreiche Anknüpfungspunkte für weitere Forschungsarbeiten.

Zum einen wäre für alle betrachteten blickbasierten Interaktionstechniken eine Evaluation mit mehr Versuchspersonen wünschenswert, um mögliche Schwachstellen zu identifizieren und die Ergebnisse auf eine breitere Datenbasis zu stellen. Dies gilt insbesondere für Blick+Fußtaste und MAGIC Button, die hier vielversprechende erste Ergebnisse lieferten, für die aber insbesondere Ermüdungseffekte von Fuß bzw. Fingern bei Nutzung über einen längeren Zeitraum untersucht werden müssten. Relevant wäre auch, den Aspekt der Ermüdung der Augen bei blickbasierter Interaktion mit bewegten Objekten in dynamischen Szenen vertieft zu untersuchen [Hir20].

Interessant wäre zudem, die Bewegtobjektselektion für kleinere Objektgrößen zu ermöglichen, die unterhalb des Auflösungsvermögens von Eyetracking liegen. Ein alternativer Ansatz für die Bewegtobjektselektion ist die Interaktionstechnik *Pursuits* (vgl. Abschnitt 2.4.3.1), wo Blickposition und Objektposition über ein Zeitfenster korreliert werden. Dabei wird ein Objekt nur aufgrund seines Bewegungsmusters als Selektionsobjekt identifiziert. Die notwendigen Trajektorien der Objektbewegung werden bspw. dann vom System bereitgestellt, wenn ein Multiobjekttracking-Verfahren die Bildfolge analysiert [Som21]. Neuere Arbeiten zeigen, dass Pursuits+linker Mausklick akzeptable Selektionsfehlerquoten und Selektionszeiten liefern [Zha20]. Statt einer Korrelation scheint auch die mittlere euklidische Distanz zwischen Blickposition und Objektposition ein vielversprechender Ansatz zu sein [Hil19].

Um blickbasierter Interaktion zum Durchbruch zu verhelfen, sind kostengünstige Eyetracker erforderlich. Diese existieren inzwischen für die Nutzung als Peripheriegerät für Computerspiele, bspw. der Tobii Eye Tracker 4C bzw. der neuere Tobii Eye Tracker 5. Ob damit vergleichbare Ergebnisse erzielt werden können, muss erprobt werden. Im Gegensatz zum hochpreisigen Eyetracker aus der vorliegenden Arbeit liegen für diese kostengünstigen Geräte keine technischen Spezifikationen der räumlichen Genauigkeit vor. Erste Arbeiten zeigen vielversprechende Ergebnisse für den Tobii Eye Tracker 4C [Zha20, Hil19].

Die Verbesserung von Blick+EEG mit dem (Fern-)Ziel einer gebrauchstauglichen Benutzungsschnittstelle betrachten u.a. Putze u. a. [Put16] und Salous u. a. [Sal18] für die Robustifizierung der P300-Erkennung bzgl. der Klassifikation von „Zielobjekt“ versus „Kein Zielobjekt“. Brouwer u. a. [Bro17] nennen als vielversprechend eine Ergänzung der EEG-Erkennung durch Augenkorrelate für Zielobjektdetektion wie die Fixationsdauer sowie Pupillenweite.

Ein weiteres reizvolles Forschungsfeld ist die Theoriebildung zu Bewegtobjektselektion sowie zu blickbasierter Interaktion. Die Frage, ob blickbasierte Interaktion dem Modell von Fitts folgt, ist umstritten [Sib00, Zha10] und rückte zuletzt wieder mehr in den Fokus der Forschungsgemeinde [Sch19, Zha21]. Ähnliches gilt für die Bewegtobjektselektion [Hua18, Par20], von der Park

u. a. [Par20, S.3] schreiben, dass ihre Erforschung in der MMI gerade erst begonnen habe. Vermutlich gilt dies noch mehr für die Kombination aus beiden, die blickbasierte Bewegtobjektselektion.

Der Aspekt der Auswirkung der Informationsfusion als Kombination von Analyseergebnissen des Menschen und automatischer Verfahren ist auch für andere automatische Bildanalyseverfahren relevant, bspw. für die automatische Änderungsdetektion [Hil18a], aber auch für das komplexe Multiobjekttracking [Som21].

Für eine mögliche Verbesserung der Klassifikation von Tätigkeiten bei der Bildfolgenanalyse auf Basis von Blickbewegungen gibt es mehrere Aspekte. Zum einen könnte der Merkmalsvektors um Merkmale erweitert werden, die die zeitliche Struktur des Blickpfads besser erfassen [Kan15, Le 17], zum anderen könnten andere Klassifikationsverfahren untersucht werden. Eine weitere Möglichkeit wäre, neben dem Blickverhalten auch Bildmerkmale für die Klassifikation zu nutzen [Bor21, Boi16].

Literatur

[AB05] AB, Tobii Technology: Tobii 50 Series. Product Description. 2005 (siehe S. 117, 119).

[AB08] AB, Tobii Technology: Tobii X60/X120 Eye Trackers. User Manual. 2008 (siehe S. 119).

[AB10] AB, Tobii Technology: Tobii T/X series Eye Trackers. Product Description. 2010 (siehe S. 120–122).

[Afk14] Afkari, Hoorieh; Eivazi, Shahram; Bednarik, Roman and Mäkelä, Susanne: „The potentials for hands-free interaction in micro-neurosurgery". In: *Proceedings of the 8th Nordic Conference on Human-Computer Interaction: Fun, Fast, Foundational.* 2014, S. 401–410 (siehe S. 70).

[Afk18] Afkari, Hoorieh: „Interaction in the micro-surgery operating room: the potentials for gaze-based interaction with the surgical microscope". Diss. Itä-Suomen yliopisto, 2018 (siehe S. 70).

[Alo13] Alonso, Roland; Causse, Mickaël; Vachon, François; Parise, Robert; Dehais, Frédéric and Terrier, Patrice: „Evaluation of head-free eye tracking as an input device for air traffic control". In: *Ergonomics* 56.2 (2013), S. 246–255 (siehe S. 178).

[Alp14] Alpaydin, Ethem: Introduction to machine learning. MIT press, 2014 (siehe S. 114, 115).

[Aru02] Arulampalam, M Sanjeev; Maskell, Simon; Gordon, Neil and Clapp, Tim: „A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking". In: *IEEE Transactions on signal processing* 50.2 (2002), S. 174–188 (siehe S. 60).

[Bad11] Bader, Thomas and Beyerer, Jürgen: „Influence of user's mental model on natural gaze behavior during human-computer interaction". In: *2nd Workshop on Eye Gaze in Intelligent Human Machine Interaction*. 2011, S. 25–32 (siehe S. 67).

[Bad14] Bader, Thomas: Multimodale Interaktion in Multi-Display-Umgebungen. Bd. 9. KIT Scientific Publishing, 2014 (siehe S. 67).

[Bau03] Baudisch, Patrick; Cutrell, Edward; Czerwinski, Mary; Robbins, Daniel C; Tandler, Peter; Bederson, Benjamin B and Zierlinger, Alex: „Drag-and-Pop and Drag-and-Pick: Techniques for Accessing Remote Screen Content on Touch-and Pen-Operated Systems." In: *Interact*. Bd. 3. 2003, S. 57–64 (siehe S. 51).

[Bec00] Bechara, Antoine; Damasio, Hanna and Damasio, Antonio R: „Emotion, decision making and the orbitofrontal cortex". In: *Cerebral cortex* 10.3 (2000), S. 295–307 (siehe S. 86).

[Bed09] Bednarik, Roman; Gowases, Tersia and Tukiainen, Markku: „Gaze interaction enhances problem solving: Effects of dwell-time based, gaze-augmented, and mouse interaction on problem-solving strategies and user experience". In: *Journal of Eye Movement Research* 3.1 (2009) (siehe S. 69, 102, 130).

[Bez05] Bezerianos, Anastasia and Balakrishnan, Ravin: „The vacuum: facilitating the manipulation of distant objects". In: *Proceedings of the SIGCHI conference on Human factors in computing systems*. 2005, S. 361–370 (siehe S. 51).

[Bie10] Bieg, Hans-Joachim; Chuang, Lewis L; Fleming, Roland W; Reiterer, Harald and Bülthoff, Heinrich H: „Eye and pointer coordination in search and selection tasks". In: *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications*. 2010, S. 89–92 (siehe S. 175).

[Boi16] Boisvert, Jonathan FG and Bruce, Neil DB: „Predicting task from eye movements: On the importance of spatial distribution, dynamics, and image features". In: *Neurocomputing* 207 (2016), S. 653–668 (siehe S. 88, 90, 113, 114, 308).

[Bol82] Bolt, Richard A: „Eyes at the interface“. In: *Proceedings of the 1982 conference on Human factors in computing systems*. 1982, S. 360–362 (siehe S. 54).

[Bor09] Bortz, Jürgen and Döring, Nicola: Forschungsmethoden und Evaluation (4. Auflage). Springer, 2009 (siehe S. 105).

[Bor14] Borji, Ali and Itti, Laurent: „Defending Yarbus: Eye movements reveal observers' task“. In: *Journal of vision* 14.3 (2014), S. 29–29 (siehe S. 90, 91, 113).

[Bor21] Borji, Ali: „Saliency prediction in the deep learning era: Successes and limitations“. In: *IEEE transactions on pattern analysis and machine intelligence* 43.2 (2021), S. 679–700 (siehe S. 90, 308).

[Bow04] Bowyer, Richard: Campaign Dictionary of Military Terms. Macmillan, 2004 (siehe S. 111).

[Bra19] Brandes, Ralf; Lang, Florian and Schmidt, Robert F: Physiologie des Menschen. Springer, 2019 (siehe S. 20, 24, 28, 29, 31–33).

[Bra21] BrainProducts. Letzter Zugriff am 28-07-2021. 2021. url: https://www.brainproducts.com/productdetails?id=42 (siehe S. 86, 126).

[Bre01] Breiman, Leo: „Random forests“. In: *Machine learning* 45.1 (2001), S. 5–32 (siehe S. 114).

[Bro17] Brouwer, Anne-Marie; Hogervorst, Maarten A; Oudejans, Bob; Ries, Anthony J and Touryan, Jonathan: „EEG and eye tracking signatures of target encoding during structured visual search“. In: *Frontiers in human neuroscience* 11 (2017), S. 264 (siehe S. 307).

[Bro21] Brouwer, Anne-Marie: „Challenges and opportunities in consumer neuroergonomics“. In: *Frontiers in Neuroergonomics* 2 (2021), S. 3 (siehe S. 86).

[Brü12] Brüstle, Stefan and Heinze, Norbert: „Archiving image sequences considering associated geographical and nongeographical attributes“. In: *Sensors, and Command, Control, Communications, and Intelligence (C3I) Technologies for Homeland Security and Homeland Defense XI*. Bd. 8359. International Society for Optics und Photonics. 2012, 83590Y (siehe S. 228).

[Buc14] Buchner, Helmut; Classen, Joseph; Curio, G; Ferbert, A; Haupt, WF; Hecht, M and Wessel, K: Praxisbuch Evozierte Potenziale: Grundlagen, Befundung, Beurteilung und differenzialdiagnostische Abgrenzung. Thieme Stuttgart, 2014 (siehe S. 85, 86).

[Bus35] Buswell, Guy Thomas: How people look at pictures: a study of the psychology and perception in art. Univ. Chicago Press, 1935 (siehe S. 88).

[Car88] Carpenter, Roger H.S.: Movements of the Eyes. 2. Aufl. Pion Limited, 1988 (siehe S. 32).

[Cas08] Castellina, Emiliano and Corno, Fulvio: „Multimodal gaze interaction in 3D virtual environments“. In: *Cogain* 8.2008 (2008), S. 33–37 (siehe S. 65).

[Cas09] Castelhano, Monica S; Mack, Michael L and Henderson, John M: „Viewing task influences eye movement control during active scene perception“. In: *Journal of vision* 9.3 (2009), S. 6–6 (siehe S. 89).

[Cle94] Cleveland, Nancy R: „Eyegaze human-computer interface for people with disabilities“. In: *First Automation Technology and Human Performance Conference*. 1994, S. 7–8 (siehe S. 75).

[Das01] Dasarathy, Belur V.: „Information Fusion - What, Where, Why, When, and How?“ In: *Information Fusion* 2.2 (2001), S. 75–76 (siehe S. 108).

[DeA09] DeAngelus, Marianne and Pelz, Jeff B: „Top-down control of eye movements: Yarbus revisited“. In: *Visual Cognition* 17.6-7 (2009), S. 790–811 (siehe S. 89, 113, 288).

[DIN00] DIN: Ergonomische Anforderungen für Bürotätigkeiten mit Bildschirmgeräten - Teil 9: Anforderungen an Eingabemittel - ausgenommen Tastaturen. Norm. 2000 (siehe S. 93, 232).

[DIN14] DIN: Ergonomie der Mensch-System-Interaktion - Teil 411: Bewertungsverfahren für die Gestaltung von physikalischen Eingabegeräten. 2014 (siehe S. 68, 93, 96, 104, 159, 183, 197, 241, 243, 258, 259, 296).

[DIN18a] DIN: Ergonomische Anforderungen für Bürotätigkeiten mit Bildschirmgeräten - Teil 11: Anforderungen an die Gebrauchstauglichkeit - Leitsätze. Norm. 2018 (siehe S. 93).

[DIN18b] DIN: Ergonomie der Mensch-System-Interaktion - Teil 11: Gebrauchstauglichkeit: Begriffe und Konzepte. Norm. 2018 (siehe S. 92, 93, 96).

[Dja15] Djamasbi, Soussan and Mortazavi, Siavash: „Generation Y, baby boomers, and gaze interaction experience in gaming". In: *2015 48th Hawaii International Conference on System Sciences*. IEEE. 2015, S. 482–490 (siehe S. 65, 69).

[Dor12] Dorr, Michael; Vig, Eleonora and Barth, Erhardt: „Eye movement prediction and variability on natural video data sets". In: *Visual cognition* 20.4-5 (2012), S. 495–514 (siehe S. 90).

[Dor20] Dorsch, F and Huber, Verlag Hans: Dorsch—Lexikon der Psychologie (M. A. Wirtz, Hrsg.) 19. Aufl. Hogrefe, Bern, 2020 (siehe S. 139).

[Dre09] Drewes, Heiko and Schmidt, Albrecht: „The MAGIC touch: Combining MAGIC-pointing with a touch-sensitive mouse". In: *IFIP Conference on Human-Computer Interaction*. Springer. 2009, S. 415–428 (siehe S. 145, 157).

[Ebe73] Ebeling, F; Johnson, R and Goldhor, R: Infrared light beam xy position encoder for display devices. US Patent 3,775,560. Nov. 1973 (siehe S. 48).

[Eng67] English, William K; Engelbart, Douglas C and Berman, Melvyn L: „Display-selection techniques for text manipulation". In: *IEEE Transactions on Human Factors in Electronics* 1 (1967), S. 5–15 (siehe S. 49).

[Fit64] Fitts, Paul M and Peterson, James R: „Information capacity of discrete motor responses." In: *Journal of experimental psychology* 67.2 (1964), S. 103 (siehe S. 50).

[Gäe80] Gäertner, K-P and Holzhausen, K-P: „Controlling air traffic with touch sensitive screen". In: *Applied ergonomics* 11.1 (1980), S. 17–22 (siehe S. 48).

[Gei06] Geisler, Jürgen: Leistung des Menschen am Bildschirmarbeitsplatz. Das Kurzzeitgedächtnis als Schranke menschlicher Belastbarkeit in der Konkurrenz von Arbeitsaufgabe und Systembedienung. 2006 (siehe S. 111).

[Ger06] Gerson, Adam D; Parra, Lucas C and Sajda, Paul: „Cortically coupled computer vision for rapid image search". In: *IEEE Transactions on neural systems and rehabilitation engineering* 14.2 (2006), S. 174–179 (siehe S. 86).

[Gez97] Gezeck, Stefan; Fischer, Burkhart and Timmer, Jens: „Saccadic reaction times: a statistical analysis of multimodal distributions". In: *Vision research* 37.15 (1997), S. 2119–2131 (siehe S. 203).

[Gil12] Gill, Dennis: „MAGIC pointing für die Selektion sich bewegender Objekte". Bachelorarbeit. Karlsruher Institut für Technologie (KIT), 2012 (siehe S. 150).

[Göb13] Göbel, Fabian; Klamka, Konstantin; Siegel, Andreas; Vogt, Stefan; Stellmach, Sophie and Dachselt, Raimund: „Gaze-supported foot interaction in zoomable information spaces". In: *CHI'13 Extended Abstracts on Human Factors in Computing Systems*. 2013, S. 3059–3062 (siehe S. 69, 70).

[Gol15] Goldstein, Eugen Bruce: Wahrnehmungspsychologie: Der Grundkurs. Springer, 2015 (siehe S. 20–22, 24, 34, 36, 37).

[Gre12] Greene, Michelle R; Liu, Tommy and Wolfe, Jeremy M: „Reconsidering Yarbus: A failure to predict observers' task from eye movement patterns". In: *Vision research* 62 (2012), S. 1–8 (siehe S. 90, 91, 113, 114).

[Gro05] Grossman, Tovi and Balakrishnan, Ravin: „The bubble cursor: enhancing target acquisition by dynamic resizing of the cursor's activation area". In: *Proceedings of the SIGCHI conference on Human factors in computing systems*. 2005, S. 281–290 (siehe S. 50).

[Gut15] Guthmann, Micha-Jamie: „Der Einfluss von Lerneffekten auf die Leistungsfähigkeit des Systemnutzers bei blickbasierter Bewegtobjektselektion". Masterarbeit. Karlsruher Institut für Technologie (KIT), 2015 (siehe S. 174, 182).

[Har04] Hartley, Andrew and Zisserman, Richard: Multiple view geometry in computer vision. Cambridge University Press, 2004 (siehe S. 252).

[Har06] Hart, Sandra G: „NASA-task load index (NASA-TLX); 20 years later". In: *Proceedings of the human factors and ergonomics society annual meeting*. Bd. 50. 9. Sage publications Sage CA: Los Angeles, CA. 2006, S. 904–908 (siehe S. 93, 259).

[Har88] Hart, Sandra G and Staveland, Lowell E: „Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research". In: *Advances in psychology*. Bd. 52. Elsevier, 1988, S. 139–183 (siehe S. 93, 259).

[Has11] Hasan, Khalad; Grossman, Tovi and Irani, Pourang: „Comet and target ghost: techniques for selecting moving targets". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 2011, S. 839–848 (siehe S. 1, 3, 53, 54, 130, 162, 295).

[Hei08] Heinze, Norbert; Esswein, Martin; Krüger, Wolfgang and Saur, Günter: „Automatic image exploitation system for small UAVs". In: *Airborne intelligence, surveillance, reconnaissance*

(ISR) systems and applications V. Bd. 6946. International Society for Optics und Photonics. 2008, 69460G (siehe S. 4).

[Hei10] Heinze, Norbert; Esswein, Martin; Krüger, Wolfgang and Saur, Günter: „Image exploitation algorithms for reconnaissance and surveillance with UAV". In: *Airborne Intelligence, Surveillance, Reconnaissance (ISR) Systems and Applications VII*. Bd. 7668. International Society for Optics und Photonics. 2010, 76680U (siehe S. 4, 208).

[Hen13] Henderson, John M; Shinkareva, Svetlana V; Wang, Jing; Luke, Steven G and Olejarczyk, Jenn: „Predicting cognitive state from eye movements". In: *PloS one* 8.5 (2013), e64937 (siehe S. 88, 90, 91).

[Hil08] Hillaire, Sébastien; Lécuyer, Anatole; Cozot, Rémi and Casiez, Géry: „Using an eye-tracking system to improve camera motions and depth-of-field blur effects in virtual environments". In: *2008 IEEE virtual reality conference*. IEEE. 2008, S. 47–50 (siehe S. 77).

[Hil12] Hild, Jutta; Peinsipp-Byma, Elisabeth and Klaus, Edmund: „Improving usability for video analysis using gaze-based interaction". In: *Full Motion Video (FMV) Workflows and Technologies for Intelligence, Surveillance, and Reconnaissance (ISR) and Situational Awareness*. Bd. 8386. International Society for Optics und Photonics. 2012, S. 838605 (siehe S. 89, 110).

[Hil13a] Hild, J; Müller, E; Klaus, E; Peinsipp-Byma, E and Beyerer, J: „Evaluating multi-modal eye gaze interaction for moving object selection". In: *Proc. ACHI* (2013), S. 454–459 (siehe S. 132, 218, 219, 224).

[Hil13b] Hild, Jutta; Brüstle, Stefan; Heinze, Norbert and Peinsipp-Byma, Elisabeth: „Gaze interaction in UAS video exploitation". In: *Motion Imagery Technologies, Best Practices, and Workflows for Intelligence, Surveillance, and Reconnaissance (ISR), and Situational Awareness*. Bd. 8740. International Society for Optics und Photonics. 2013, 87400H (siehe S. 228, 241, 258, 282, 283).

[Hil13c] Hild, Jutta; Fischer, Yvonne; Peinsipp-Byma, Elisabeth and Beyerer, Jürgen: „Gaze-based interaction for real-time video surveillance systems“. In: *Proceedings of the 8th Security Research Conference (Future Security)*. Fraunhofer Verlag, 2013, S. 63–72 (siehe S. 16, 137).

[Hil14a] Hild, Jutta; Gill, Dennis and Beyerer, Jürgen: „Comparing mouse and magic pointing for moving target acquisition“. In: *Proceedings of the Symposium on Eye Tracking Research and Applications*. 2014, S. 131–134 (siehe S. 144, 218, 224, 241, 258).

[Hil14b] Hild, Jutta; Putze, Felix; Kaufman, David; Kühnle, Christian; Schultz, Tanja and Beyerer, Jürgen: „Spatio-temporal event selection in basic surveillance tasks using eye tracking and EEG“. In: *Proceedings of the 7th Workshop on Eye Gaze in Intelligent Human Machine Interaction: Eye-Gaze & Multimodality*. 2014, S. 3–8 (siehe S. 208, 215, 216, 218).

[Hil15] Hild, Jutta; Krüger, Wolfgang; Brüstle, Stefan; Trantelle, Patrick; Unmüßig, Gabriel; Heinze, Norbert; Peinsipp-Byma, Elisabeth and Beyerer, Jürgen: „Collaborative real-time motion video analysis by human observer and image exploitation algorithms“. In: *Motion Imagery: Standards, Quality, and Interoperability*. Bd. 9463. International Society for Optics und Photonics. 2015, S. 94630C (siehe S. 265, 282, 283).

[Hil16a] Hild, Jutta; Kühnle, Christian and Beyerer, Jürgen: „Gaze-based moving target acquisition in real-time full motion video“. In: *Proceedings of the Ninth Biennial ACM Symposium on Eye Tracking Research & Applications*. 2016, S. 241–244 (siehe S. 236, 258, 282, 283).

[Hil16b] Hild, Jutta; Petersen, Patrick and Beyerer, Jürgen: „Moving target acquisition by gaze pointing and button press using hand or foot“. In: *Proceedings of the Ninth Biennial ACM Symposium on Eye Tracking Research & Applications*. 2016, S. 257–260 (siehe S. 158, 218).

[Hil17] Hild, Jutta; Krüger, Wolfgang; Brüstle, Stefan; Trantelle, Patrick; Unmüßig, Gabriel; Voit, Michael; Heinze, Norbert; Peinsipp-Byma, Elisabeth and Beyerer, Jürgen: „Pilot study on real-time motion detection in UAS video data by human observer and image exploitation algorithm". In: *Geospatial Informatics, Fusion, and Motion Video Analytics VII*. Bd. 10199. International Society for Optics und Photonics. 2017, S. 1019903 (siehe S. 251, 282, 283).

[Hil18a] Hild, Jutta; Saur, Günter; Petersen, Patrick; Voit, Michael; Peinsipp-Byma, Elisabeth and Beyerer, Jürgen: „Evaluating user interfaces supporting change detection in aerial images and aerial image sequences". In: *International Conference on Human Interface and the Management of Information*. Springer. 2018, S. 383–402 (siehe S. 308).

[Hil18b] Hild, Jutta; Voit, Michael; Kühnle, Christian and Beyerer, Jürgen: „Predicting observer's task from eye movement patterns during motion image analysis". In: *Proceedings of the 2018 ACM symposium on eye tracking research & applications*. 2018, S. 1–5 (siehe S. 284).

[Hil19] Hild, Jutta; Peinsipp-Byma, Elisabeth; Voit, Michael and Beyerer, Jürgen: „Suggesting Gaze-based Selection for Surveillance Applications". In: *2019 16th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*. IEEE. 2019, S. 1–8 (siehe S. 277, 282, 283, 307).

[Hir20] Hirzle, Teresa; Cordts, Maurice; Rukzio, Enrico and Bulling, Andreas: „A Survey of Digital Eye Strain in Gaze-Based Interactive Systems". In: *ACM Symposium on Eye Tracking Research and Applications*. 2020, S. 1–12 (siehe S. 306).

[Hof13] Hoffman, Robert R; Johnson, Matthew; Bradshaw, Jeffrey M and Underbrink, Al: „Trust in automation". In: *IEEE Intelligent Systems* 28.1 (2013), S. 84–88 (siehe S. 13, 302).

[Hof91] Hoffmann, Errol R: „Capture of moving targets: a modification of Fitts' Law". In: *Ergonomics* 34.2 (1991), S. 211–220 (siehe S. 295).

[Hol11] Holmqvist, Kenneth; Nyström, Marcus; Andersson, Richard; Dewhurst, Richard; Jarodzka, Halszka and Van de Weijer, Joost: Eye tracking: A comprehensive guide to methods and measures. OUP Oxford, 2011 (siehe S. 26, 32, 41–46, 54, 61, 122).

[Hua13] Huang, Baihan; Lo, Anthony HP and Shi, Bertram E: „Integrating EEG information improves performance of gaze based cursor control". In: *2013 6th International IEEE/EMBS Conference on Neural Engineering (NER)*. IEEE. 2013, S. 415–418 (siehe S. 87).

[Hua18] Huang, Jin; Tian, Feng; Fan, Xiangmin; Zhang, Xiaolong and Zhai, Shumin: „Understanding the uncertainty in 1D unidirectional moving target selection". In: *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 2018, S. 1–12 (siehe S. 307).

[Hut89] Hutchinson, Thomas E; White, K Preston; Martin, Worthy N; Reichert, Kelly C and Frey, Lisa A: „Human-computer interaction using eye-gaze input". In: *IEEE Transactions on systems, man, and cybernetics* 19.6 (1989), S. 1527–1534 (siehe S. 76).

[Ili09] Ilich, Michael Victor: „Moving target selection in interactive video". Diss. University of British Columbia, 2009 (siehe S. 52).

[IOS21] IOSB, Fraunhofer: Automatisierte Bildauswertung für Unbemannte Luftfahrzeuge (ABUL). Zuletzt abgerufen am 19.07.2021. 2021. url: https://www.iosb.fraunhofer.de/de/projekte-produkte/abul-bildauswertung-unbemannte-luftfahrzeuge.html (siehe S. 4).

[Irw92] Irwin, David E: „Visual memory within and across fixations". In: *Eye movements and visual cognition*. Springer, 1992, S. 146–165 (siehe S. 215).

[Isa18] Isachenko, Andrey V; Zhao, Darisy G; Melnichuk, Eugeny V; Dubynin, Ignat A; Velichkovsky, Boris M and Shishkin, Sergei L: „The pursuing gaze beats mouse in non-pop-out target selection". In: *2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. IEEE. 2018, S. 3518–3523 (siehe S. 83, 84).

[Iso06] Isokoski, Poika and Martin, Benoît: „Eye tracker input in first person shooter games". In: *Proceedings of the 2nd Conference on Communication by Gaze Interaction: Communication by Gaze Interaction-COGAIN 2006: Gazing into the Future*. 2006, S. 78–81 (siehe S. 78).

[Iso07] Isokoski, Poika; Hyrskykari, Aulikki; Kotkaluoto, Sanna and Martin, Benoit: „Gamepad and eye tracker input in FPS games: data for the first 50 minutes". In: *Proc. of the 3rd Conference on Communication by Gaze Interaction (COGAIN 2007)*. 2007, S. 78–81 (siehe S. 78).

[Iso09] Isokoski, Poika; Joos, Markus; Spakov, Oleg and Martin, Benoît: „Gaze controlled games". In: *Universal Access in the Information Society* 8.4 (2009), S. 323–337 (siehe S. 76, 79).

[Ist09] Istance, Howell; Hyrskykari, Aulikki; Vickers, Stephen and Chaves, Thiago: „For your eyes only: Controlling 3d online games by eye-gaze". In: *IFIP Conference on Human-Computer Interaction*. Springer. 2009, S. 314–327 (siehe S. 65).

[Jac90] Jacob, Robert JK: „What you look at is what you get: eye movement-based interaction techniques". In: *Proceedings of the SIGCHI conference on Human factors in computing systems*. 1990, S. 11–18 (siehe S. 54, 55, 63).

[Jac91] Jacob, Robert JK: „The use of eye movements in human-computer interaction techniques: what you look at is what you get". In: *ACM Transactions on Information Systems (TOIS)* 9.2 (1991), S. 152–169 (siehe S. 64, 66).

[Jac93] Jacob, Robert JK: „Eye movement-based human-computer interaction techniques: Toward non-command interfaces“. In: *Advances in human-computer interaction* 4 (1993), S. 151–190 (siehe S. 74).

[Jac95] Jacob, Robert JK: „Eye tracking in advanced interface design“. In: *Virtual environments and advanced interface design* 258 (1995), S. 288 (siehe S. 64, 75).

[Jag80] Jagacinski, Richard J; Repperger, Daniel W; Ward, Sharon L and Moran, Martin S: „A test of Fitts' law with moving targets“. In: *Human Factors* 22.2 (1980), S. 225–233 (siehe S. 51, 295).

[Jar10] Jarodzka, Halszka; Holmqvist, Kenneth and Nyström, Marcus: „A vector-based, multidimensional scanpath similarity measure“. In: *Proceedings of the 2010 symposium on eye-tracking research & applications*. 2010, S. 211–218 (siehe S. 113, 114).

[Jön05] Jönsson, Erika: „If looks could kill–an evaluation of eye tracking in computer games“. In: *Unpublished Master's Thesis, Royal Institute of Technology (KTH), Stockholm, Sweden* (2005) (siehe S. 77, 78).

[Jus76] Just, Marcel Adam and Carpenter, Patricia A: „The role of eye-fixation research in cognitive psychology“. In: *Behavior Research Methods & Instrumentation* 8.2 (1976), S. 139–143 (siehe S. 54, 89).

[Kab95] Kabbash, Paul and Buxton, William AS: „The "prince" technique: Fitts' law and selection using area cursors“. In: *Proceedings of the SIGCHI conference on Human factors in computing systems*. 1995, S. 273–279 (siehe S. 50).

[Kam08] Kammerer, Yvonne; Scheiter, Katharina and Beinhauer, Wolfgang: „Looking my way through the menu: the impact of menu design and multimodal input on gaze-based menu selection“. In: *Proceedings of the 2008 Symposium on Eye Tracking Research & Applications*. 2008, S. 213–220 (siehe S. 55).

[Kan14] Kanan, Christopher; Ray, Nicholas A; Bseiso, Dina NF; Hsiao, Janet H and Cottrell, Garrison W: „Predicting an observer's task using multi-fixation pattern analysis". In: *Proceedings of the symposium on eye tracking research and applications*. 2014, S. 287–290 (siehe S. 90, 91, 113).

[Kan15] Kanan, Christopher; Bseiso, Dina NF; Ray, Nicholas A; Hsiao, Janet H and Cottrell, Garrison W: „Humans have idiosyncratic and task-specific scanpaths for judging faces". In: *Vision research* 108 (2015), S. 67–76 (siehe S. 294, 308).

[Kar15] Kardan, Omid; Berman, Marc G; Yourganov, Grigori; Schmidt, Joseph and Henderson, John M: „Classifying mental states from eye movements during scene viewing." In: *Journal of Experimental Psychology: Human Perception and Performance* 41.6 (2015), S. 1502 (siehe S. 90, 92, 113, 114).

[Kau20] Kaufmann, Herbert and Steffen, Heimo: Strabismus. Georg Thieme Verlag, 2020 (siehe S. 20, 30–34).

[Koh09] Koh, Do Hyong; Munikrishne Gowda, Sandeep A and Komogortsev, Oleg V: „Input evaluation of an eye-gaze-guided interface: kalman filter vs. velocity threshold eye movement identification". In: *Proceedings of the 1st ACM SIGCHI symposium on Engineering interactive computing systems*. 2009, S. 197–202 (siehe S. 61).

[Kom07] Komogortsev, Oleg V and Khan, Javed I: „Kalman filtering in the design of eye-gaze-guided computer interfaces". In: *International Conference on Human-Computer Interaction*. Springer. 2007, S. 679–689 (siehe S. 59, 61).

[Kru08] Krusienski, Dean J; Sellers, Eric W; McFarland, Dennis J; Vaughan, Theresa M and Wolpaw, Jonathan R: „Toward enhanced P300 speller performance". In: *Journal of neuroscience methods* 167.1 (2008), S. 15–21 (siehe S. 86).

[Küh17] Kühnle, Christian: „Klassifikation von Aufgaben aus Blickbewegungsmustern bei der Analyse dynamischer Szenen“. Mastearbeit. Karlsruher Institut für Technologie (KIT), 2017 (siehe S. 284).

[Kum07a] Kumar, Manu; Garfinkel, Tal; Boneh, Dan and Winograd, Terry: „Reducing shoulder-surfing by using gaze-based password entry“. In: *Proceedings of the 3rd symposium on Usable privacy and security*. 2007, S. 13–19 (siehe S. 55).

[Kum07b] Kumar, Manu; Paepcke, Andreas and Winograd, Terry: „Eyepoint: practical pointing and selection using gaze and keyboard“. In: *Proceedings of the SIGCHI conference on Human factors in computing systems*. 2007, S. 421–430 (siehe S. 58, 66).

[Kum07c] Kumar, Manu; Winograd, Terry and Paepcke, Andreas: „Gaze-enhanced scrolling techniques“. In: *CHI'07 Extended Abstracts on Human Factors in Computing Systems*. 2007, S. 2531–2536 (siehe S. 55).

[Kum08] Kumar, Manu; Klingner, Jeff; Puranik, Rohan; Winograd, Terry and Paepcke, Andreas: „Improving the accuracy of gaze input for interaction“. In: *Proceedings of the 2008 symposium on Eye tracking research & applications*. 2008, S. 65–68 (siehe S. 56, 58, 60, 67, 102, 164, 182, 243, 248, 250, 257, 273, 297).

[Lan00] Lankford, Chris: „Effective eye-gaze input into windows“. In: *Proceedings of the 2000 symposium on Eye tracking research & applications*. 2000, S. 23–27 (siehe S. 76).

[Le 17] Le Meur, Olivier; Coutrot, Antoine; Liu, Zhi; Rämä, Pia; Le Roch, Adrien and Helo, Andrea: „Your gaze betrays your age“. In: *2017 25th European Signal Processing Conference (EUSIPCO)*. IEEE. 2017, S. 1892–1896 (siehe S. 113, 294, 308).

[Lei15] Leigh, R John and Zee, David S: The neurology of eye movements. Contemporary Neurology, 2015 (siehe S. 28).

[Ley04] Leyba, J and Malcolm, J: „Eye tracking as an aiming device in a computer game“. In: *Course work (CPSC 412/612 Eye Tracking Methodology and Applications by A. Duchowski), Clemson University* 14 (2004) (siehe S. 77).

[Lip17] Lippert, Herbert; Herbold, Désirée and Lippert-Burmester, Wunna: Anatomie: Text und Atlas; deutsche und lateinische Bezeichnungen. Elsevier, Urban&FischerVerlag, 2017 (siehe S. 20, 24, 25).

[Low74] Lowe, JF: „Computer creates custom control panel“. In: *Design News* 29.22 (1974), S. 54–55 (siehe S. 48).

[Mac12] MacKenzie, I Scott: Human-computer interaction: An empirical research perspective. Newnes, 2012 (siehe S. 1, 48, 49).

[Maj02] Majaranta, Päivi and Räihä, Kari-Jouko: „Twenty years of eye typing: systems and design issues“. In: *Proceedings of the 2002 symposium on Eye tracking research & applications*. 2002, S. 15–22 (siehe S. 55, 64, 102).

[Maj09] Majaranta, Päivi; Ahola, Ulla-Kaija and Špakov, Oleg: „Fast gaze typing with an adjustable dwell time“. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 2009, S. 357–360 (siehe S. 64).

[Mar09] Martinez-Conde, Susana; Macknik, Stephen L; Troncoso, Xoana G and Hubel, David H: „Microsaccades: a neurophysiological analysis“. In: *Trends in neurosciences* 32.9 (2009), S. 463–475 (siehe S. 45).

[McG02] McGuffin, Michael and Balakrishnan, Ravin: „Acquisition of expanding targets“. In: *Proceedings of the SIGCHI conference on Human factors in computing systems*. 2002, S. 57–64 (siehe S. 51).

[McG05] McGuffin, Michael J and Balakrishnan, Ravin: „Fitts' law and expanding targets: Experimental studies and designs for user interfaces“. In: *ACM Transactions on Computer-Human Interaction (TOCHI)* 12.4 (2005), S. 388–422 (siehe S. 51).

[Min04] Miniotas, Darius; Špakov, Oleg and MacKenzie, I Scott: „Eye gaze interaction with expanding targets“. In: *CHI'04 extended abstracts on Human factors in computing systems*. 2004, S. 1255–1258 (siehe S. 130).

[Min06] Miniotas, Darius; Špakov, Oleg; Tugoy, Ivan and MacKenzie, I Scott: „Speech-augmented eye gaze interaction with small closely spaced targets“. In: *Proceedings of the 2006 symposium on Eye tracking research & applications*. 2006, S. 67–72 (siehe S. 70, 130).

[Mon05] Monden, Akito; Matsumoto, Ken-ichi and Yamato, Masatake: „Evaluation of gaze-added target selection methods suitable for general GUIs“. In: *International journal of computer applications in technology* 24.1 (2005), S. 17–24 (siehe S. 69, 102, 130).

[Nac10] Nacke, Lennart E; Stellmach, Sophie; Sasse, Dennis and Lindley, Craig A: „Gameplay experience in a gaze interaction game“. In: *arXiv preprint arXiv:1004.0259* (2010) (siehe S. 77).

[Ols12a] Olsen, Anneli: „The Tobii I-VT fixation filter“. In: *Tobii Technology* 21 (2012) (siehe S. 61).

[Ols12b] Olsen, Anneli and Matos, Ricardo: „Identifying parameter values for an I-VT fixation filter suitable for handling data sampled with various sampling frequencies“. In: *proceedings of the symposium on Eye tracking research and applications*. 2012, S. 317–320 (siehe S. 61).

[Orr68] Orr, NW and Hopkin, VD: The role of the touch display in air traffic control. Techn. Ber. RAF INST OF AVIATION MEDICINE FARNBOROUGH (ENGLAND), 1968 (siehe S. 48).

[Par01] Partala, Timo; Aula, Anne and Surakka, Veikko: „Combined voluntary gaze direction and facial muscle activity as a new pointing technique“. In: *Proceedings of INTERACT 2001, Tokyo, Japan*. IOS Press, 2001, S. 100–107 (siehe S. 71).

[Par20] Park, Eunji and Lee, Byungjoo: „An Intermittent Click Planning Model“. In: *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 2020, S. 1–13 (siehe S. 307, 308).

[Par21] Parisay, Mohsen; Poullis, Charalambos and Kersten-Oertel, Marta: „EyeTAP: Introducing a Multimodal Gaze-based Technique using Voice Inputs with a Comparative Analysis of Selection Techniques“. In: *International Journal of Human-Computer Studies* (2021), S. 102676 (siehe S. 70).

[Pet15] Petersen, Patrick: „Bewegtobjektselektion mittels Blick+ Tastendruck-Interaktionstechniken mit Hand- oder Fußtaste“. Bachelorarbeit. Karlsruher Institut für Technologie (KIT), 2015 (siehe S. 164).

[Phi97] Philipson, Warren R: Manual of photographic interpretation. Asprs Publications, 1997 (siehe S. 1, 2, 20, 35).

[Pos78] Posner, Michael I; Nissen, Mary Jo and Ogden, William C: „Attended and unattended processing modes: The role of set for spatial location“. In: *Modes of perceiving and processing information* 137.158 (1978), S. 2 (siehe S. 36).

[Pub96] Publius Ovidius Naso (* 43 v.u.Z. bis 17 n.u.Z.), dt. Übersetzung von Hermann Breitenbach: Metamorphoses. Verwandlungen. Zweisprachig. Deutscher Taschenbuchverlag, 1996 (siehe S. 63).

[Put13] Putze, Felix; Hild, Jutta; Kärgel, Rainer; Herff, Christian; Redmann, Alexander; Beyerer, Jürgen and Schultz, Tanja: „Locating user attention using eye tracking and EEG for spatio-temporal event selection“. In: *Proceedings of the 2013 international conference on Intelligent user interfaces*. 2013, S. 129–136 (siehe S. 196, 205, 206, 216, 218).

[Put16] Putze, Felix; Popp, Johannes; Hild, Jutta; Beyerer, Jürgen and Schultz, Tanja: „Intervention-free selection using EEG and eye tracking“. In: *Proceedings of the 18th ACM International Conference on Multimodal Interaction*. 2016, S. 153–160 (siehe S. 307).

[Rag20] Ragan, Eric D; Pachuilo, Andrew; Goodall, John R and Bacim, Felipe: „Preserving Contextual Awareness during Selection of Moving Targets in Animated Stream Visualizations“. In: *Proceedings of the International Conference on Advanced Visual Interfaces*. 2020, S. 1–9 (siehe S. 52, 130, 295).

[Raj16] Rajanna, Vijay and Hammond, Tracy: „Gawschi: Gaze-augmented, wearable-supplemented computer-human interaction“. In: *Proceedings of the Ninth Biennial ACM Symposium on Eye Tracking Research & Applications*. 2016, S. 233–236 (siehe S. 70).

[Ram19] Ramirez Gomez, Argenis and Gellersen, Hans: „Looking outside the box: reflecting on gaze interaction in gameplay“. In: *Proceedings of the Annual Symposium on Computer-Human Interaction in Play*. 2019, S. 625–637 (siehe S. 76, 77, 102, 130).

[Ram21] Ramirez Gomez, Argenis Ramirez; Clarke, Christopher; Sidenmark, Ludwig and Gellersen, Hans: „Gaze+ Hold: Eyes-only Direct Manipulation with Continuous Gaze Modulated by Closure of One Eye“. In: *ACM Symposium on Eye Tracking Research and Applications*. 2021, S. 1–12 (siehe S. 65).

[Ras99] Rasmusson, D; Chappell, R and Trego, M: „Quick Glance: Eye-tracking access to the Windows95 operating environment“. In: *Proceedings of the Fourteenth International Conference on Technology and Persons with Disabilities (CSUN'99)*. 1999 (siehe S. 65).

[Ren02] Rensink, Ronald A: „Change detection“. In: *Annual review of psychology* 53.1 (2002), S. 245–277 (siehe S. 37).

[Ren97] Rensink, Ronald A; O'regan, J Kevin and Clark, James J: „To see or not to see: The need for attention to perceive changes in scenes“. In: *Psychological science* 8.5 (1997), S. 368–373 (siehe S. 37).

[Ric97] Richardson, Peter and Larsen, James: „Repetitive strain injuries in the information age workplace“. In: *Human Resource Management: Published in Cooperation with the School of Business Administration, The University of Michigan and in alliance with the Society of Human Resources Management* 36.4 (1997), S. 377–384 (siehe S. 10).

[Sal00] Salvucci, Dario D and Goldberg, Joseph H: „Identifying fixations and saccades in eye-tracking protocols“. In: *Proceedings of the 2000 symposium on Eye tracking research & applications*. 2000, S. 71–78 (siehe S. 56, 58, 60–62, 132, 150, 203, 214, 288).

[Sal18] Salous, Mazen; Putze, Felix; Schultz, Tanja; Hild, Jutta and Beyerer, Jürgen: „Investigating static and sequential models for intervention-free selection using multimodal data of EEG and eye tracking“. In: *Proceedings of the Workshop on Modeling Cognitive Processes from Multimodal Data*. 2018, S. 1–6 (siehe S. 307).

[Sau91] Sauter, D; Martin, BJ; Di Renzo, N and Vomscheid, C: „Analysis of eye tracking movements using innovations generated by a Kalman filter“. In: *Medical and biological Engineering and Computing* 29.1 (1991), S. 63–69 (siehe S. 59).

[Sch07] Schlögl, Alois; Keinrath, Claudia; Zimmermann, Doris; Scherer, Reinhold; Leeb, Robert and Pfurtscheller, Gert: „A fully automated correction method of EOG artifacts in EEG recordings“. In: *Clinical neurophysiology* 118.1 (2007), S. 98–104 (siehe S. 126).

[Sch19] Schuetz, Immo; Murdison, T Scott; MacKenzie, Kevin J and Zannoli, Marina: „An Explanation of Fitts' Law-like Performance in Gaze-Based Selection Tasks Using a Psychophysics Approach“. In: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 2019, S. 1–13 (siehe S. 307).

[Scy21] Scythe. Letzter Zugriff am 26-07-2021. 2021. url: https://www.scythe-eu.com/produkte/pc-zubehoer/usb-foot-switch-ii.html%5C#specs (siehe S. 126).

[Sei09] Seiffert, Chris; Khoshgoftaar, Taghi M; Van Hulse, Jason and Napolitano, Amri: „RUSBoost: A hybrid approach to alleviating class imbalance“. In: *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans* 40.1 (2009), S. 185–197 (siehe S. 91).

[She08] Shenoy, Pradeep and Tan, Desney S: „Human-aided computing: Utilizing implicit human processing to classify images“. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 2008, S. 845–854 (siehe S. 86).

[She20] Sheskin, David J: Handbook of parametric and nonparametric statistical procedures. 5. Aufl. Chapman und Hall/CRC, 2020 (siehe S. 137).

[Shi08] Shic, Frederick; Scassellati, Brian and Chawarska, Katarzyna: „The incomplete fixation measure“. In: *Proceedings of the 2008 symposium on Eye tracking research & applications*. 2008, S. 111–114 (siehe S. 61).

[Shi94] Shi, Jianbo and Tomasi, Carlo: „Good features to track“. In: *1994 Proceedings of IEEE conference on computer vision and pattern recognition*. IEEE. 1994, S. 593–600 (siehe S. 251, 267).

[Shn18] Shneiderman, Ben; Plaisant, Catherine; Cohen, Maxine S; Jacobs, Steven; Elmqvist, Niklas and Diakopoulos, Nicholas: Designing the user interface: strategies for effective human-computer interaction. Pearson, 2018 (siehe S. 47–49, 62, 94).

[Sib00] Sibert, Linda E and Jacob, Robert JK: „Evaluation of eye gaze interaction“. In: *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*. 2000, S. 281–288 (siehe S. 64, 73, 102, 307).

[Sim99] Simons, Daniel J and Chabris, Christopher F: „Gorillas in our midst: Sustained inattentional blindness for dynamic events“. In: *perception* 28.9 (1999), S. 1059–1074 (siehe S. 37).

[Smi06] Smith, J David and Graham, TC Nicholas: „Use of eye movements for video game control“. In: *Proceedings of the 2006 ACM SIGCHI international conference on Advances in computer entertainment technology*. 2006, 20–es (siehe S. 77, 78).

[Som21] Sommer, Lars; Kruger, Wolfgang and Teutsch, Michael: „Appearance and Motion Based Persistent Multiple Object Tracking in Wide Area Motion Imagery“. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2021, S. 3878–3888 (siehe S. 307, 308).

[Špa05a] Špakov, Oleg: „EyeChess: the tutoring game with visual attentive interface“. In: *Alternative Access: Feelings & Games* 5 (2005) (siehe S. 65).

[Špa05b] Špakov, Oleg and Miniotas, Darius: „Gaze-based selection of standard-size menu items“. In: *Proceedings of the 7th international conference on Multimodal interfaces*. 2005, S. 124–128 (siehe S. 55).

[Špa12] Špakov, Oleg: „Comparison of eye movement filters used in HCI“. In: *Proceedings of the symposium on eye tracking research and applications*. 2012, S. 281–284 (siehe S. 56, 59, 60).

[Sta95] Stampe, Dave M and Reingold, Eyal M: „Selection by looking: A novel computer interface and its application to psychological research“. In: *Studies in visual information processing*. Bd. 6. Elsevier, 1995, S. 467–478 (siehe S. 65, 102, 130).

[Str17] Strentzsch, Gunnar; Camp, Florian van de and Stiefelhagen, Rainer: „Digital map table VR: Bringing an interactive system to virtual reality“. In: *International Conference on Virtual, Augmented and Mixed Reality*. Springer. 2017, S. 54–71 (siehe S. 48).

[Sur04] Surakka, Veikko; Illi, Marko and Isokoski, Poika: „Gazing and frowning as a new human–computer interaction technique“. In: *ACM Transactions on Applied Perception (TAP)* 1.1 (2004), S. 40–56 (siehe S. 71).

[Szu00] Szunejko, Monika Halina: „Managing repetitive strain injuries in bibliographic services departments“. In: *Technical Services Quarterly* 18.1 (2000), S. 33–45 (siehe S. 10).

[Tei74] Teichner, Warren H: „The detection of a simple visual signal as a function of time of watch“. In: *Human factors* 16.4 (1974), S. 339–352 (siehe S. 212).

[Teu12] Teutsch, Michael and Krüger, Wolfgang: „Detection, segmentation, and tracking of moving objects in UAV videos“. In: *2012 IEEE Ninth International Conference on Advanced Video and Signal-Based Surveillance*. IEEE. 2012, S. 313–318 (siehe S. 251, 267).

[The03] Theiler, James P and Cai, D Michael: „Resampling approach for anomaly detection in multispectral images“. In: *Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery IX*. Bd. 5093. International Society for Optics und Photonics. 2003, S. 230–240 (siehe S. 111).

[Tho14] Thompson, Atalie C; Prill, Marnie J Kremer; Biswal, Sandip; Rebner, Murray; Rebner, Rachel E; Thomas, William R; Edwards, Sonya D; Thompson, Matthew O and Ikeda, Debra M: „Factors associated with repetitive strain, and strategies to reduce injury among breast-imaging radiologists“. In: *Journal of the American College of Radiology* 11.11 (2014), S. 1074–1079 (siehe S. 10).

[Tre86] Treisman, Anne: „Features and objects in visual processing“. In: *Scientific American* 255.5 (1986), 114B–125 (siehe S. 37).

[Tre88] Treisman, Anne: „Features and objects: The fourteenth Bartlett memorial lecture“. In: *The Quarterly Journal of Experimental Psychology Section A* 40.2 (1988), S. 201–237 (siehe S. 37).

[Tre99] Treisman, Anne: „Solutions to the binding problem: progress through controversy and convergence“. In: *Neuron* 24.1 (1999), S. 105–125 (siehe S. 37).

[Tui16] Tuisku, Outi; Rantanen, Ville; Špakov, Oleg; Surakka, Veikko and Lekkala, Jukka: „Pointing and selecting with facial activity“. In: *Interacting with Computers* 28.1 (2016), S. 1–12 (siehe S. 71).

[Van11] Van der Kamp, Jan and Sundstedt, Veronica: „Gaze and voice controlled drawing“. In: *Proceedings of the 1st conference on novel gaze-controlled applications*. 2011, S. 1–8 (siehe S. 70).

[Vel15a] Velloso, Eduardo; Oechsner, Carl; Sachmann, Katharina; Wirth, Markus and Gellersen, Hans: „Arcade+ A Platform for Public Deployment and Evaluation of Multi-Modal Games“. In: *Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play*. 2015, S. 271–275 (siehe S. 79, 80, 102, 130).

[Vel15b] Velloso, Eduardo; Schmidt, Dominik; Alexander, Jason; Gellersen, Hans and Bulling, Andreas: „The feet in human–computer interaction: A survey of foot-based interaction“. In: *ACM Computing Surveys (CSUR)* 48.2 (2015), S. 1–35 (siehe S. 69).

[Vel16] Velloso, Eduardo and Carter, Marcus: „The emergence of eyeplay: a survey of eye interaction in games“. In: *Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play*. 2016, S. 171–185 (siehe S. 76, 82, 103).

[Vel17] Velloso, Eduardo; Carter, Marcus; Newn, Joshua; Esteves, Augusto; Clarke, Christopher and Gellersen, Hans: „Motion correlation: Selecting objects by matching their movement“. In: *ACM Transactions on Computer-Human Interaction (TOCHI)* 24.3 (2017), S. 1–35 (siehe S. 82).

[Ver08] Vertegaal, Roel: „A Fitts Law comparison of eye tracking and manual input in the selection of visual targets“. In: *Proceedings of the 10th international conference on Multimodal interfaces*. 2008, S. 241–248 (siehe S. 41, 64, 66, 68, 93, 102, 130, 193, 231).

[Vid13] Vidal, Mélodie; Bulling, Andreas and Gellersen, Hans: „Pursuits: spontaneous interaction with displays based on smooth pursuit eye movement and moving targets“. In: *Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing*. 2013, S. 439–448 (siehe S. 80–82, 103).

[Vid15] Vidal, Melodie; Bismuth, Remi; Bulling, Andreas and Gellersen, Hans: „The royal corgi: Exploring social gaze interaction for immersive gameplay“. In: *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. 2015, S. 115–124 (siehe S. 77).

[War86] Ware, Colin and Mikaelian, Harutune H: „An evaluation of an eye tracker as a device for computer input2“. In: *Proceedings of the SIGCHI/GI conference on Human factors in computing systems and graphics interface*. 1986, S. 183–188 (siehe S. 64–67, 102, 131).

[Wer10] Wermke, Matthias; Kunkel-Razum, Kathrin and Scholze-Stubenrecht, Werner: „Der Duden in zwölf Bänden, Das Standardwerk zur deutschen Sprache, Band 10 Bedeutungswörterbuch, 4“. In: *Auflage, Mannheim* (2010) (siehe S. 111, 112).

[Wid84] Widdel, Heino: „Operational problems in analysing eye movements“. In: *Advances in psychology*. Bd. 22. Elsevier, 1984, S. 21–29 (siehe S. 62).

[Wig07] Wigdor, Daniel; Forlines, Clifton; Baudisch, Patrick; Barnwell, John and Shen, Chia: „Lucid touch: a see-through mobile device“. In: *Proceedings of the 20th annual ACM symposium on User interface software and technology*. 2007, S. 269–278 (siehe S. 48).

[Wil08] Wilcox, Tom; Evans, Mike; Pearce, Chris; Pollard, Nick and Sundstedt, Veronica: „Gaze and voice based game interaction: the revenge of the killer penguins.“ In: *SIGGRAPH Posters* 81.10.1145 (2008), S. 1400885–1400972 (siehe S. 65, 70).

[Yar67] Yarbus, Alfred L: „Eye movements during perception of complex objects“. In: *Eye movements and vision*. Springer, 1967, S. 171–211 (siehe S. 89).

[Yon11] Yong, Xinyi; Fatourechi, Mehrdad; Ward, Rabab K and Birch, Gary E: „The design of a point-and-click system by integrating a self-paced brain–computer interface with an Eye-tracker“. In: *IEEE Journal on Emerging and Selected Topics in Circuits and Systems* 1.4 (2011), S. 590–602 (siehe S. 87).

[Zan10] Zander, Thorsten O; Gaertner, Matti; Kothe, Christian and Vilimek, Roman: „Combining eye gaze input with a brain–computer interface for touchless human–computer interaction“. In: *Intl. Journal of Human–Computer Interaction* 27.1 (2010), S. 38–51 (siehe S. 87).

[Zha03] Zhai, Shumin: „What's in the Eyes for Attentive Input“. In: *Communications of the ACM* 46.3 (2003), S. 34–39 (siehe S. 54).

[Zha07] Zhang, Xuan and MacKenzie, I Scott: „Evaluating eye tracking with ISO 9241-Part 9“. In: *International Conference on Human-Computer Interaction*. Springer. 2007, S. 779–788 (siehe S. 66–68, 93, 102, 146, 152, 183, 197, 241, 258).

[Zha10] Zhang, Xinyong; Ren, Xiangshi and Zha, Hongbin: „Modeling dwell-based eye pointing target acquisition“. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 2010, S. 2083–2092 (siehe S. 307).

[Zha20] Zhao, Darisy G; Karikov, Nikita D; Melnichuk, Eugeny V; Velichkovsky, Boris M and Shishkin, Sergei L: „Voice as a mouse click: Usability and effectiveness of simplified hands-free gaze-voice selection“. In: *Applied Sciences* 10.24 (2020), S. 8791 (siehe S. 84, 307).

[Zha21] Zhang, Xinyong: „An Evaluation of Eye-Foot Input for Target Acquisitions“. In: *International Conference on Human-Computer Interaction*. Springer. 2021, S. 499–517 (siehe S. 70, 307).

[Zha99] Zhai, Shumin; Morimoto, Carlos and Ihde, Steven: „Manual and gaze input cascaded (MAGIC) pointing“. In: *Proceedings of the SIGCHI conference on Human factors in computing systems*. 1999, S. 246–253 (siehe S. 10, 66, 71, 73, 74, 132, 144, 157, 159, 299).

Eigene Publikationen

Dieser Abschnitt enthält ein vollständiges Verzeichnis der eigenen Veröffentlichungen.

[1] Hild, Jutta; Eckel, Susanne and Geisler, Jürgen: „Representing Uncertainty in Situation Maps for Disaster Management“. In: *Mobile Response Workshop at the 6th International Conference on Information Systems for Crisis Response and Management (ISCRAM)*. 2009.

[2] Ott, Jonathan; Hild, Jutta and Bauer, Alexander: „Decision Support to Facilitate Cost-Optimal Response in Time-and-Safety-Critical Situations“. In: *Proceedings of the 4th Security Research Conference (Future Security)*. 2009, S. 322–338.

[3] Hild, Jutta; Ott, Jonathan; Fischer, Yvonne and Glökler, Christian: „Markov based decision support for cost-optimal response in security management.“ In: *Proceedings of the 7th International Conference on Information Systems for Crisis Response and Management (ISCRAM)*. 2010, S. 1–5.

[4] Hild, Jutta; Ott, Jonathan and Peinsipp-Byma, Elisabeth: „Toward intelligent decision support for security staff: evaluation of an interactive resource management system based on a CMDP model“. In: *Sensors, and Command, Control, Communications, and Intelligence (C3I) Technologies for Homeland Security and Homeland Defense X*. Bd. 8019. SPIE. 2011, S. 53–67.

[5] Hild, Jutta; Peinsipp-Byma, Elisabeth and Klaus, Edmund: „Improving usability for video analysis using gaze-based interaction“.

In: *Full Motion Video (FMV) Workflows and Technologies for Intelligence, Surveillance, and Reconnaissance (ISR) and Situational Awareness*. Bd. 8386. International Society for Optics und Photonics. 2012, S. 838605.

[6] Hild, J; Müller, E; Klaus, E; Peinsipp-Byma, E and Beyerer, J: „Evaluating multi-modal eye gaze interaction for moving object selection". In: *Proc. ACHI* (2013), S. 454–459.

[7] Hild, Jutta; Brüstle, Stefan; Heinze, Norbert and Peinsipp-Byma, Elisabeth: „Gaze interaction in UAS video exploitation". In: *Motion Imagery Technologies, Best Practices, and Workflows for Intelligence, Surveillance, and Reconnaissance (ISR), and Situational Awareness*. Bd. 8740. International Society for Optics und Photonics. 2013, 87400H.

[8] Hild, Jutta; Fischer, Yvonne; Peinsipp-Byma, Elisabeth and Beyerer, Jürgen: „Gaze-based interaction for real-time video surveillance systems". In: *Proceedings of the 8th Security Research Conference (Future Security)*. Fraunhofer Verlag, 2013, S. 63–72.

[9] Putze, Felix; Hild, Jutta; Kärgel, Rainer; Herff, Christian; Redmann, Alexander; Beyerer, Jürgen and Schultz, Tanja: „Locating user attention using eye tracking and EEG for spatio-temporal event selection". In: *Proceedings of the 2013 international conference on Intelligent user interfaces*. 2013, S. 129–136.

[10] Hild, Jutta; Gill, Dennis and Beyerer, Jürgen: „Comparing mouse and magic pointing for moving target acquisition". In: *Proceedings of the Symposium on Eye Tracking Research and Applications*. 2014, S. 131–134.

[11] Hild, Jutta; Putze, Felix; Kaufman, David; Kühnle, Christian; Schultz, Tanja and Beyerer, Jürgen: „Spatio-temporal event selection in basic surveillance tasks using eye tracking and EEG". In: *Proceedings of the 7th Workshop on Eye Gaze in Intelligent Human Machine Interaction: Eye-Gaze & Multimodality*. 2014, S. 3–8.

[12] Hild, Jutta; Krüger, Wolfgang; Brüstle, Stefan; Trantelle, Patrick; Unmüßig, Gabriel; Heinze, Norbert; Peinsipp-Byma, Elisabeth and Beyerer, Jürgen: „Collaborative real-time motion video analysis by human observer and image exploitation algorithms". In: *Motion Imagery: Standards, Quality, and Interoperability*. Bd. 9463. International Society for Optics und Photonics. 2015, S. 94630C.

[13] Balthasar, Sebastian; Martin, Manuel; Camp, Florian van de; Hild, Jutta and Beyerer, Jürgen: „Combining low-cost eye trackers for dual monitor eye tracking". In: *International Conference on Human-Computer Interaction*. Springer. 2016, S. 3–12.

[14] Camp, Florian van de; Gill, Dennis; Hild, Jutta and Beyerer, Jürgen: „An Analysis of Accuracy Requirements for Automatic Eye-tracker Recalibration at Runtime". In: *International Conference on Human-Computer Interaction*. Springer. 2016, S. 149–154.

[15] Hild, Jutta; Krüger, Wolfgang; Heinze, Norbert; Peinsipp-Byma, Elisabeth and Beyerer, Jürgen: „Interacting with target tracking algorithms in a gaze-enhanced motion video analysis system". In: *Geospatial Informatics, Fusion, and Motion Video Analytics VI*. Bd. 9841. International Society for Optics und Photonics. 2016, 98410K.

[16] Hild, Jutta; Kühnle, Christian and Beyerer, Jürgen: „Gaze-based moving target acquisition in real-time full motion video". In: *Proceedings of the Ninth Biennial ACM Symposium on Eye Tracking Research & Applications*. 2016, S. 241–244.

[17] Hild, Jutta; Petersen, Patrick and Beyerer, Jürgen: „Moving target acquisition by gaze pointing and button press using hand or foot". In: *Proceedings of the Ninth Biennial ACM Symposium on Eye Tracking Research & Applications*. 2016, S. 257–260.

[18] Putze, Felix; Popp, Johannes; Hild, Jutta; Beyerer, Jürgen and Schultz, Tanja: „Intervention-free selection using EEG and eye tracking". In: *Proceedings of the 18th ACM International Conference on Multimodal Interaction*. 2016, S. 153–160.

[19] Hild, Jutta; Krüger, Wolfgang; Brüstle, Stefan; Trantelle, Patrick; Unmüßig, Gabriel; Voit, Michael; Heinze, Norbert; Peinsipp-Byma, Elisabeth and Beyerer, Jürgen: „Pilot study on real-time motion detection in UAS video data by human observer and image exploitation algorithm". In: *Geospatial Informatics, Fusion, and Motion Video Analytics VII*. Bd. 10199. International Society for Optics und Photonics. 2017, S. 1019903.

[20] Hild, Jutta; Klaus, Edmund; Hammer, Jan-Hendrik; Martin, Manuel; Voit, Michael; Peinsipp-Byma, Elisabeth and Beyerer, Jürgen: „A Pilot Study on Gaze-Based Control of a Virtual Camera Using 360°-Video Data". In: *International Conference on Engineering Psychology and Cognitive Ergonomics*. Springer. 2018, S. 419–428.

[21] Hild, Jutta; Saur, Günter; Petersen, Patrick; Voit, Michael; Peinsipp-Byma, Elisabeth and Beyerer, Jürgen: „Evaluating user interfaces supporting change detection in aerial images and aerial image sequences". In: *International Conference on Human Interface and the Management of Information*. Springer. 2018, S. 383–402.

[22] Hild, Jutta; Voit, Michael; Kühnle, Christian and Beyerer, Jürgen: „Predicting observer's task from eye movement patterns during motion image analysis". In: *Proceedings of the 2018 ACM symposium on eye tracking research & applications*. 2018, S. 1–5.

[23] Putze, Felix; Hild, Jutta; Sano, Akane; Kasneci, Enkelejda; Solovey, Erin and Schultz, Tanja: „Modeling cognitive processes from multimodal signals". In: *Proceedings of the 20th ACM International Conference on Multimodal Interaction*. 2018, S. 663–663.

[24] Salous, Mazen; Putze, Felix; Schultz, Tanja; Hild, Jutta and Beyerer, Jürgen: „Investigating static and sequential models for intervention-free selection using multimodal data of EEG and eye tracking". In: *Proceedings of the Workshop on Modeling Cognitive Processes from Multimodal Data*. 2018, S. 1–6.

[25] Hild, Jutta; Peinsipp-Byma, Elisabeth; Voit, Michael and Beyerer, Jürgen: „Suggesting Gaze-based Selection for Surveillance Applications“. In: *2019 16th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*. IEEE. 2019, S. 1–8.

[26] Hild, Jutta; Voit, Michael and Peinsipp-Byma, Elisabeth: „Estimating Immersed User States from Eye Movements: A Survey“. In: *International Conference on Human-Computer Interaction*. Springer. 2020, S. 337–342.

Betreute studentische Arbeiten

[1] Fabry, Daniel: „Untersuchung des Blickverhaltens während der Durchführung von Aufgaben im Bereich der bildgestützten Aufklärung “. Diplomarbeit. Karlsruher Institut für Technologie (KIT), 2011.

[2] Gill, Dennis: „MAGIC pointing für die Selektion sich bewegender Objekte“. Bachelorarbeit. Karlsruher Institut für Technologie (KIT), 2012.

[3] Müller, Elke: „Blick+Taste-Interaktion für Selektionsaufgaben bei der Videobildauswertung“. Diplomarbeit. Karlsruher Institut für Technologie (KIT), 2012.

[4] Panic, Delia Lara: „Visualisierung des Blickverhaltens als Assistenz bei Bildauswerteaufgaben“. Diplomarbeit. Karlsruher Institut für Technologie (KIT), 2012.

[5] Cicek, Tuna Murat: „Implementierung blickbasierter Selektion für die Benutzungsoberfläche des dynamischen Lagebildes des NEST-Überwachungssystems“. Studienarbeit. Karlsruher Institut für Technologie (KIT), 2014.

[6] Gill, Dennis: „Kalibrierung für ein Tisch-basiertes Blickmessgerät vor und während der Systemnutzung eines Desktop-Computers “. Masterarbeit. Karlsruher Institut für Technologie (KIT), 2015.

[7] Guthmann, Micha-Jamie: „Der Einfluss von Lerneffekten auf die Leistungsfähigkeit des Systemnutzers bei blickbasierter Bewegtobjektselektion“. Masterarbeit. Karlsruher Institut für Technologie (KIT), 2015.

[8] Petersen, Patrick: „Bewegtobjektselektion mittels Blick+ Tastendruck-Interaktionstechniken mit Hand- oder Fußtaste“. Bachelorarbeit. Karlsruher Institut für Technologie (KIT), 2015.

[9] Kühnle, Christian: „Klassifikation von Aufgaben aus Blickbewegungsmustern bei der Analyse dynamischer Szenen“. Masterarbeit. Karlsruher Institut für Technologie (KIT), 2017.

[10] Petersen, Patrick: „Konzeption, Implementierung und Evaluation von Benutzungsschnittstellen zur Änderungsdetektion in Bildern und Bildfolgen“. Masterarbeit. Karlsruher Institut für Technologie (KIT), 2017.

Abbildungsverzeichnis

Tabellenverzeichnis

Karlsruher Schriftenreihe zur Anthropomatik (ISSN 1863-6489)

Band 1 Jürgen Geisler
Leistung des Menschen am Bildschirmarbeitsplatz.
ISBN 3-86644-070-7

Band 2 Elisabeth Peinsipp-Byma
Leistungserhöhung durch Assistenz in interaktiven Systemen zur Szenenanalyse. 2007
ISBN 978-3-86644-149-1

Band 3 Jürgen Geisler, Jürgen Beyerer (Hrsg.)
Mensch-Maschine-Systeme.
ISBN 978-3-86644-457-7

Band 4 Jürgen Beyerer, Marco Huber (Hrsg.)
Proceedings of the 2009 Joint Workshop of Fraunhofer IOSB and Institute for Anthropomatics, Vision and Fusion Laboratory.
ISBN 978-3-86644-469-0

Band 5 Thomas Usländer
Service-oriented design of environmental information systems.
ISBN 978-3-86644-499-7

Band 6 Giulio Milighetti
Multisensorielle diskret-kontinuierliche Überwachung und Regelung humanoider Roboter.
ISBN 978-3-86644-568-0

Band 7 Jürgen Beyerer, Marco Huber (Hrsg.)
Proceedings of the 2010 Joint Workshop of Fraunhofer IOSB and Institute for Anthropomatics, Vision and Fusion Laboratory.
ISBN 978-3-86644-609-0

Band 8 Eduardo Monari
Dynamische Sensorselektion zur auftragsorientierten Objektverfolgung in Kameranetzwerken.
ISBN 978-3-86644-729-5

Die Bände sind unter www.ksp.kit.edu als PDF frei verfügbar oder als Druckausgabe bestellbar.

Band 9 Thomas Bader
Multimodale Interaktion in Multi-Display-Umgebungen.
ISBN 3-86644-760-8

Band 10 Christian Frese
Planung kooperativer Fahrmanöver für kognitive Automobile.
ISBN 978-3-86644-798-1

Band 11 Jürgen Beyerer, Alexey Pak (Hrsg.)
Proceedings of the 2011 Joint Workshop of Fraunhofer IOSB and Institute for Anthropomatics, Vision and Fusion Laboratory.
ISBN 978-3-86644-855-1

Band 12 Miriam Schleipen
Adaptivität und Interoperabilität von Manufacturing Execution Systemen (MES).
ISBN 978-3-86644-955-8

Band 13 Jürgen Beyerer, Alexey Pak (Hrsg.)
Proceedings of the 2012 Joint Workshop of Fraunhofer IOSB and Institute for Anthropomatics, Vision and Fusion Laboratory.
ISBN 978-3-86644-988-6

Band 14 Hauke-Hendrik Vagts
Privatheit und Datenschutz in der intelligenten Überwachung: Ein datenschutzgewährendes System, entworfen nach dem „Privacy by Design" Prinzip.
ISBN 978-3-7315-0041-4

Band 15 Christian Kühnert
Data-driven Methods for Fault Localization in Process Technology. 2013
ISBN 978-3-7315-0098-8

Band 16 Alexander Bauer
Probabilistische Szenenmodelle für die Luftbildauswertung.
ISBN 978-3-7315-0167-1

Band 17 Jürgen Beyerer, Alexey Pak (Hrsg.)
Proceedings of the 2013 Joint Workshop of Fraunhofer IOSB and Institute for Anthropomatics, Vision and Fusion Laboratory.
ISBN 978-3-7315-0212-8

Die Bände sind unter www.ksp.kit.edu als PDF frei verfügbar oder als Druckausgabe bestellbar.

Band 18 Michael Teutsch
Moving Object Detection and Segmentation for Remote Aerial Video Surveillance.
ISBN 978-3-7315-0320-0

Band 19 Marco Huber
Nonlinear Gaussian Filtering: Theory, Algorithms, and Applications.
ISBN 978-3-7315-0338-5

Band 20 Jürgen Beyerer, Alexey Pak (Hrsg.)
Proceedings of the 2014 Joint Workshop of Fraunhofer IOSB and Institute for Anthropomatics, Vision and Fusion Laboratory.
ISBN 978-3-7315-0401-6

Band 21 Todor Dimitrov
Permanente Optimierung dynamischer Probleme der Fertigungssteuerung unter Einbeziehung von Benutzerinteraktionen.
ISBN 978-3-7315-0426-9

Band 22 Benjamin Kühn
Interessengetriebene audiovisuelle Szenenexploration.
ISBN 978-3-7315-0457-3

Band 23 Yvonne Fischer
Wissensbasierte probabilistische Modellierung für die Situationsanalyse am Beispiel der maritimen Überwachung.
ISBN 978-3-7315-0460-3

Band 24 Jürgen Beyerer, Alexey Pak (Hrsg.)
Proceedings of the 2015 Joint Workshop of Fraunhofer IOSB and Institute for Anthropomatics, Vision and Fusion Laboratory.
ISBN 978-3-7315-0519-8

Band 25 Pascal Birnstill
Privacy-Respecting Smart Video Surveillance Based on Usage Control Enforcement.
ISBN 978-3-7315-0538-9

Band 26 Philipp Woock
Umgebungskartenschätzung aus Sidescan-Sonardaten für ein autonomes Unterwasserfahrzeug.
ISBN 978-3-7315-0541-9

Band 27 Janko Petereit
Adaptive State × Time Lattices: A Contribution to Mobile Robot Motion Planning in Unstructured Dynamic Environments.
ISBN 978-3-7315-0580-8

Band 28 Erik Ludwig Krempel
Steigerung der Akzeptanz von intelligenter Videoüberwachung in öffentlichen Räumen.
ISBN 978-3-7315-0598-3

Band 29 Jürgen Moßgraber
Ein Rahmenwerk für die Architektur von Frühwarnsystemen. 2017
ISBN 978-3-7315-0638-6

Band 30 Andrey Belkin
World Modeling for Intelligent Autonomous Systems.
ISBN 978-3-7315-0641-6

Band 31 Chettapong Janya-Anurak
Framework for Analysis and Identification of Nonlinear Distributed Parameter Systems using Bayesian Uncertainty Quantification based on Generalized Polynomial Chaos.
ISBN 978-3-7315-0642-3

Band 32 David Münch
Begriffliche Situationsanalyse aus Videodaten bei unvollständiger und fehlerhafter Information.
ISBN 978-3-7315-0644-7

Band 33 Jürgen Beyerer, Alexey Pak (Eds.)
Proceedings of the 2016 Joint Workshop of Fraunhofer IOSB and Institute for Anthropomatics, Vision and Fusion Laboratory.
ISBN 978-3-7315-0678-2

Band 34 Jürgen Beyerer, Alexey Pak and Miro Taphanel (Eds.)
Proceedings of the 2017 Joint Workshop of Fraunhofer IOSB and Institute for Anthropomatics, Vision and Fusion Laboratory.
ISBN 978-3-7315-0779-6

Band 35 Michael Grinberg
Feature-Based Probabilistic Data Association for Video-Based Multi-Object Tracking.
ISBN 978-3-7315-0781-9

Band 36 Christian Herrmann
Video-to-Video Face Recognition for Low-Quality Surveillance Data.
ISBN 978-3-7315-0799-4

Band 37 Chengchao Qu
Facial Texture Super-Resolution by Fitting 3D Face Models.
ISBN 978-3-7315-0828-1

Band 38 Miriam Ruf
Geometrie und Topologie von Trajektorienoptimierung für vollautomatisches Fahren.
ISBN 978-3-7315-0832-8

Band 39 Angelika Zube
Bewegungsregelung mobiler Manipulatoren für die Mensch-Roboter-Interaktion mittels kartesischer modellprädiktiver Regelung.
ISBN 978-3-7315-0855-7

Band 40 Jürgen Beyerer and Miro Taphanel (Eds.)
Proceedings of the 2018 Joint Workshop of Fraunhofer IOSB and Institute for Anthropomatics, Vision and Fusion Laboratory.
ISBN 978-3-7315-0936-3

Band 41 Marco Thomas Gewohn
Ein methodischer Beitrag zur hybriden Regelung der Produktionsqualität in der Fahrzeugmontage.
ISBN 978-3-7315-0893-9

Band 42 Tianyi Guan
Predictive energy-efficient motion trajectory optimization of electric vehicles.
ISBN 978-3-7315-0978-3

Band 43 Jürgen Metzler
Robuste Detektion, Verfolgung und Wiedererkennung von Personen in Videodaten mit niedriger Auflösung.
ISBN 978-3-7315-0968-4

Band 44 Sebastian Bullinger
Image-Based 3D Reconstruction of Dynamic Objects Using Instance-Aware Multibody Structure from Motion.
ISBN 978-3-7315-1012-3

Band 45 Jürgen Beyerer, Tim Zander (Eds.)
Proceedings of the 2019 Joint Workshop of Fraunhofer IOSB and Institute for Anthropomatics, Vision and Fusion Laboratory.
ISBN 978-3-7315-1028-4

Band 46 Stefan Becker
Dynamic Switching State Systems for Visual Tracking.
ISBN 978-3-7315-1038-3

Band 47 Jennifer Sander
Ansätze zur lokalen Bayes'schen Fusion von Informationsbeiträgen heterogener Quellen.
ISBN 978-3-7315-1062-8

Band 48 Philipp Christoph Sebastian Bier
Umsetzung des datenschutzrechtlichen Auskunftsanspruchs auf Grundlage von Usage-Control und Data-Provenance-Technologien.
ISBN 978-3-7315-1082-6

Band 49 Thomas Emter
Integrierte Multi-Sensor-Fusion für die simultane Lokalisierung und Kartenerstellung für mobile Robotersysteme.
ISBN 978-3-7315-1074-1

Band 50 Patrick Dunau
Tracking von Menschen und menschlichen Zuständen.
ISBN 978-3-7315-1086-4

Band 51 Jürgen Beyerer, Tim Zander (Eds.)
Proceedings of the 2020 Joint Workshop of Fraunhofer IOSB and Institute for Anthropomatics, Vision and Fusion Laboratory.
ISBN 978-3-7315-1091-8

Band 52 Lars Wilko Sommer
Deep Learning based Vehicle Detection in Aerial Imagery.
ISBN 978-3-7315-1113-7

Band 53 Jan Hendrik Hammer
Interaktionstechniken für mobile Augmented-Reality-Anwendungen basierend auf Blick- und Handbewegungen.
ISBN 978-3-7315-1169-4

Band 54 Jürgen Beyerer, Tim Zander (Eds.)
Proceedings of the 2021 Joint Workshop of Fraunhofer IOSB and Institute for Anthropomatics, Vision and Fusion Laboratory.
ISBN 978-3-7315-1171-7

Band 55 Ronny Hug
Probabilistic Parametric Curves for Sequence Modeling.
ISBN 978-3-7315-1198-4

Band 56 Florian Patzer
Automatisierte, minimalinvasive Sicherheitsanalyse und Vorfallreaktion für industrielle Systeme.
ISBN 978-3-7315-1207-3

Band 57 Achim Christian Kuwertz
Adaptive Umweltmodellierung für kognitive Systeme in offener Welt durch dynamische Konzepte und quantitative Modellbewertung.
ISBN 978-3-7315-1219-6

Band 58 Julius Pfrommer
Distributed Planning for Self-Organizing Production Systems.
ISBN 978-3-7315-1253-0

Band 59 Ankush Meshram
Self-learning Anomaly Detection in Industrial Production.
ISBN 978-3-7315-1257-8

Band 60 Patrick Philipp
Über die Formalisierung und Analyse medizinischer Prozesse im Kontext von Expertenwissen und künstlicher Intelligenz.
ISBN 978-3-7315-1289-9

Band 61 Mathias Anneken
Anomaliedetektion in räumlich-zeitlichen Datensätzen.
ISBN 978-3-7315-1300-1

Band 62 Jürgen Beyerer, Tim Zander (Eds.)
Proceedings of the 2022 Joint Workshop of Fraunhofer IOSB and Institute for Anthropomatics, Vision and Fusion Laboratory.
ISBN 978-3-7315-1304-9

Band 63 Fabian Dürr
Multimodal Panoptic Segmentation of 3D Point Clouds.
ISBN 978-3-7315-1314-8

Band 64 Jutta Hild
Nutzung von Blickbewegungen für die Mensch-Computer-Interaktion mit dynamischen Bildinhalten am Beispiel der Videobildauswertung.
ISBN 978-3-7315-1330-8

Lehrstuhl für Interaktive Echtzeitsysteme
Karlsruher Institut für Technologie

Fraunhofer-Institut für Optronik, Systemtechnik
und Bildauswertung IOSB Karlsruhe

Die Interaktion mit dynamischen Bildinhalten ist für Systemnutzer herausfordernd bezüglich Wahrnehmung, Kognition und Motorik. Die Ergänzung der Benutzungsschnittstelle durch Eyetracking ermöglicht die Nutzung von Blickbewegungen für die Interaktion. Die vorliegende Arbeit erforscht, ob Systemnutzer dadurch mit höherer Leistung und geringerer Belastung interagieren können.

Beim Eyetracking erfasst ein Eyetracker die Blickbewegungen und liefert so einen Hinweis auf den Fokus der visuellen Aufmerksamkeit. Blickbasierte Interaktion gilt daher als intuitiv für Zeigeoperationen, da der Mensch gewöhnlich an die Stelle blickt, an der eine Systemeingabe erfolgt.

Die vorliegende Arbeit identifiziert geeignete blickbasierte Interaktionstechniken zur Selektion bewegter Objekte in Bildfolgen mithilfe zahlreicher Querschnitt- und einer Längsschnittstudie. Zudem wird untersucht, wie blickbasierte Interaktion zusammen mit automatischen Bildauswerteverfahren die Systemnutzer bei der Videobildauswertung unterstützt und ob blickbasierte Klassifikation der Benutzertätigkeit für typische Bildauswertetätigkeiten möglich ist.

ISSN 1863-6489
ISBN 978-3-7315-1330-8