Lehrstuhl für Mensch-Maschine-Kommunikation
Technische Universität München

Berührungslose Bedienung von Infotainment-Systemen im Fahrzeug

Michael Geiger

Vollständiger Abdruck der von der Fakultät für Elektrotechnik und Informationstechnik der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktor-Ingenieurs (Dr.-Ing.)

genehmigten Dissertation.

Vorsitzender:	Univ.-Prof. Dr.-Ing. Ulrich Wagner
Prüfer der Dissertation:	1. Univ.-Prof. Dr. rer. nat. Manfred Karl Lang, i.R.
	2. Univ.-Prof. Dr.-Ing. Klaus Diepold

Die Dissertation wurde am 16.04.2003 bei der Technischen Universität München eingereicht und durch die Fakultät für Elektrotechnik und Informationstechnik am 26.09.2003 angenommen.

Vorwort

Die vorliegende Arbeit ist das Ergebnis meiner Forschungsarbeiten am Lehrstuhl für Mensch-Maschine-Kommunikation der Technischen Universität München.

An aller erster Stelle möchte ich meinem Doktorvater Herrn Prof. Manfred Lang für die Möglichkeit zur Durchführung dieser Arbeit an seinem Institut großen Dank aussprechen. Das von ihm geförderte, angenehme Forschungsklima sowie seine stete Bereitschaft, bei auftauchenden Problemen für fachliche Diskussionen zur Verfügung zu stehen, trugen maßgeblich zum Gelingen dieses Projektes bei.

Außerordentlich gerne erinnere ich mich an die freundschaftliche wie fruchtbare Zusammenarbeit im ADVIA-Team - mit meinem nervenstarken Zimmerkollegen Martin Zobl, sowie Ralf Nieschulz, Marc Hofmann, Björn Schuller und Robert Neuss.

Ein großes Dankeschön richte ich auch an Frank Wallhoff für die Bereitstellung seiner HTK-Ressourcen sowie seines breitgefächerten Wissens über den akrobatischen Umgang mit Hidden-Markov-Modellen. In Spezialfragen zur *C++* Programmierung war oftmals Robert Lieb meine letzte Rettung.

Für seine wertvolle Unterstützung bei der Umsetzung eines ansprechenden Bedienoberflächendesigns möchte ich mich bei Alexander Fürlinger vielmals bedanken.

Peter Brand danke ich für Bereitstellung einer allzeit bestens gepflegten technischen Infrastruktur, sowie Heiner Hundhammer und Ernst Ertl für die immer freundliche Unterstützung in jeglichen Hardware-Belangen.

Für ihre wertvollen Beiträge bedanke ich mich bei meinen Werkstudenten und Diplomanden, insbesondere bei Andrea Nobbe, Phillip Fleischmann und Wolfgang Fraedrich.

Besonders ausgezeichneten Dank spreche ich meinem ehemaligen Kollegen und Freund Prof. Joschi Hunsinger aus, an dessen kongenialen Intellekt ich mich in unzähligen Debatten erfreuen durfte und der mir durch seine unversiegbaren Ratschläge nicht selten Passagen durch augenscheinlich unbegehbares Terrain aufzeigte.

Nicht zuletzt möchte ich an dieser Stelle meinen Eltern danken, die mir das Studium ermöglichten und mich zudem in meiner Entscheidung bekräftigten, den Weg zu gehen, der zu dieser Arbeit führte.

Für ihre großartige Hilfsbereitschaft möchte ich mich schließlich ganz herzlich bei meiner Korrekturleserin Britta Janowetz bedanken.

München, im April 2003

Michael Geiger

Zusammenfassung

Die vorliegende Arbeit beschreibt die Realisierung eines Gesamtkonzepts zur berührungslosen Bedienung eines *Infotainmentsystems* im Kraftfahrzeug. Von zentralem Interesse ist hierbei die Nutzung eines - in dieser Domäne - neuartigen Informationskanals: die visuelle Interaktion mittels dynamischer Hand- und Kopfgesten. Dieses Vorhaben erfolgt in der Absicht, dem Fahrer die Bedienung stetig anwachsender Funktionsumfänge zu erleichtern oder überhaupt erst zu ermöglichen. Erreicht wird dies durch die ergonomische Auslastung der menschlichen Kapazitäten, wobei eine Anlehnung an die natürlichen Dialogformen vollzogen wird.

Motiviert durch positive Befunde aus der gezielten Untersuchung kognitiver Ablenkungseffekte bei der gestischen Bedienung sowie aus grundlegenden Benutzerstudien erfolgte die Entwicklung einer gestenoptimierten Bedienumgebung im Fahrzeug. Diese wurde mit der Anbindung eines sprachverstehenden Systems zu einer *multimodalen* Applikation erweitert. Die Grundphilosophie des hier verfolgten Ansatzes ist die optimale Nutzung der jeweiligen modalitätenspezifischen Stärken von Gestik und Sprache. Eine abschließende Evaluierung erfolgte durch Usability-Tests im lehrstuhleigenen Fahrsimulator und erbrachte vielversprechende Ergebnisse: Die gestenoptimierte Bedienumgebung ist selbst von ungeübten Benutzern - auch während einer (simulierten) Autofahrt - weitgehend intuitiv bedienbar und findet hohe Akzeptanz.

Ein weiterer Schwerpunkt dieser Arbeit besteht in der Entwicklung einer neuartigen Technologie zur automatischen Erkennung dynamischer Hand- und Kopfgesten unter Berücksichtigung der restriktiven Randbedingungen, die deren potenzieller Einsatz im Fahrzeug mit sich bringt. Diese sind unter anderem: Wirtschaftlichkeit bezüglich Gesamtkosten und Bedarf an Rechenleistung, Robustheit trotz extremer Fremdlichteinflüsse und variierender komplexer Bildhintergründe, Echtzeitfähigkeit sowohl während der Erkennung als auch in der Trainingsphase, sowie der Verzicht auf körperkontaktierende Technik.

Im Gegensatz zu zahlreich vorhandenen videobasierten Ansätzen beruht der vorgestellte Lösungsweg auf der Merkmalgewinnung durch Infrarot-Distanz-Mess-Sensoren, welche zu Arrays verschaltet werden. Die auf diese Weise gewonnenen räumlichen Gesteninformationen werden vorverarbeitet und einem Mustererkennungsverfahren unterzogen, bei dem sie mittels *Dynamic-Time-Warping* (DTW) klassifiziert werden.

Die Evaluierung erfolgt unter Verwendung eines Gesteninventars, das sich aus zwölf Handgesten sowie acht Kopfgesten zusammensetzt. Zur Abschätzung der Erkennungsleistung des hier verfolgten DTW-Ansatzes wird ein durchgängiger Vergleich zur rechenaufwändigeren Erkennung mittels *Hidden-Markov-Modellen* (HMMs) gezogen. Das Gestenerkennungssystem erbringt hierbei Erkennungsraten von weit über 90 % und weist dabei ein deutlich besseres Kosten-Nutzen-Verhältnis auf als die HMM-Erkennung. Zudem verhält es sich robust gegen Fremdeinflüsse wie z.B. Sonnenlicht und erfüllt die Forderung nach Echtzeitfähigkeit bei äußerst geringem Bedarf an Rechenleistung.

Inhaltsverzeichnis

1

Einleitung

1.1 Ausgangssituation

> *„Die Kehrseite des Fortschritts: Noch nie haben wir uns so oft verirrt und verrannt."*
> ERNST FERSTL (Lehrer, Dichter und Aphoristiker).

Übertragen auf heutige Systeme der Informations- und Kommunikationstechnik beschreibt dieses Zitat deren häufig schwer durchschaubare sowie benutzerunfreundliche Bedienung auf sehr treffende Weise.

Obgleich es sich hierbei um ein allgemein und alt bekanntes Dilemma handelt, wurde es bis vor wenigen Jahren nur spärlich anerkannt und aus diesem Grunde wenig ernst genommen. Das heute deutlich vorhandene Bestreben, diesem entgegenzuwirken, ist nicht zuletzt auf den Umstand zurückzuführen, dass hochkomplexe Systeme in zunehmendem Maße Einzug in das alltägliche Leben erlangen. Da es sich bei den hier anzutreffenden Benutzergruppen zumeist um Nicht-Experten handelt, wurde die Optimierung der Mensch-Maschine-Schnittstelle (*Man Machine Interface*, MMI) zur unabdingbaren Notwendigkeit, um die Bedienbarkeit solcher Systeme zu vereinfachen oder überhaupt erst zu ermöglichen.

Als Konsequenz wurden seitens der Industrie die Prioritäten verlagert, so dass einerseits zunehmender Forschungs- und Entwicklungsaufwand zugunsten der Mensch-Maschine-Kommunikation (MMK) betrieben wird, so wie andererseits ein wachsender Anteil der zur Verfügung stehenden Systemressourcen für diesen Zweck bereitgestellt werden. Dieser Wandel erfolgte nicht zuletzt auch aufgrund der Erkenntnis, dass der *Usability*-Aspekt (Gebrauchstauglichkeit) aus Sicht des Benutzers mittlerweile eine durchaus tragende Rolle für die Kaufentscheidung spielt.

Derartige Zusammenhänge treffen in besonders ausgeprägter Weise auf die Kraftfahrzeugdomäne zu, auf die sich die vorliegende Arbeit bezieht. Speziell hier zeichnet sich heute - insbesondere bei Fahrzeugen der Ober- und Luxusklasse - die Tendenz zur stark zunehmenden Integration zahlreicher technischer Einrichtungen ab. Dies sind neben klassischen Geräten (z.B. Audio-Anlage, Bordcomputer etc.) bereits heutzutage Einrichtungen der Kommunikations- und Informationselektronik (z.B. Mobiltelefon/-telefax, Internetdienste), sowie Komfort-, Sicherheits- und Fahrerassistenz-

systeme (z.B. Klimaanlage, elektronische Abstandsüberwachung, Navigationssystem etc.). Als übergreifende Bezeichnung trifft man in diesem Zusammenhang häufig auf den Begriff *Infotainment* (zusammengesetzt aus *Information* und *Entertainment*).

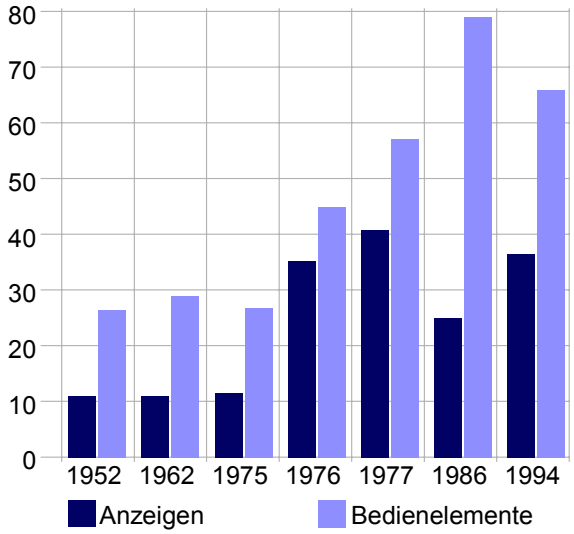


Abb. 1.1: Anstieg der in BMW-Fahrzeugen vorhandenen Anzeige- und Bedienelemente (Quelle: BMW GROUP).

Zu Beginn dieses Techniktrends in den frühen 80er Jahren bestand die Vorgehensweise der Automobilhersteller in der Anhäufung von Einzelkomponenten (wie Radio, Telefon, Navigationssystem etc.), welche meist von verschiedenen Zuliefererfirmen entwickelt wurden. Dies führte einerseits zu einer unüberschaubaren Anzahl an Bedien- und Anzeigeelementen (siehe Abb. 1.1). Andererseits wurde die Bedienbarkeit durch die Koexistenz unterschiedlicher herrstellerabhängiger Bedienkonzepte zusätzlich erschwert. Es wurde somit dringend notwendig, andere Strategien zu verfolgen, um dem Autofahrer die Handhabung trotz stetig wachsender Funktionsumfänge zu ermöglichen. Seit Mitte der 90er Jahre besteht der Trend, die umfangreichen Infotainment-Funktionalitäten in Form eines *Integrierten Bedienkonzepts* bereitzustellen. Hierbei werden Bedien- und Anzeigeelemente in ein einziges MMI integriert, wodurch dem Autofahrer eine einheitliche und anwenderfreundliche Bedienoberfläche geboten werden soll. Üblicherweise dient ein hochauflösendes Farbdisplay als multifunktionale Hauptanzeige. Darüber hinaus zeichnet sich eine Tendenz ab, bei der versucht wird, die Anzahl der haptischen Bedienelemente (Knöpfe, Schalter, Tasten etc.) auf ein Minimum zu reduzieren. Als besonders drastisches Beispiel sei hier das Infotainment-System IDRIVE der BMW GROUP (siehe Kap. 2) genannt, welches nur ein einziges multifunktionales Bedienelement - den sogenannten CONTROLLER - vorsieht. Es handelt sich hierbei um eine Symbiose aus Dreh-/Drücksteller und digitalem Joystick (8 Richtungen), woraus ein Dreh-/Schiebe-/Drücksteller hervorgeht.

1.2 Motivation und Zielsetzung

Wenn auch die oben genannten Entwicklungstrends der Benutzerfreundlichkeit in vielen Aspekten entgegenkommen, werfen sie neue Problematiken auf. So führt die drastische Reduzierung der haptischen Bedienelemente unweigerlich auch dazu, dass einzelne Funktionen, welche zuvor direkt durch eigene Tasten zugänglich waren, nun in entsprechend tief verzweigten Menüstrukturen „ver-

borgen“ werden müssen. Dies wirkt sich durch die mittlerweile enorme Anzahl an implementierten Einzelfunktionen (>700 bei iDRIVE) besonders nachteilig aus.

Zudem leidet die Selbsterklärungsfähigkeit und somit die Intuitivität der Bedienung nicht zuletzt unter der Implementierung zu vieler Freiheitsgrade in ein einziges haptisches Bedienelement (siehe auch iDRIVE, Kap. 2). Dadurch ist es dem Autofahrer im Allgemeinen nicht möglich, allein durch den Anblick des Bedienelements zu erahnen, welche Auswirkungen seine Betätigung hat bzw. in welcher Weise die Betätigung überhaupt zu erfolgen hat. Die Implementierung derartiger Multifunktions-Bedienelemente sollte nach [GEI98] vermieden werden.

Zwar findet man bereits heute erste Anwendungen von kommandoartiger Sprachbedienung im Fahrzeug, die Informations-*Eingabe* erfolgt jedoch nach wie vor hauptsächlich durch haptische Bedienung. Speziell in der Fahrzeugumgebung ergeben sich hieraus weitere Probleme, welche von [FAS98] folgendermaßen beschrieben werden:

Einerseits geht die haptische Bedienung während der Fahrt häufig mit Blickabwendungen vom Straßenverkehr einher, da die Bewegungsabläufe zur Betätigung der Bedienelemente in manchen Fällen visuell beobachtet werden müssen. Andererseits kann eine „motorische Konkurrenz“ auftreten, da der taktile Kanal bereits von der Fahraufgabe selbst stark beansprucht wird.

Zudem kann die haptische Interaktion im Allgemeinen nicht zu den natürlichen Kommunikationsformen des Menschen gezählt werden. Die Realisierbarkeit einer intuitiven Benutzerschnittstelle durch rein haptische Bedienung erscheint daher - zumindest ab einer gewissen Systemkomplexität – prinzipiell fragwürdig.

Schließlich konnte im Rahmen dieser Arbeit belegt werden, dass insbesondere die haptische Bedienung mit extremen Ablenkungseffekten einhergeht. Diese Erkenntnis spricht gegen die übermäßige Nutzung dieser Eingabemethode im Fahrzeug, da Infotainment-Systeme typischerweise auch während der Autofahrt bedient werden (müssen). Es ist jedoch davon auszugehen, dass sich die hohe kognitive Beanspruchung bei der haptischen Bedienung sehr nachteilig auf die Verkehrssicherheit auswirkt.

Betrachtet man hingegen die Informations-*Ausgabe* im Fahrzeug, welche in erster Linie durch optische Anzeigen erfolgt, so ist auch hier Verbesserungsbedarf anzumerken. Die Nachteile der ausschließlich visuellen Informationsdarbietung liegen auf der Hand: So ist der visuelle Kanal des Fahrers bereits stark ausgelastet durch die Verarbeitung des gesehenen Verkehrsgeschehens. Darüber hinaus geht das Ablesen eines Displays zwangsläufig einher mit einer Blickabwendung vom Straßenverkehr und erhöht somit das Unfallrisiko.

Das Hauptanliegen dieser Arbeit besteht nun in der Entwicklung, Umsetzung und Evaluierung eines ganzheitlichen Bedienkonzepts mit der Zielsetzung, den Autofahrer bei der Systembedienung zu entlasten. Eine zentrale Rolle spielt hierbei die Erschließung neuer bzw. derzeit wenig genutzter Eingabekanäle durch die Bereitstellung von Gestik und natürlicher Sprache als Eingabemodalitäten. Darüber hinaus soll die visuelle Informationsausgabe durch den Einsatz akustischer Rückmeldungen - z.B. in Form von Sprachausgabe - weitgehend unterstützt werden.

Der hier verfolgte Ansatz geschieht in Anlehnung an die natürlichen Dialogformen des Menschen. Diese beruhen in erster Linie auf dem Informationsaustausch mittels Sprache (auditiv) und Gestik (visuell).

Während zur Implementierung einer automatischen Spracherkennung auf bereits existierende Systeme zurückgegriffen werden kann, ist die Erkennung dynamischer Hand- und Kopfgesten derzeit noch Gegenstand der Forschung. Die Nutzung des visuellen Kanals hat sich jedoch als intuitive Eingabemodalität erwiesen, die den Mensch-Maschine-Dialog entscheidend verbessern kann [MOR00]. Aus diesem Grunde liegen die Schwerpunkte dieser Arbeit bei der Erforschung der *gestischen* Bedienung im Fahrzeug: Die hierbei gewonnenen Erkenntnisse finden einerseits Anwendung in der Entwicklung eines gestisch bedienbaren Infotainmentsystems sowie andererseits in einer neuartigen Technologie zur fahrzeugtauglichen Gestenerkennung.

Abb. 1.2 zeigt eine schematische Darstellung der heute üblichen Mensch-Maschine-Interaktion im Fahrzeug sowie die in dieser Arbeit angestrebte Bereitstellung natürlicher Informationskanäle. In diesem Zusammenhang wird die Bedeutung der *Berührungslosen Bedienung*, die den Titel dieser Arbeit prägt, anschaulich: Sie ist das gemeinsame Unterscheidungsmerkmal, welches die hier angewandte Gestik- sowie Sprachbedienung von der konventionellen haptischen Bedienung abgrenzt.

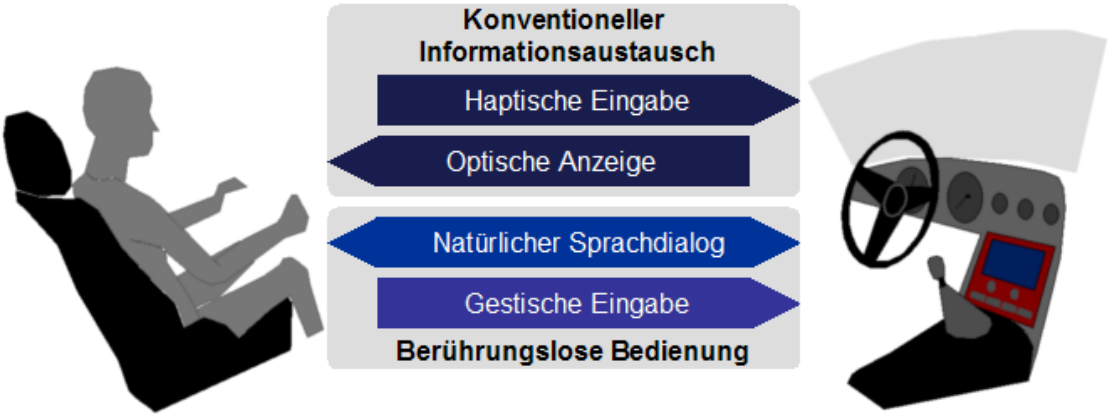


Abb. 1.2: Informationsaustausch im Fahrzeug. Konventioneller Ansatz (Kasten oben) und Erweiterung durch berührungslose Bedienung (Kasten unten).

Durch die Einführung der berührungslosen Bedienung wird der konventionelle Informationsaustausch also erweitert. Man spricht nun von einem sogenannten *Multimodalen System*, da gleichzeitig mehrere Kommunikationskanäle (Modalitäten) des Menschen für die Interaktion genutzt werden. Hierbei wird eine ergonomische Auslastung der menschlichen Kapazitäten angestrebt, wodurch beispielsweise der taktile Kanal entlastet werden soll, der während einer Autofahrt ohnehin stark beansprucht wird.

Die Grundphilosophie des hier verfolgten multimodalen Ansatzes ist die optimale Nutzung der jeweiligen Stärken der Modalitäten Gestik und Sprache.

1.3 Aufbau der Arbeit

Kap. 2 gibt zunächst einen Einblick in den aktuellen Stand von Technik und Forschung bezüglich der hier relevanten Themengebiete. Danach wird in Kap. 3 ein Überblick über das im Rahmen der vorliegenden Arbeit entwickelte Gesamtsystem gegeben, wobei jeweils kurz auf die vorhandenen Teilkomponenten eingegangen wird. Anhand mehrerer Benutzerstudien wird in Kap. 4 auf die Zusammenstellung eines intuitiven Gesten-„Vokabulars“ eingegangen, welches sich für die Bedienung eines Fahrzeug-Infotainmentsystems eignet. In Kap. 5 wird eine umfangreiche Studie zur vergleichenden Untersuchung von kognitiven Ablenkungseffekten bei der gestischen bzw. haptischen Bedienung vorgestellt. Kap. 6 beleuchtet die Entwicklung des multimodalen Infotainmentsystems GECOM unter Anwendung der zuvor beschriebenen Forschungsergebnisse, das anschließend durch Usability-Tests erprobt wird. Es folgt in Kap. 7 die detaillierte technische Beschreibung des im Rahmen dieser Arbeit entwickelten Gestenerkennungssystems zusammen mit einer Evaluierung der Erkennungsleistung. Ein abschließendes Résumé in Kap. 8 beinhaltet eine Diskussion zur durchgeführten Arbeit sowie weiterführende Ausblicke.

2

Stand der Technik

Die vorliegende Arbeit berührt ein weit gefächertes Themenspektrum, das sich von grundsätzlichen Studien zur Erschließung neuer Eingabemethoden über die technische Umsetzung ihrer maschinellen Verarbeitung bis hin zu einer konkreten Realisierung einer adäquaten Fahrzeugbedienumgebung erstreckt. Der nachfolgende Überblick soll und kann daher nicht den Anspruch auf Vollständigkeit erheben, sondern dem Leser einen schnellen Einstieg in die hier relevante Materie ermöglichen. Dazu wird zunächst der aktuelle Stand bereits verfügbarer Technik anhand zweier marktbestimmender Fahrzeug-Infotainmentsysteme erörtert. Weiterführend werden anschließend mehrere Forschungskonzepte beleuchtet, in denen ebenfalls der Ansatz der multimodalen - insbesondere gestischen - Bedienung beschritten wird. Abschließend erfolgt eine Darstellung der wichtigsten Differenzierungsmerkmale, welche den Neuwert der vorliegenden Arbeit unterstreichen sollen.

2.1 Kommerziell verfügbare Infotainmentsysteme

Der heutige Stand der Technik wird im Folgenden anhand zweier Infotainmentsysteme aus dem Fahrzeug-Luxusklassensegment erörtert: iDRIVE des 7er-BMW und MMI[1] im AUDI A8.

Die beiden Bedienkonzepte zeigen in den Grundzügen des Informationsaustauschs zwischen System und Fahrer einige Übereinstimmungen: Zur visuellen Informationsausgabe dient jeweils ein hochauflösendes Farbdisplay (iDRIVE: Bilddiagonale 6,5 Zoll, Format 16:9; MMI: Bilddiagonale 7 Zoll, Format 4:3), welches in beiden Fällen im zentralen oberen Armaturenbereich positioniert ist (siehe Abb. 2.1). Es zeigt sich hier der Trend, Anzeigeelemente möglichst hoch zu platzieren, um die Kopfbewegungen beim Ablesen minimal zu halten.

[1] Die Abkürzung *MMI* steht hier nicht wie üblich für *Man Machine Interface*, sondern für *Multi Media Interface*.

Abb. 2.1: Aktuelle Infotainmentsysteme im Fahrzeug. iDRIVE im 7er-BMW (links) und MMI im AUDI A8 (rechts).

Neben dem Hauptbildschirm existiert ein verhältnismäßig kleines („Kombi“-) Display, welches sich innerhalb der Instrumententafel befindet. Hier wird ein gefilterter Minimalumfang besonders wichtiger Informationen, wie z.B. Navigationshinweise in Pfeildarstellung dargeboten. Diese Strategie der Informationsaufteilung in einen primären sowie einen sekundären Sichtbereich ist mittlerweile ein weitverbreiteter Ansatz, der sich in vielen Fahrzeugkonzepten wiederfindet (siehe auch Abb. 2.2).

Abb. 2.2: Futuristische Designstudie der Firma SIEMENS VDO.

Die Informationseingabe erfolgt in erster Linie über haptische Bedienelemente. Auch hierbei zeigen sich tendenzielle Übereinstimmungen der beiden Konzepte: Mit der Zielsetzung, die Gesamtzahl an Schaltern und Tasten drastisch zu reduzieren, werden multifunktionale Bedienelemente eingesetzt, welche darüber hinaus räumlich von den Anzeigen getrennt platziert sind. Letztere Maßnahme soll dem ergonomischen Aspekt der optimalen Erreichbarkeit Genüge tragen: Gemäß Abb. 2.1 befinden

sich die Haupteingabeelemente innerhalb des Greifraumes, wenn der Arm des Fahrers auf der rechten Armstütze abgelegt wird.

Als zentrales Bedienelement wird ein sogenannter *Dreh-/Drücksteller* eingesetzt (siehe Abb. 2.3). Die Bedienphilosophie ist dabei folgende: Durch Drehen erfolgt die Navigation innerhalb der im Display dargestellten Menüpunkte und zur Bestätigung einer Auswahl wird der Knopf nach unten gedrückt.

Abb. 2.3: Haptische Bedienelemente. iDRIVE (links) und MMI (rechts).

Das AUDI MMI besitzt zusätzlich vier sogenannte *Softkeys*, welche zusammen mit dem Dreh-Drücksteller eine Einheit bilden und kontextabhängig mit verschiedenen Funktionen belegt werden. Durch spezielle Gestaltungskriterien (Design und Platzierung, siehe Abb. 2.3 rechts) soll dem Benutzer der Zusammenhang zwischen Visualisierung und der aktuellen Softkey-Belegung verdeutlicht werden: Das Betätigen einer der vier Tasten aktiviert die im Display in der entsprechenden Ecke dargestellte Funktion (siehe Abb. 2.4 rechts). Eine vergleichbare Maßnahme zur Steigerung der Selbsterklärungsfähigkeit ist bei iDRIVE nicht zu ersehen. Des Weiteren stellt das AUDI MMI acht *Hardkeys* (Tasten mit fester Funktionsbelegung) bereit, die den direkten Zugriff auf Hauptfunktionen wie z.B. Radio, Telefon oder Navigationssystem ermöglichen, sowie eine „RETURN"-Taste. Die Möglichkeit, jegliche Eingaben jederzeit rückgängig machen zu können, wurde bereits in [GEI98] aufgrund intensiver Benutzertests als notwendig befunden. Es sei jedoch an dieser Stelle angemerkt, dass die hier gewählte Bezeichnung „RETURN" für eine Rückgängig-Funktion eher missverständlich erscheint, da Verwechslungen mit der Return-Taste früherer Computerterminals (Bedeutung: Eingabe bestätigen) zu erwarten sind.

Abb. 2.4: Display Gestaltung. iDRIVE (links) und MMI (rechts).

Hinsichtlich der Reduzierung haptischer Bedienelemente geht BMW mit dem iDRIVE-Konzept einen deutlich drastischeren Weg: Neben dem Haupteingabegerät, dem sogenannten CONTROLLER, existieren keine weiteren Bedienelemente. Stattdessen wurden hier weitere Freiheitsgrade imple-

mentiert: Der Controller kann - neben dem erwähnten Dreh-Drück-Konzept - in acht Richtungen (orthogonal und diagonal) translatorisch bewegt bzw. verschoben werden. Auf diese Weise lassen sich je nach Kontext verschiedene Funktionen aufrufen. Es handelt sich hierbei also ebenfalls um eine Softkey-Variante, wobei diese Bedienmöglichkeit durch reines Betrachten des CONTROLLERS nicht ersichtlich ist und einen klaren Mangel an Selbsterklärungsfähigkeit darstellt. Eine weitere Besonderheit des CONTROLLERS liegt in seiner variablen Haptik (*Force Feedback*): Es handelt sich um ein aktives Bedienelement, dessen haptischen Eigenschaften an den jeweiligen Kontext angepasst werden. Durch einen integrierten Elektromotor wird das zum Drehen erforderliche Moment dynamisch variiert. Dies erlaubt es beispielsweise, die Anzahl der Rasterstufen (pro Umdrehung) an die Anzahl der aktuell angezeigten Menüpunkte anzupassen, die Realisierung von Endanschlägen beim Erreichen des Anfangs bzw. des Endes einer Liste, sowie die Simulation eines Rastwerks mit Mittelstellung und Zentrierfeder.

2.1.1 Sprachbedienung

Beide betrachteten Systeme können auf Kundenwunsch (Sonderausstattung) mit einem Spracherkennungsmodul ausgestattet werden. Dieses ist auf die Verarbeitung isolierter Kommandos (Einzelworterkennung[2]) ausgelegt und wird per Tastendruck (*Push-To-Talk-Taste* am Lenkrad, PTT) aktiviert. Auf diese Weise lassen sich etwa bei IDRIVE manche Bedienabfolgen per Sprache als Alternative zur Haptik bedienen. Es werden dem Benutzer jedoch nur Teilbereiche des Gesamtumfangs via Spracheingabe zugänglich gemacht. Das implementierte Vokabular umfasst etwa 270 Kommandowörter. Die Bedienlogik der Sprachsteuerung existiert parallel zum Strukturbaum des haptisch bedienbaren Funktionsumfanges. Dies bedeutet, dass der Benutzer nicht etwa die auf dem Hauptdisplay dargestellte Menüstruktur sprachlich bedienen kann, sondern hierfür einer eigenen Bedienlogik folgen muss, welche lediglich auf dem kleinen Kombi-Display visualisiert wird. Für die Benutzerfreundlichkeit bedeutet dies ein schwerwiegendes Manko, welches sich in vergleichbarer Weise auch im MMI wiederfindet. Ein sogenanntes *Speak-What-You-See*-Konzept würde hier Abhilfe schaffen. Hierfür ist die konsistente Einbettung der Sprachsteuerung in die gesamte Bedienumgebung eine Grundvoraussetzung.

2.2 Infotainmentsysteme im Forschungsstadium

Bei der Optimierung der Mensch-Maschine-Schnittstelle im Fahrzeug handelt es sich um eine Thematik, die gerade in jüngster Zeit zum Gegenstand zahlreicher Forschungsarbeiten wurde. Nachfolgend wird der Fokus auf eine Auswahl an Konzepten gerichtet, welche die Nutzung der Gestik als Eingabemodalität vorsehen.

In [CAI00A] wird ein Bedienkonzept zur gestischen Steuerung sekundärer Fahrzeugfunktionen vorgestellt. Dabei handelt es sich unter anderem ebenfalls um die hier betrachteten Infotainment-Umfänge. Der Fahrer aktiviert bestimmte Funktionen, indem er mit dem Zeigefinger auf beschriftete Buttons zeigt, die auf einem Display dargestellt werden. Zur Ermittlung des intendierten Zeige-

[2] Eine Ausnahme bildet die Eingabe von Telefonnummern, bei der neben einzelnen Ziffern auch komplette Zahlenblöcke zusammenhängend ausgesprochen werden können.

Ziels wird die Lage der Hand im Raum mit einem sogenannten *6DOF-Sensor* (*Six-Degree-Of-Freedom*) ermittelt. Eine derartige Realisierung ist für den kommerziellen Einsatz im Fahrzeug jedoch ungeeignet, da dem Fahrer nicht zugemutet werden kann, eine spezielle Apparatur am Körper befestigen zu müssen.

Einen ähnlichen Ansatz verfolgt das IWAVE-Konzept [ALP02], wobei es sich hier lediglich um eine Bedienumgebung ohne technische Realisierung für die Erkennung der Zeigegesten handelt. Zur Visualisierung ist ein *Head-Up-Display* (HUD) vorgesehen, d.h., die Bedienoberfläche wird in die Frontscheibe projiziert (siehe Abb. 2.5 rechts).

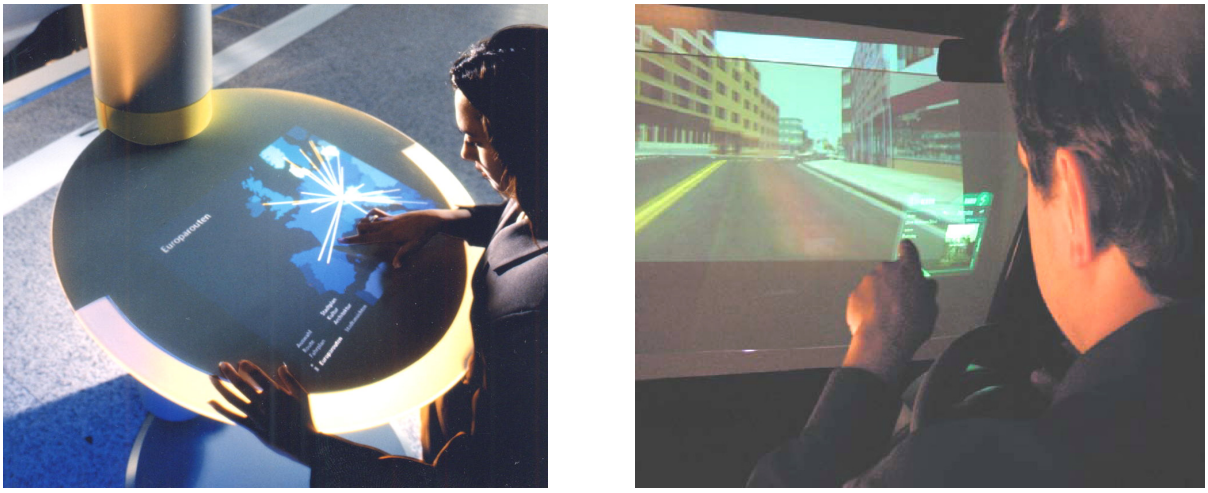


Abb. 2.5: Statische Erkennung der Zeigefinger-Position. SIVIT (SIEMENS VIRTUAL TOUCHSCREEN, links) und IWAVE-Konzept (rechts).

Die Erkennung von Zeigegesten im Sinne einer Auswertung der statischen Position des Zeigefingers wird von der SIEMENS AG mit dem SIVIT (SIEMENS VIRTUAL TOUCHSCREEN) als kommerzielles Produkt angeboten (siehe Abb. 2.5 links). Der Gestenraum wird hierbei von einer Videokamera erfasst. Durch die Auswertung der Handkontur wird die Position der Zeigefingerspitze bestimmt. Die Bedienoberfläche wird mittels eines Beamers auf eine horizontale Fläche projiziert und die Auswahl eines dort dargestellten Menüpunktes erfolgt, indem der Benutzer für eine bestimmte Zeit mit dem Zeigefinger an der entsprechenden Position verweilt.

Von der Nutzung derartiger Zeigegesten zur Adressierung von Intentionszielen wird in der vorliegenden Arbeit bewusst abgesehen, da sie mit erheblichen Ablenkungseffekten einhergeht (siehe Kap. 5) und daher zur haptischen Bedienung keinen Mehrwert aufweist.

2.3 Gestenerkennung im Fahrzeug

Der reale Einsatz einer automatischen Gestenerkennung im Fahrzeug ist mit der Anbringung spezieller Gerätschaften oder gar Farbmarken am Körper des Fahrers prinzipiell unvereinbar. Bisherige Ansätze, die dennoch derartige Maßnahmen beinhalten, werden in dieser Arbeit nicht weiter betrachtet.

[ZOB03A] beschreibt ein videobasiertes Gestenerkennungssystem, welches aufbauend auf den Arbeiten von [MOR00] für den Einsatz im Fahrzeug erweitert wurde. Hierbei wird der Gestenraum von einer konventionellen CCD-Kamera erfasst, die sich am Fahrzeughimmel befindet und auf die Mittelkonsole gerichtet ist. Zur Reduzierung störender Einflüsse von externen Lichtquellen ist die Kamera einerseits mit einer Tageslicht-Sperrfilter-Scheibe ausgestattet, während die Bildszene andererseits von mehreren Infrarot-LED-Arrays ($\lambda = 950$ nm) ausgeleuchtet wird. Die Einzelbilderfassung erfolgt mit 25 fps (*frames per second*) bei einer Auflösung von 384×144 Pixel. Damit eine Hand vom Szenenhintergrund zur weiteren Verarbeitung getrennt werden kann, wird folgendes Wissen angewandt: Die Hand ist ein vergleichsweise großes Objekt im Bild, das aufgrund der starken IR-Beleuchtung eine hohe Helligkeit besitzt. Die einzelnen Verarbeitungsschritte der örtlichen Segmentierung gehen aus Abb. 2.6 hervor.

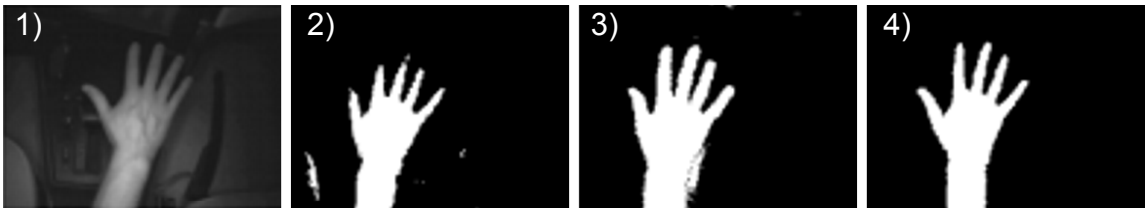


Abb. 2.6: Segmentierung der Hand nach [ZOB03A]: 1) Originalbild, 2) adaptive Schwellwertbildung, 3) Hintergrundausblendung und 4) segmentierte Hand aus Kombination von 2) und 3).

Aus jedem vorverarbeiteten Einzelbild wird ein Merkmalvektor berechnet, dessen Komponenten - Fläche, Schwerpunkt und HU-Momente - die segmentierte Handform beschreiben. Eine dynamische Handgeste besteht somit aus einer Sequenz derartiger Merkmalvektoren. Zur Gestenklassifikation werden semikontinuierliche *Hidden-Markov-Modelle* (HMMs) eingesetzt. Das beschriebene System wurde zu Demonstrationszwecken an die Bedienumgebung GECOM angebunden, welche im Rahmen dieser Arbeit entwickelt wurde (siehe Kap. 6). Dabei erstreckte sich das implementierte Gesteninventar über 22 dynamische sowie 6 statische Handgesten (d.h. Handformen).

Eine der Mindestvoraussetzungen für die robuste Erkennung dynamischer Handgesten besteht offensichtlich in einer zuverlässigen Modellierung statischer Handformen, deren automatische Erkennung am Rande auch in [ZOB03A] besprochen wird. Wesentlich ausführlicher widmet sich [AKY00] der Erkennung ausschließlich statischer Gesten zur Bedienung eines „Nachrichten-Containers“ im Fahrzeug. Dieser enthält gespeicherte Tonkonserven, wie z.B. Verkehrsnachrichten, deren Wiedergabe unter Verwendung der in Abb. 2.7 dargestellten Handformen gesteuert wird. Inwiefern diese Bedienung mit *statischen* Gesten einem natürlicheren Mensch-Maschine-Dialog beizutragen vermag, wird dort jedoch nicht untersucht.

Abb. 2.7: Statische Handgesten und zugeordnete Funktionen von [AKY00].

Hinsichtlich ihrer Fahrzeugtauglichkeit ergeben sich für beide der angesprochenen Systeme einige Probleme, die in Kap. 7 ausführlich diskutiert werden.

2.3.1 Neuwert der vorliegenden Arbeit

Die vorliegende Arbeit befasst sich in erster Linie mit der Bereitstellung der Eingabemodalität *Gestik* im Fahrzeug. Im Gegensatz zu anderen Forschungsarbeiten mit ähnlichen Zielsetzungen besteht hier das Hauptanliegen darin, zunächst die Gebrauchstauglichkeit dieser neuen Bedienmethode objektiv zu beleuchten, um die gewonnenen Erkenntnisse anschließend vorteilhaft in einem Gesamtsystem zu realisieren.

Insbesondere die im Fahrzeugeinsatz hochrelevanten Aspekte der modalitätenspezifischen Ablenkungseffekte wurden in dieser Arbeit bereits frühzeitig detailliert untersucht. Die resultierenden Befunde belegen erstmals signifikante Vorteile der gestischen Bedienung gegenüber der haptischen.

Die Entwicklung einer gestenoptimierten Bedienumgebung basiert ebenfalls auf ausführlichen Usability-Studien, die schritthaltend zur Umsetzung durchgeführt wurden. Diese spezielle Vorgehensweise differenziert das resultierende Gesamtsystem von vielen anderen Arbeiten. Die starke Beeinflussbarkeit des gestischen Interaktionsverhaltens von der Darbietungsweise der Bedienumgebung zeichnete sich bereits in grundlegenden Voruntersuchungen ab und verdeutlicht umso mehr die Wichtigkeit entwicklungsbegleitender Benutzertests. Die Innovation liegt hierbei weniger in der *Realisierung* einer gestisch bedienbaren Anwendung als vielmehr in deren *Optimierung* auf genau diese Eingabemodalität.

Schließlich wird ein einsatzfähiges System zur automatischen Erkennung dynamischer Hand- und Kopfgesten vorgestellt. Dabei liegt der Hauptschwerpunkt auf dem potenziellen Einsatz im Fahrzeug unter Einhaltung der hier gegebenen schwierigen Randbedingungen. Es resultiert eine neuartige, nicht videobasierte Technologie, welche eine robuste Gestenerkennung erlaubt und erstmals ohne aufwändige Hard- und Software für Bildverarbeitungsprozesse auskommt.

3

Systemüberblick

3.1 Aufbau

Abb. 3.1 enthält eine schematische Darstellung des im Rahmen der vorliegenden Arbeit entwickelten Gesamtsystems. Die zentrale Komponente ist die gestenoptimierte Bedienumgebung GECOM (*Ge*sture *Co*ntrolled *M*MI). Es handelt sich hierbei um ein Fahrzeug-MMI, dessen Funktionsumfang sich über domänentypische Infotainmentkomponenten (Audio, Telefon, Navigation etc.) erstreckt.

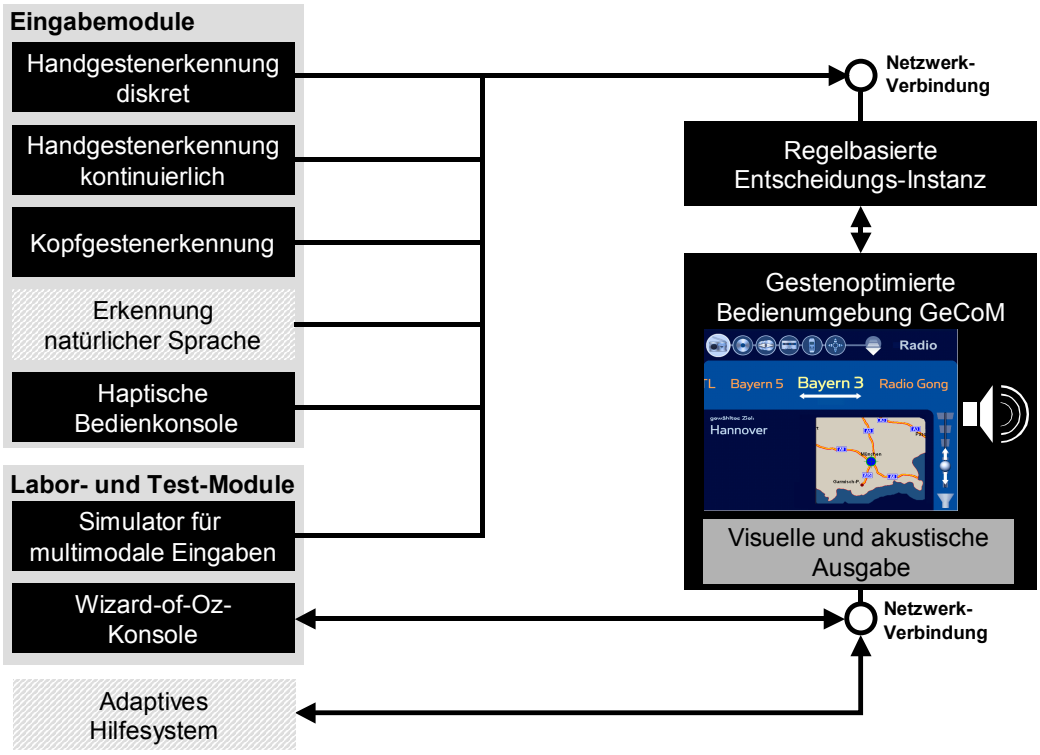


Abb. 3.1: Schematischer Aufbau des Gesamtsystems.

Die Kommunikation zwischen GECOM und externen Komponenten erfolgt über Netzwerkverbindungen (*TCP/IP Sockets*). Diese verteilte Architektur ermöglicht einerseits die flexible Anbindung weiterer Module und bietet andererseits die Möglichkeit einer optimalen Verteilung der benötigten Einzelkomponenten auf verschiedene Rechner. Hierbei handelt es sich hauptsächlich um Eingabemodule, die dem Benutzer zur Interaktion mit GECOM bereit stehen. Darüber hinaus existieren Labor- bzw. Testmodule, die zur Durchführung von Benutzerstudien oder zu Diagnosezwecken benötigt werden. Des Weiteren kommt ein adaptives Hilfesystem zum Einsatz, welches dem Benutzer unter Berücksichtigung von dessen aktuellem Wissensstand kontextabhängige Hilfeinformationen bereitstellt.

Die beiden in Abb. 3.1 schraffiert hinterlegten Komponenten wurden nicht im Rahmen dieser Arbeit entwickelt, sondern lediglich in das Gesamtsystem integriert. Nähere Informationen hierzu sind [HOF01] (Erkennung natürlicher Sprache) und [NIE01] (Adaptives Hilfesystem) zu entnehmen.

3.2 Gestenoptimierte Bedienumgebung GECOM

Das prototypische Infotainmentsystem GECOM enthält den Zustandsautomaten zur Simulation der implementierten Geräte (Radio, CD-Spieler, Telefon, Navigationssystem etc.) sowie das Interface zur Steuerung der Netzwerkfunktionalitäten und ist somit die Kernkomponente des Gesamtsystems. Als zentrales Ausgabegerät für die grafische Bedienoberfläche dient ein TFT-Farbdisplay (etwa 10 Zoll bzw. 25 cm Bilddiagonale). Darüber hinaus wird neben dem visuellen auch der auditive Kanal für die Informationsausgabe genutzt. Das akustische Feedback umfasst sowohl Sprachausgaben als auch Signaltöne und spielt eine wichtige Rolle für die im Fahrzeug angestrebte *Blindbedienbarkeit* (siehe auch [GEI98], S. 17).

Sowohl die Entwicklung der Bedienlogik als auch die Gestaltung der Bedienoberfläche erfolgten mit der Zielsetzung, dem Benutzer den gesamten Funktionsumfang durch rein gestische Bedienung zugänglich zu machen. Durch die zusätzliche Anbindung einer sprachverstehenden Komponente wird GECOM zum multimodalen Bediensystem erweitert. Die Wahl der bevorzugten Eingabemodalität ist dem Benutzer hierbei freistellt, so dass er von den jeweiligen Stärken der vorhandenen Eingabekanäle profitieren kann (aber nicht muss).

Um dem Benutzer bei Bedarf Informationen - insbesondere zur noch ungewohnten Gestenbedienung - bereitzustellen, wurde ein audiovisuelles Hilfesystem (siehe Kap. 6.4.4) implementiert. Dabei wird die Ausführung einer Geste einerseits auf dem Display visualisiert, während diese andererseits durch Sprachausgabe erklärt wird.

Da das Interaktionsverhalten des Anwenders z.B. bei der Auswertung von Benutzertests von großem Interesse ist, besteht die Möglichkeit, komplette Bedienszenarien zeitgenau in Form von *Log-Dateien* abzuspeichern, wodurch sie jederzeit reproduziert werden können.

Das Bediensystem GECOM fungiert bei der TCP/IP-Kommunikation als zentraler *Server*, an den sich externe Module als *Clients* anmelden können. Die Implementierung erfolgte in der Skriptsprache Tcl/Tk (Version 8.3). Zur näheren Erläuterung der gestenoptimierten Bedienumgebung GECOM siehe Kap. 6.

3.3 Eingabemodule

Das Hauptanliegen dieser Arbeit besteht darin, dem Autofahrer die Bedienung von Geräten im Fahrzeug durch den Einsatz natürlicher Kommunikationsformen zu erleichtern. Eine zentrale Rolle spielt daher die Bereitstellung der entsprechenden Eingabekanäle.

3.3.1 Hand- und Kopfgestenerkennung

Zur Bereitstellung des visuellen Eingabekanals wurde eine neuartige Technologie zur Gestenerkennung entwickelt (siehe Kap. 7). Dabei wurden typische fahrzeugspezifische Randbedingungen berücksichtigt (Robustheit, Wirtschaftlichkeit etc.). Das Grundprinzip beruht auf der Gewinnung von räumlicher Information mittels Infrarot-Distanz-Mess-Sensorik sowie deren Klassifikation unter Verwendung eines klassischen Mustererkennungsverfahrens. Das entwickelte Verfahren erweist sich als besonders leistungsfähig bei der automatischen Erkennung von Teilkörpergestik im Fahrzeug.

Aufgrund der Ergebnisse umfangreicher Benutzerstudien (siehe [Zob01] und [Zob02]) konnte ein intuitives „Vokabular“ an *diskreten* Handgesten extrahiert werden. Dieses Gesteninventar wird dem Benutzer mit der Implementierung der automatischen Handgestenerkennung nun tatsächlich zur Gerätebedienung bereitgestellt. Bei dieser Art der Informationseingabe handelt es sich um die sogenannte *indirekte Manipulation*, d.h., die Ausführung *eines* abgeschlossenen Bewegungsablaufs (diskrete Geste) hat genau *eine* bestimmte Systemreaktion zur Folge. Darüber hinaus wird auch die *kontinuierliche* Handgestik zur *direkten Manipulation* bereitgestellt. Diese erweist sich z.B. für das stufenlose Einstellen der Musiklautstärke durch kontinuierliche Handbewegungen als komfortable Eingabeform. Für nähere Erläuterungen zu den verschiedenen Gestentypen (diskret, kontinuierlich etc.) siehe Kap. 4.1.

Die Implementierung eines Kopfgestenerkenners ermöglicht das intuitive Beantworten intentionaler Systemrückfragen (Entscheidungsfragen wie z.B.: „Soll die Zielführung gestartet werden?“) durch Kopfnicken bzw. -schütteln. Zusätzlich ist das System in der Lage, typische Blickbewegungen, die nicht in kommunikativer Absicht erfolgen, explizit als solche zu erkennen.

Eine detaillierte Beschreibung der implementierten Gestenerkennungssysteme erfolgt in Kap. 7.

3.3.2 Erkennung natürlicher Sprache

Für den direkten Zugriff auf nahezu alle Systemfunktionen (im Sinne von *shortcuts*) wurde das sprachverstehende System INSENSE eingebunden, welches im Rahmen der Arbeiten von [HOF00] und [HOF01] entwickelt wurde. Dem Benutzer wird dadurch die Möglichkeit geboten, spontansprachliche Eingaben - im Idealfall ohne vorangehende Lernphase - zu tätigen. Das System INSENSE zeichnet sich insbesondere durch seine hohe Robustheit bei *Out-of-Vocabulary-Fällen* aus (Auftreten von Wortäußerungen, die nicht im Vokabular des Spracherkenners enthalten sind). Darüber hinaus ist es in der Lage, mehrere Systemparameter aus umfangreichen Benutzeräußerungen zu extrahieren. Dazu ein typisches Beispiel: „Ich möchte jetzt bitte mal von der vierten CD das dritte Lied hören und zwar - äh - bei mittlerer Lautstärke“. Selbst für derart umgangssprachliche For-

mulierungen, die sich für existierende Spracherkennungssysteme meist als überaus problematisch erweisen, gibt [HOF03] Erkennungsraten von bis zu 91 % an.

3.3.3 Haptische Bedienkonsole

Zunächst widerspricht das Vorhandensein einer haptischen Bedienkonsole der Grundphilosophie der vorliegenden Arbeit, da gerade diese Art der Bedienung *nicht* berührungslos erfolgt. Die Implementierung der konventionellen haptischen Eingabemethode neben Gesten- und Sprachbedienung erfolgte lediglich zu Vergleichszwecken für einige Benutzertests (siehe z.B. Kap. 5). Die haptische Bedienkonsole besteht aus einer Anordnung festbelegter Funktionstasten (*Hardkeys*) für den direkten Zugriff auf die vorhandenen Infotainment-Geräte. Darüber hinaus existieren zwei Dreh-/Drücksteller, welche der Menüsteuerung bzw. der Lautstärkeregelung dienen (siehe Abb. 3.2).

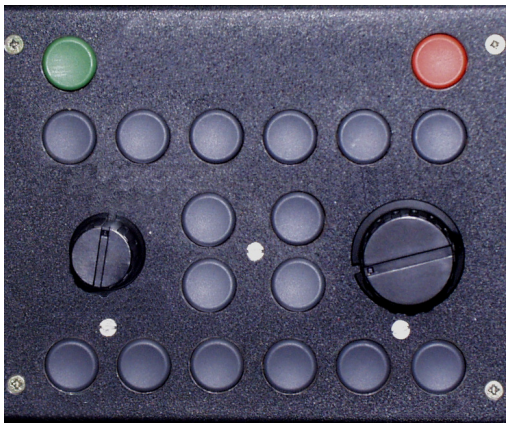


Abb. 3.2: Haptische Bedienkonsole (Tasten hier unbeschriftet).

3.4 Labor- und Test-Module

3.4.1 Wizard-of-Oz-Konsole

Die im Rahmen dieser Arbeit entwickelte Bedienumgebung wurde in diversen *Usability-Tests* evaluiert und sukzessive verbessert. Eine zentrale Rolle spielte dabei die Anwendung der sogenannten *Wizard-of-Oz*-Methode nach [NIE94]. Hierbei beobachtet ein Versuchsleiter, der *Wizard*, eine Versuchsperson bei der Bedienung eines technischen Gerätes und simuliert dabei das gesamte Systemverhalten oder zumindest einige seiner Teilkomponenten. Die Versuchsperson gewinnt dadurch den Eindruck, mit einem vollständig funktionstüchtigen System konfrontiert zu sein.

In der vorliegenden Arbeit bestand die Aufgabe des Wizards in erster Linie darin, automatische Erkennungssysteme (Gesten- und Spracherkennung) zu ersetzen bzw. zu simulieren. Dies war einerseits notwendig, da die entsprechenden Systeme zu Beginn der Arbeiten größtenteils noch nicht existierten. Die durchgeführten Usability-Studien lieferten daher wichtige Erkenntnisse z.B. über die Anforderungen an einen realen Gestenerkenner im Fahrzeug. Andererseits treten bei der Anwendung der Wizard-of-Oz-Methode keine Fehlerkennungen auf, wie sie für reale Mustererkennungssysteme typisch sind. Dieser Aspekt ist besonders wichtig, da Fehlerkennungen aus Sicht des Benutzers kaum nachvollziehbar sind und dessen Interaktionsverhalten daher auf unerwünschte sowie unvorhersehbare Weise beeinflussen können. Um jedoch auch derartige Benutzerreaktionen

explizit zu untersuchen, werden bei der Wizard-of-Oz-Methode absichtliche Fehlerkennungen gezielt eingesetzt, die im Gegensatz zu realen Fehlerkennungen reproduzierbar sind (siehe auch [NIE02A]).

Der Wizard muss also absolute Kontrolle über das System GECOM besitzen, um dessen Reaktionen entsprechend des beobachteten Benutzerverhaltens beeinflussen zu können. Dazu stehen ihm mehrere Eingabekanäle zur Verfügung, welche hier mit dem Begriff *Wizard-of-Oz-Konsole* zusammengefasst und im Folgenden erläutert werden.

Grafische Bedienoberfläche WIZCON

Abb. 3.3 zeigt die grafische Bedienoberfläche WIZCON (*Wi*zard *Con*sole) zur haptischen Fernsteuerung sämtlicher Funktionalitäten der Bedienumgebung GECOM durch den Wizard.

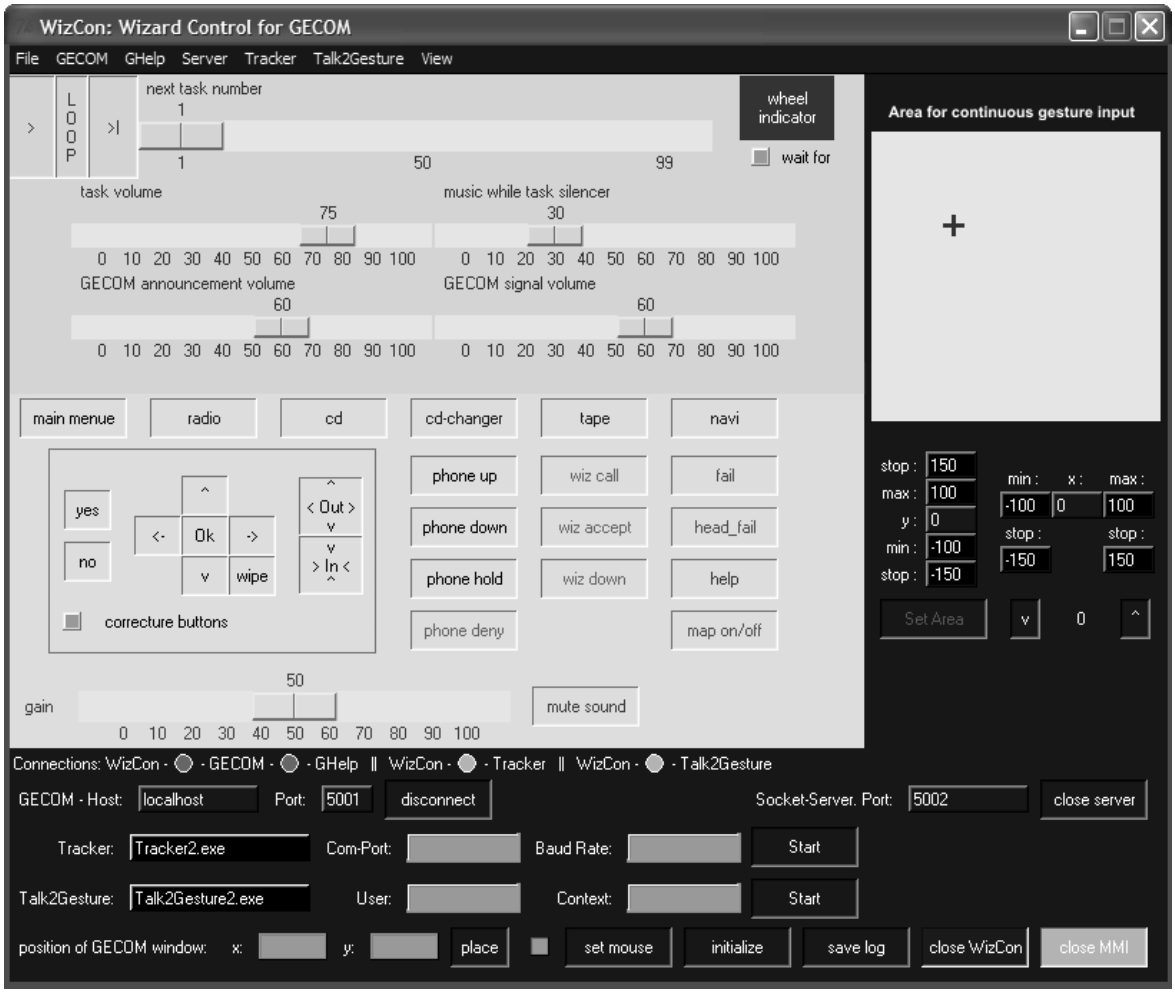


Abb. 3.3: Grafische Bedienoberfläche WIZCON zur Fernsteuerung von GECOM.

Die Ansteuerung der GECOM-Geräte erfolgt über Tastenbedienung. Darüber hinaus können mit WIZCON auch simulierte gestische Eingaben übermittelt werden. Mehrere Schieberegler dienen zur Einstellung diverser Lautstärken (z.B. Lautstärke für Audiowiedergabe, Systemansagen, Signaltöne etc.). Des Weiteren existieren diverse Anzeigen, welche über den Netzwerk-Verbindungsstatus externer Module sowie deren Systemzustände informieren.

Ein spezielles Eingabefeld dient der Simulation von kontinuierlichen Gesten (siehe Kap. 4.1). Dabei werden x- und y-Position eines angezeigten Cursors zyklisch an GECOM gesendet. Dort werden sie als aktuelle Handposition des Benutzers interpretiert. Der Wizard kann diesen Cursor mittels Joystick oder Mouse innerhalb des Eingabefeldes bewegen und dadurch z.B. horizontale oder vertikale Handbewegungen einer Versuchsperson simulieren.

Ferner enthält die Wizard-Oberfläche Bedien- und Anzeigeelemente für eine integrierte Versuchsablaufsteuerung (siehe auch [GEI01B]). Die Bedienaufgaben, welche einer Versuchsperson im Laufe eines Usability-Experiments gestellt werden, liegen als abgespeicherte Ton-Dateien vor, und können auf Knopfdruck in der vorgegebenen Reihenfolge abgespielt bzw. bei Bedarf jederzeit wiederholt werden. Diese Methodik gewährleistet die Einhaltung identischer Rahmenbedingungen für jede Versuchsperson und erleichtert es dem Wizard, den Überblick über den Versuchsfortschritt zu behalten. Unter realen Versuchsbedingungen werden die Kapazitäten des Wizard meist stark beansprucht: Er muss die Versuchsperson permanent via Monitor beobachten und deren Aktionen in entsprechende Systemreaktionen umsetzen. Dabei ist die zeitliche Verzögerung zwischen Benutzeraktion und Systemreaktion möglichst minimal zu halten, da ansonsten mit abweichendem Verhalten der Versuchsperson zu rechnen ist (z.B. Kontrollblicke zum Display bei ausbleibender Systemreaktion). Die oben beschriebene Art der haptischen Versuchssteuerung über die grafische Bedienkonsole ist daher ungeeignet, da hierfür einerseits Blickabwendungen vom Versuchsgeschehen nötig sind und andererseits kein direktes Systemfeedback gewährleistet ist. Sie wird während laufender Versuche lediglich für Eingaben in unvorhersehbaren „Notfällen“ verwendet und ansonsten zu Testzwecken eingesetzt. Zur Einhaltung der genannten Anforderungen (permanente Beobachtung der Versuchsperson *und* echtzeitnahe Systemreaktionen) werden dem Wizard weitere Eingabekanäle bereitgestellt:

Wizard-Sprachsteuerung

Um dem Wizard die Umsetzung der beobachteten Aktionen ohne Blickabwendungen vom Versuchsszenario zu ermöglichen, bot sich die Anbindung einer Sprachsteuerung an. Daher wurde das Spracherkennungssystem WISPER (*Wiz*ard *Spe*ech *R*ecognition; siehe [GEI01B] und [NIE02B]) entwickelt. Es handelt sich hierbei um einen Einzelwort-Erkenner, der durch Tastendruck (*Push-To-Talk*, PTT) aktiviert wird und erkannte Kommandos an die Wizard-Konsole übergibt. Der Wizard „übersetzt“ also beobachtete Aktionen (z.B. Gesten) in Sprachkommandos. Hierzu ein Beispiel: Eine Versuchsperson führt eine Winkbewegung nach oben aus, um die Musiklautstärke zu erhöhen. Der Wizard beobachtet dies, betätigt die PTT-Taste und spricht das Kommando „hoch“. WISPER erkennt das Kommando und veranlasst die Wizard-Konsole, einen entsprechenden Systembefehl an GECOM zu senden. Die Reaktion von GECOM ist dabei identisch mit jener, die nach Knopfdruck der entsprechenden Taste (Pfeil nach oben) in der grafischen Wizard-Bedienoberfläche erfolgt wäre. Diese Methodik für Wizard-Eingaben bewährte sich bei ihrer Anwendung in zahlreichen Benutzerstudien.

3D Tracker und Datenhandschuh

Eine besondere Schwierigkeit ergibt sich bei der Untersuchung kontinuierlicher Gesten. Hierbei müssen Handpositionen bzw. -bewegungen erfasst und an das Bediensystem übertragen werden. Für die Simulation kontinuierlicher Gesten erwies sich die oben genannte Joystickbedienung via

Eingabefeld als ungeeignet, da es kaum möglich ist, die kontinuierlichen Handbewegungen der Versuchsperson bei gleichzeitiger Mitführung des Cursors im Eingabefeld zu beobachten. Es wurde daher eine zweckdienlichere Eingabemethode entwickelt, welche in zwei verschiedenen Varianten angewandt wurde.

Der erste Ansatz bestand darin, die räumliche Position der rechten Hand des Wizards unter Verwendung eines elektromagnetischen 3D-Trackers permanent zu ermitteln (siehe auch Kap. 4.3). Die Simulation kontinuierlicher Gesten erfolgte nun, indem der Wizard die beobachteten Handbewegungen der Versuchsperson mit seiner rechten Hand nachvollzog bzw. imitierte. Die Daten des Trackers wurden vorverarbeitet und an die gestisch gesteuerte Applikation weitergeleitet. Die Versuchsperson sollte dadurch den Eindruck gewinnen, das System direkt durch kontinuierliche Handbewegungen manipulieren zu können. Der Wizard hatte zusätzlich die Aufgabe, zu entscheiden, ob die Versuchsperson entweder *diskrete* Einzelgesten ausführt oder die *kontinuierliche* Steuerung beabsichtigt. Je nach vermuteter Absicht der Versuchsperson wurde das System durch den Wizard in den entsprechenden Modus (diskret bzw. kontinuierlich) versetzt. Dieser Moduswechsel erfolgte durch die Auswertung der Handstellung des Wizards mittels eines Datenhandschuhs (Erkennung statischer Handgesten). Bei dieser Vorgehensweise erwies es sich als nachteilig, dass die Systemantwortzeit zwangsläufig mit der Reaktionszeit des Wizards behaftet ist. Von vielen Versuchspersonen wurde das System genau diesbezüglich bemängelt, da das Feedback bei kontinuierlichen Handbewegungen teilweise nicht direkt genug erfolgte.

Dieses Manko sollte in einem zweiten Ansatz vermieden werden. Um ein direktes Systemfeedback gewährleisten zu können, wurde der „Umweg“ über den Wizard vermieden, indem der 3D-Tracker an der Hand der Versuchsperson selbst befestigt wurde (siehe Abb. 3.4 rechts).

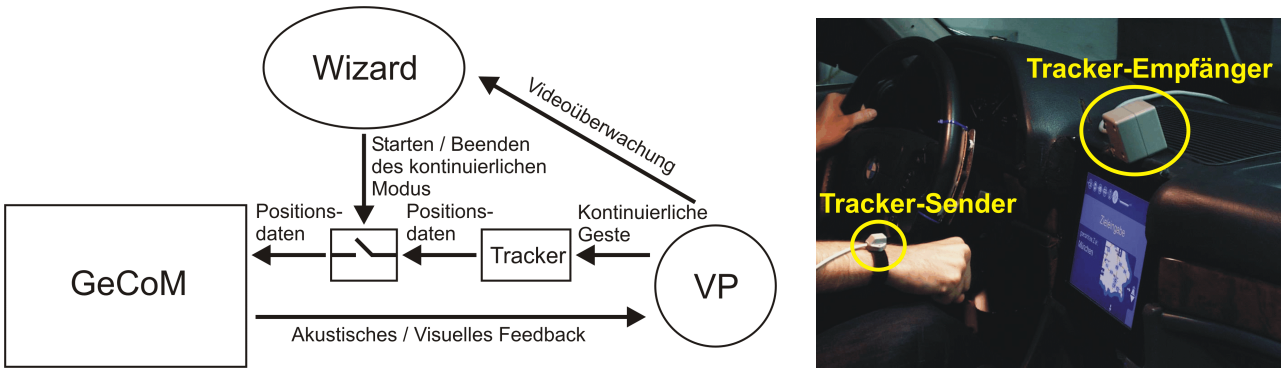


Abb. 3.4: Schematische Darstellung für die Eingabe kontinuierlicher Gesten (links) und Positionierung der Tracker-Komponenten (rechts).

Die aktuellen Positionsdaten werden dabei permanent an die Wizard-Konsole gesendet. Die Aufgabe des Wizards besteht in diesem Fall lediglich darin, zu erkennen, wann die Versuchsperson beabsichtigt, kontinuierliche Gestik anzuwenden. In diesem Fall aktiviert er durch Knopfdruck den kontinuierlichen Modus, wodurch die Positionsdaten an die Bedienumgebung GECOM weitergeleitet werden und dort z.B. eine stufenlose Änderung der Musiklautstärke bewirken. Entsprechend beendet der Wizard diesen Modus, wenn er das Ende einer kontinuierlichen Eingabe erkennt; er steuert also lediglich den Datenfluss der kontinuierlichen Positionsdaten (siehe Abb. 3.4 links). Diese Vor-

gehensweise erbrachte letztlich das erwünschte direkte Systemverhalten bei kontinuierlichen Bedienung.

3.4.2 Simulator für multimodale Eingaben

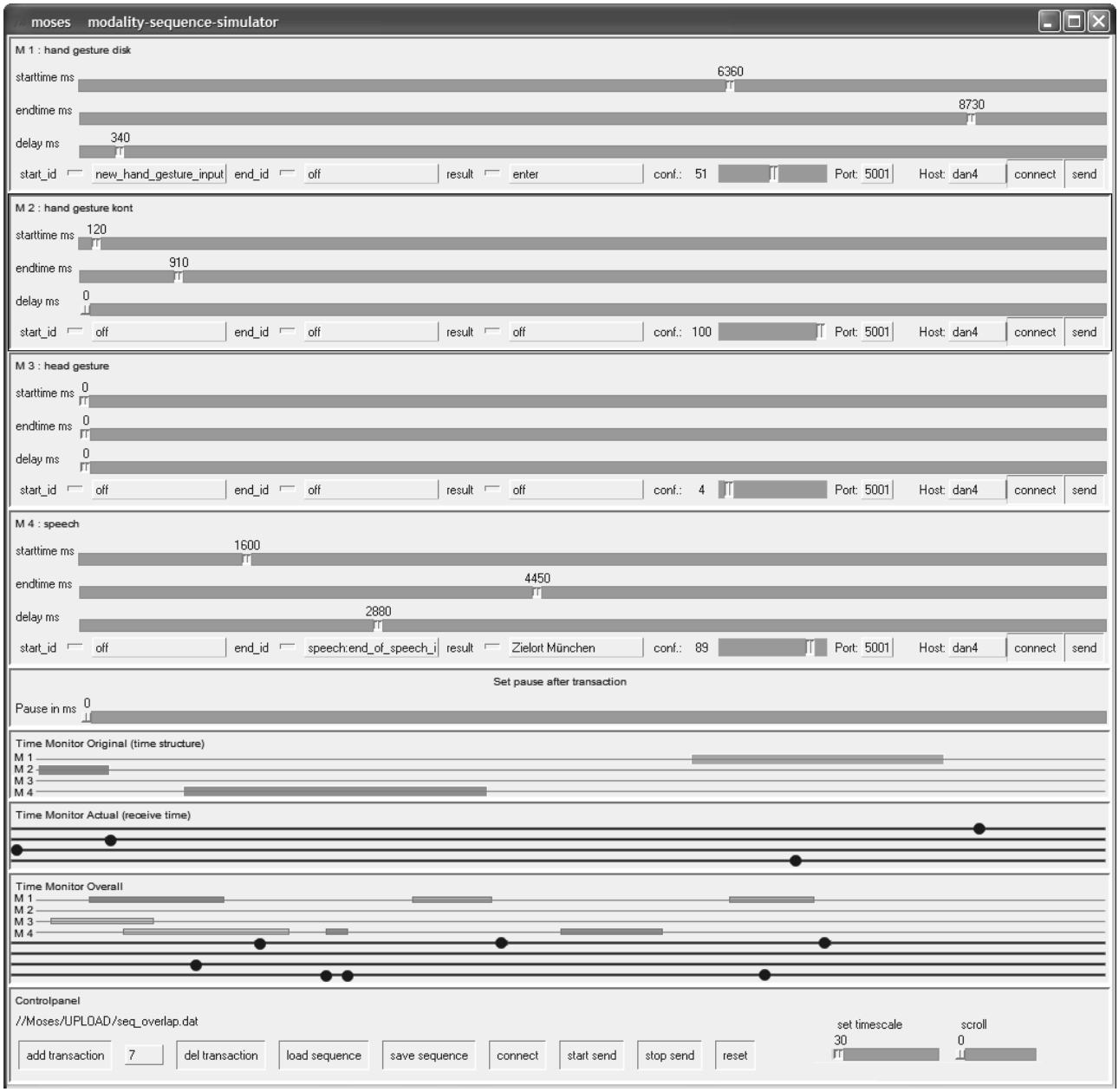


Abb. 3.5: Bedienoberfläche des Simulators für multimodale Eingabesequenzen.

Um das Systemverhalten einer multimodalen Applikation, in diesem Falle die Bedienumgebung GECOM, zu testen, ist es hilfreich, beliebige Bedienszenarien in reproduzierbarer Form simulieren zu können. Aus diesem Grunde wurde ein spezieller Simulator für multimodale Eingabesequenzen entwickelt (siehe Abb. 3.5). Es handelt sich hierbei um ein Software-Tool, welches in der Lage ist, das Verhalten automatischer Erkennungssysteme zu simulieren. Die gewünschten Erkennungsmodule lassen sich flexibel in die Bedienoberfläche des Simulators integrieren, wobei deren spezifische Eigenschaften (z.B. Netzwerkprotokoll) über die vorhandenen Kontrollelemente eingestellt

werden. Das in Abb. 3.5 dargestellte Setup enthält beispielsweise die vier Erkennungsmodule bzw. *Modalitäten*, die für die Ansteuerung der Bedienumgebung GECOM von zentralem Interesse sind:

- *M* 1: diskrete Handgestenerkennung
- *M* 2: kontinuierliche Handgestenerkennung
- *M* 3: Kopfgestenerkennung
- *M* 4: Spracherkennung

Bei der Erstellung von Eingabesequenzen lassen sich insbesondere die zeitlichen Zusammenhänge typischer Ereignisse durch Schieberegler exakt vorgeben. Von besonderer Wichtigkeit sind dabei die folgenden drei *Ereignisse*, welche auch bei der Interaktion mit realen Erkennungssystemen auftreten:

- *E* 1: Beginn der Benutzereingabe (z.B. Ausführungsbeginn einer Geste)
- *E* 2: Ende der Benutzereingabe
- *E* 3: Ausgabe des Klassifikationsergebnisses nach einer bestimmten Verarbeitungszeit

Somit lassen sich sehr flexibel beliebige Benutzerszenarien unter Berücksichtigung der spezifischen Eigenschaften der real vorhandenen Erkennungssysteme konstruieren. Dabei sind insbesondere jene Fälle von Interesse, bei denen sich Eingaben verschiedener Modalitäten zeitlich überlappen. Derartige Überschneidungen können beispielsweise aus unterschiedlichen Verarbeitungszeiten der Erkennungssysteme resultieren (siehe auch [NEU00]) oder bei der multimodalen Bedienung bewusst vom Benutzer eingesetzt werden (z.B. Kopfnicken während einer Spracheingabe). Durch die Simulation derartiger Sonderfälle wird überprüft, ob die angesteuerte Applikation (GECOM) in gewünschter Weise reagiert.

Darüber hinaus kann jedes simulierte Erkennungsergebnis mit einem *Konfidenzmaß* versehen werden. Bei der Wiedergabe einer Sequenz werden die enthaltenen Ereignisse zeitgenau über Netzwerkverbindungen an die Applikation geschickt, aus deren „Sicht“ sich die eintreffenden Datenstrukturen nicht von denen der realen Erkennungssysteme unterscheiden.

Als weiteres Feature ist der multimodale Simulator in der Lage, gespeicherte Log-Dateien (siehe Kap: 3.2) aus real durchgeführten Versuchsszenarien einzulesen, wodurch diese Bediensequenzen durch erneutes Abspielen jederzeit rekonstruiert werden können. Außerdem lassen sich derartige Sequenzen für Testzwecke durch Editieren in beliebiger Weise nachträglich modifizieren.

Der entwickelte Simulator ist also ein nützliches Test- und Diagnosewerkzeug für Forschungsarbeiten an multimodalen Applikationen.

3.5 Adaptives Hilfesystem

Wie weiter oben erörtert (siehe Kap. 3.2), beinhaltet GECOM ein audiovisuelles Hilfesystem. Im Rahmen der Arbeiten von [NIE01] und [NIE02B] wurde eine adaptive Komponente (GHELP, *G*esture *Help*) zur Ansteuerung dieses Hilfesystems entwickelt, welche im Idealfall genau diejenige Hilfeinformation ermittelt, die dem unsicheren Benutzer in seiner aktuellen Situation Klarheit verschafft. Dazu werden alle zur Verfügung stehenden Hilfestellungen unter Berücksichtigung von

Interaktionshistorie, Systemzustand und statistischen Befunden aus Benutzertests mit Prioritäten gewichtet. Dem Benutzer wird schließlich die Hilfestellung mit der höchsten Gesamtpriorität dargeboten. Das Hilfesystem „beobachtet“ also das Interaktionsverhalten des Benutzers und leistet bei Bedarf diejenige Hilfestellung, die optimal sowohl zum Benutzertyp als auch zur aktuellen Situation passt.

In laufenden Forschungsarbeiten von [NIE02B] wird das adaptive Hilfesystem automatisiert, so dass Unsicherheiten des Benutzers selbstständig erkannt werden, wodurch die explizite Hilfeanforderung künftig entfallen kann. Als Entscheidungsmerkmale für Hilfebedarf spielen bei der angewandten Methodik die Konfidenzmaße der automatischen Erkennungssysteme eine zentrale Rolle. Unter der Voraussetzung, dass sich z.B. die unsaubere Ausführung einer Geste in einem entsprechend niedrigen Konfidenzmaß niederschlägt, kann letzteres Rückschlüsse auf eine eventuelle Unsicherheit des Benutzers erlauben. Das in Kap. 7.4.5 eingeführte Verfahren zur Berechnung von Konfidenzmaßen für die Bewertung von Klassifikationsentscheidungen berücksichtigt diese Anforderung.

3.6 Regelbasierte Entscheidungs-Instanz

Wie Abb. 3.1 zeigt, erfolgt die Steuerung der Bedienumgebung GECOM über mehrere Eingabekanäle. Zur Vermeidung bzw. Auflösung von Konflikten, wie sie etwa durch zeitlich überlagerte Datenströme aus verschiedenen Modalitäten entstehen können, wird eine regelbasierte Entscheidungs-Instanz eingesetzt. Die vorhandenen Eingabekomponenten senden ihre Daten zunächst an dieses Modul, welches anhand eines Regelwerks entscheidet, ob bzw. welche Informationen an GECOM weitergeleitet werden. Eine Besonderheit stellt dabei das implementierte Regelwerk selbst dar. So kann dieses einerseits auf konventionelle Weise nach „If-Then-Schema“ manuell erstellt werden. Andererseits besteht jedoch auch die Möglichkeit der automatischen Generierung unter der Verwendung lernender Algorithmen, die aus Gründen der Portabilität auf erweiterte Funktionsumfänge implementiert wurde.

3.6.1 Automatische Generierung eines Regelwerks

Hierfür wird der sogenannte *Decision-Tree-Algorithmus* (Version C4.5) nach R. QUINLAN [QUI93] eingesetzt. Es handelt sich hierbei um ein Verfahren, das dem Fachgebiet des *überwachten maschinellen Lernens* zuzuordnen ist. Aus gegebenen Trainingsbeispielen („*Samples*“) wird ein Entscheidungsbaum generiert, der es erlaubt, neue, zuvor unbeobachtete Szenarien einer vorgegebenen Zielklasse zuzuweisen. Dabei entspricht jede dieser Klassen einer möglichen Entscheidung („*Goal*“), die aufgrund eines beobachteten Samples getroffen wird. Ein Sample setzt sich dabei aus *Attribut-Wert-Paaren* zusammen, wobei jedes Attribut entweder einen diskreten oder einen kontinuierlichen Wert beinhalten kann. Da die finale Entscheidung auf diesen Attributen beruht, ist die Güte der Entscheidungen maßgeblich von der Auswahl der Attribute abhängig. Hierbei ist also darauf zu achten, dass die Attribute Informationen tragen, welche für die zu treffenden Entscheidungen von Belang sind. Dies soll anhand nachfolgender Beispiele (siehe Tab. 3.1) veranschaulicht werden. Es handelt sich hierbei um drei von insgesamt zwölf eingesetzten Attributen, aufgrund derer eine Entscheidung darüber getroffen werden soll, wie GECOM auf die eingehenden Daten eines Erkennungssystems reagieren soll.

Attribut	Wertebereich	Typ
Cause	*Start Input, End Input, Recognized Input*	diskret
Confidence	*0, …, 1*	kontinuierlich
Application State	*Main Menu, Radio, Phone Book, Phone Call Incoming, CD Player, CD Changer, …*	diskret

Tab. 3.1: Attribute eines Samples anhand dreier Beispiele.

Das Attribut *Cause* gibt an, welcher Kategorie die eingehende Botschaft eines Erkennungssystems angehört. Dabei wird unterschieden zwischen *Start Input* bzw. *End Input* (Erkenner hat den Beginn bzw. das Ende einer Benutzereingabe detektiert) und *Recognized Input* (Erkenner hat das Erkennungsergebnis einer abgeschlossenen Benutzereingabe gesendet). Ein weiteres wichtiges Attribut ist das Konfidenzmaß (*Confidence*), welches angibt, mit welcher Zuverlässigkeit eine Benutzeraktion erkannt wurde. Der zulässige Wertebereich ist dabei kontinuierlich (von Null bis Eins). Des Weiteren wird auch der aktuelle Systemzustand der zu steuernden Applikation, in diesem Falle GECOM, durch ein eigenes Attribut in den Entscheidungsprozess integriert. Der Wertebereich erstreckt sich hierbei über insgesamt zwölf diskrete Zustände, von denen ein exemplarischer Auszug in obiger Tabelle genannt wird (*Application State*).

Der Entscheidungsraum umfasst zehn Klassen (Goals). Einige Beispiele sind exemplarisch in Tab. 3.2 aufgeführt.

Goal	Vorausgehendes Erkennungsereignis	Konsequenz
Start OK	*Beginn einer neuen Benutzereingabe*	Daten weiterleiten an GECOM
Start Deny	*Beginn einer neuen Benutzereingabe*	Daten verwerfen
End OK	*Ende einer Benutzereingabe*	Daten weiterleiten an GECOM
End Deny	*Ende einer Benutzereingabe*	Daten verwerfen
Recognition OK	*Übergabe eines Erkennungsresultates*	Daten weiterleiten an GECOM
Recognition Deny	*Übergabe eines Erkennungsresultates*	Daten verwerfen
Recognition Bad	*Übergabe eines Erkennungsresultates*	Daten verwerfen

Tab. 3.2: Beispiele für implementierte Entscheidungsklassen (Goals).

Das angewandte Decision-Tree-Verfahren zeichnet sich insbesondere dadurch aus, dass es selbst für zuvor unbeobachtete Situationen in der Lage ist, adäquate Entscheidungen zu treffen. Dabei ist mit einer gewissen Fehlerrate zu rechnen, die mit zunehmendem Umfang an Beobachtungen abnimmt. Damit ein Regelwerk automatisch generiert werden kann, ist also ein Training erforderlich, d.h., dem Algorithmus muss eine erschöpfende Reihe von beobachteten Situationen zur Verfügung gestellt werden. Diese Samples werden jeweils mit einer Entscheidung versehen, die ein menschlicher Experte zu treffen hat (*überwachtes Training*).

Bei der vorliegenden Anwendung ergibt sich folgende Vorgehensweise für das Training: Mit dem Ziel, möglichst charakteristische Bediensequenzen zu sammeln, agiert ein Benutzer multimodal mit

GECOM. Hierbei handelt es sich um einen Experten, d.h. um eine Person mit klaren Vorstellungen über das gewünschte Zielverhalten von GECOM bei multimodalen Eingaben. Bei der Bedienung werden vom Experten insbesondere auch Ausnahmesituationen wie etwa die oben erwähnten Zeitüberlappungen „provoziert“. Nach jeder Einzelaktion wird der Experte aufgefordert, eine Entscheidung darüber zu treffen, in welcher Weise GECOM reagieren soll, d.h. er wählt eines der vorhandenen Goals aus. Dabei erweist sich die Möglichkeit, sämtliche Bediensequenzen für das Training nicht tatsächlich ausführen zu müssen, sondern diese stattdessen unter Verwendung des Simulators für multimodale Eingaben (siehe Kap. 3.4.2) größtenteils künstlich generieren zu können, als große Erleichterung. Die simulierten Bedienbeispiele werden zusammen mit den Expertenentscheidungen abgespeichert. Der Decision-Tree-Algorithmus generiert aus dieser Datenbasis automatisch einen Entscheidungsbaum, der schließlich das Regelwerk für die implementierte Entscheidungsinstanz darstellt.

Der Hauptvorteil dieser Vorgehensweise liegt in der flexiblen Erweiterbarkeit des Systemverhaltens bei multimodaler Bedienung. Ergibt sich beispielsweise nachträglich der Wunsch, durch eine bestimmte Kombination von Eingaben unterschiedlicher Modalität eine besondere Systemreaktion zu erzielen, kann dies erreicht werden, ohne den Quellcode antasten zu müssen. Es ist lediglich notwendig, die Trainingsdatenbasis um die entsprechenden Samples zu erweitern, woraufhin automatisch ein neues Regelwerk generiert werden kann. Folgendes Beispiel soll eine derartige, nachträgliche Adaption veranschaulichen:

GECOM befinde sich im Zustand *Radio*. Der Benutzer beginnt eine Spracheingabe, entscheidet jedoch nach kurzer Zeit, diese wieder abzubrechen und macht aus diesem Grunde eine horizontale Wischbewegung mit der Hand. GECOM würde darauf wie folgt reagieren: Zunächst wird das - vermutlich unsinnige - Sprachkommando ausgeführt, woraufhin durch die Wisch-Geste der Ton stummgeschaltet wird. Das ursprünglich von Benutzer beabsichtigte Systemverhalten kann erreicht werden, indem das oben genannte Szenario in die Trainingsdatenbasis aufgenommen wird. Der Decision-Tree-Algorithmus generiert daraufhin ein neues Regelwerk, in dem dieser Sonderfall entsprechend berücksichtigt wird.

Für eine detaillierte Beschreibung der Regelwerk-Generierung sowie der implementierten Algorithmen siehe [FIS03] und [QUI93].

4

Intuitives Gesteninventar

4.1 Gestik als Eingabemodalität

Zunächst muss geklärt werden, in welchem Sinne der Begriff *Gestik* im Zusammenhang dieser Arbeit zu verstehen ist. Folgende Abschnitte sollen daher einen schematischen Überblick geben, wobei eine Einteilung der Modalität Gestik in verschiedene Kategorien vorgenommen wird.

In [PAY00] wird eine Geste definiert „als eine Handlung, die einem Zusehenden ein optisches Signal übermittelt (Geste = beobachtete Handlung)“. Etwas restriktiver ist die Auslegung nach [KEN86]: Hier wird eine Geste als eine *sichtbare Handlung* verstanden, welche der Kommunikationsteilnehmer routinemäßig „herausfiltert“ und zudem davon ausgeht, dass eine kommunikative *Absicht* vorliegt.

Diese kommunikative Absicht ist ein wichtiger Aspekt zur Abgrenzung einer Geste im Sinne dieser Arbeit von willkürlichen Körperbewegungen. Bezüglich dieser Betrachtungsweise unterscheidet [MOR81] wie folgt:

- *Primäre Gesten*: Gesten, die in rein kommunikativer Absicht hervorgebracht werden.
- *Sekundäre Gesten*: Handlungen, die zwar beiläufig Information übermitteln, aber nicht in erster Linie mit dieser Absicht erfolgen (z.B. Verdecken der Nase beim Niesen).

Zur Nutzung der Gestik als Eingabemodalität werden im Folgenden lediglich primäre Gesten in Betracht gezogen. Diese lassen sich durch weitere Kategorisierung (siehe auch [STU92] und [FLE01]) grob unterteilen in:

- *Ganzkörpergesten*: Der ganze Körper wird hierbei als Kommunikationsmittel genutzt (z.B. Pantomime).
- *Teilkörpergesten*: Die wohl häufigste Form menschlicher Gesten erfolgt durch Hand- und Kopfbewegungen.

- *Dynamische Gesten*: Wie die Bezeichnung nahe legt, handelt es sich hierbei um Körper*bewegungen*. Im Allgemeinen wird der Hauptanteil der Information, welche durch diesen Gestentyp übermittelt werden soll, durch den Bewegungsablauf selbst transportiert (z.B. Kopfnicken).

- *Statische Gesten*: Bei diesem Gestentyp ist für die Informationsübermittlung keine Bewegung nötig sondern sie erfolgt allein durch die Formgebung bzw. Haltung eines Körperteils. Das Fingeralphabet der Gehörlosensprache (siehe Abb. 4.1) beruht beispielsweise größtenteils auf statischen Handgesten.

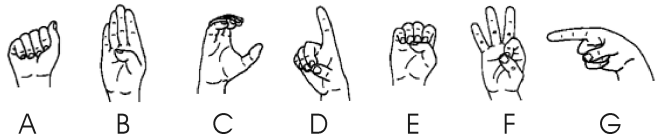


Abb. 4.1: Deutsches Fingeralphabet der Gebärdensprache (Auszug).

In Anlehnung an möglichst natürliche Kommunikationsformen, wird in dieser Arbeit *dynamische Teilkörpergestik*, nämlich Hand- und Kopfgestik, für die Informationseingabe im Mensch-Maschine-Dialog genutzt. Der Einsatz statischer Gesten wird hingegen nicht in Betracht gezogen, da es sich hierbei um eine eher künstliche Art der Kommunikation handelt und somit der hier angestrebten Grundphilosophie widerspricht. Darüber hinaus wurden in zahlreichen Untersuchungen zur gestischen Interaktion (siehe [ZOB01] und [ZOB02]) nahezu keine statischen Gesten beobachtet.

Zur weiteren Verfeinerung dynamischer Teilkörpergestik bietet sich die Betrachtung existierender Taxonomien von [EFR72], [MOR81] sowie [MOR00] an. Daraus werden im folgenden Abschnitt lediglich diejenigen Arten von Gesten aufgeführt, deren Einsatz für den angestrebten gestischen Mensch-Maschine-Dialog potenziell sinnvoll erscheint. Aus anschließend genannten Gründen werden hiervon jedoch nicht alle Kategorien in der vorliegenden Arbeit tatsächlich verwendet.

- *Mimische Gesten*: Hierbei handelt es sich um Gesten, die Personen, Dinge oder Vorgänge bzw. deren Eigenschaften imitieren (Beispiele: Spiralbewegung zur Beschreibung einer Wendeltreppe oder imitiertes Abheben eines Telefonhörers).

- *Schematische Gesten*: Auch diese Gestenart beruht auf imitierenden Bewegungen. Im Gegensatz zur mimischen Gestik handelt es sich hierbei jedoch meist um standardisierte Kürzel, welche nur dann verstanden werden können, wenn sie zuvor erlernt wurden. Sie können dennoch als intuitiv bezeichnet werden, da sie als allgemein bekannte Ausdrucksmittel jedermann geläufig sind (z.B. „Geldschein-Geste“, d.h. Reiben des Daumens an Zeige- und Mittelfinger). Zur Vermeidung von Missverständnissen ist jedoch zu beachten, dass schematische Gesten teilweise kulturabhängig sind (siehe [AXT98]).

- *Kinemimische Gesten*: Dies sind Gesten, bei deren Ausführung eine Bewegungsrichtung nachgeahmt wird (z.B. Winken nach rechts, links, oben oder unten).

- *Symbolische Gesten:* Darunter versteht man Gesten, die abstrakte Eigenschaften beschreiben. Es kann sich hierbei um Emotionen, Stimmungen oder Gedanken handeln, welche sich

oftmals nicht durch reale Objekte beschreiben lassen. Wie die schematischen Gesten müssen sie zwar erlernt werden, sind jedoch üblicherweise allgemein bekannt (Beispiele: „Daumen nach oben“ im Sinne einer Bestätigung bzw. einer positiven Haltung oder „Den Vogel zeigen“ für Dummheit). Auch bei diesem Gestentyp existieren interkulturelle Unterschiede.

- *Deiktische Gesten:* Hierbei handelt es sich um objektbezogene Gesten, die sichtbar vorhandene Objekte adressieren. Dies geschieht meist durch direktes Zeigen mit dem Finger auf das intendierte Ziel.

- *Technische Gesten:* Dies sind Gesten, die von Experten zur Verständigung genutzt werden. Meist geschieht dies aufgrund besonderer Umgebungsbedingungen, welche die sprachliche Kommunikation nicht zulassen (Beispiele: Kommunikation bei Tauchern oder in lauter Umgebung, siehe Abb. 4.2).

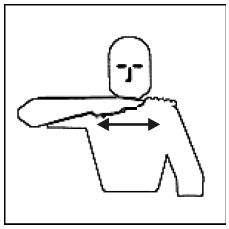


Abb. 4.2: Technische Geste für die Bedienung landwirtschaftlicher Maschinen mit der Bedeutung: „Ausschalten“ bzw. „Motor abstellen“.

- *Kodierte Gesten:* Wie bei den technischen Gesten dienen kodierte Gesten der Kommunikation zwischen Experten. Meist handelt es sich hierbei um eine eigene Zeichensprache, die nur dann verstanden werden kann, wenn sie explizit erlernt wurde (z.B. Gebärdensprache, siehe Abb. 4.1).

Theoretisch könnten alle der vorangehend aufgeführten Gestenarten für die Mensch-Maschine-Interaktion eingesetzt werden. Eine erste Selektion erfolgt jedoch allein aufgrund des Hauptanliegens dieser Arbeit, indem ein Gestentyp nur dann in Betracht gezogen wird, wenn seine Verwendung zu einer intuitiveren Bedienung beizutragen verspricht. Die letzten beiden Kategorien (technische und kodierte Gesten) scheiden aus diesem Grunde aus, da ihnen kein allgemein verständliches Gesteninventar zugrunde liegt.

Darüber hinaus wird auch von der Implementierung deiktischer Gesten abgesehen. Diese Entscheidung beruht einerseits auf der Tatsache, dass ein durch Zeigen adressiertes Gerät im Allgemeinen zunächst visuell fixiert werden muss - bei der Anwendung im Fahrzeug wäre somit eine Blickabwendung vom Straßenverkehr unvermeidbar. Darüber hinaus wurde bereits in früheren Forschungsarbeiten von [PAP99] belegt, dass Zeigebewegungen mit extremen kognitiven Belastungen einhergehen. Diese Befunde konnten durch eigene Untersuchungen (siehe Kap. 5) bestätigt werden. Der Einsatz deiktischer Gestik widerspräche dem hier verfolgten Ansatz zur gestischen Bedienung im Fahrzeug. Schließlich soll gerade die Ablenkung vom Verkehrsgeschehen verringert anstatt erhöht werden.

In der vorliegenden Arbeit wird die dynamische Handgestik für zwei grundsätzlich verschiedene Eingabemethoden eingesetzt. Daher erscheint es sinnvoll, diesen Gestentyp in folgender Weise abermals zu unterteilen:

- *Diskrete dynamische Gesten*: Im Sinne des Begriffs *diskret*[3] ist unter diesem Gestentyp *ein* abgeschlossener Bewegungsablauf zu verstehen, welcher *einen* bestimmten Bedeutungsinhalt übermittelt und somit genau *eine* bestimmte Systemreaktion zur Folge hat. Diese Art der gestischen Bedienung wird von [MOR00] auch als *indirekte Manipulation* bezeichnet.
- *Kontinuierliche dynamische Gesten*: Wie der Begriff *kontinuierlich* nahe legt, wird der aktuelle Systemzustand hierbei direkt *während* einer (Hand-) Bewegung verändert (*direkte Manipulation* [MOR00]). Im Gegensatz zur diskreten Gestik liegt der Informationsgehalt sowohl in der Bewegungsrichtung als auch in der Bewegungsamplitude bzw. der absoluten Position der Hand im Raum. Der Einsatz der kontinuierlichen Gestik erlaubt somit die stufenlose Beeinflussung von Regelgrößen wie z.B. der Musiklautstärke oder der Position eines grafisch dargestellten Objekts. Es handelt sich hierbei im Sinne der obigen Kategorisierung um kinemimische Gesten.

Prinzipiell wäre in dieser Arbeit auch der Einsatz kontinuierlicher Kopfgestik denkbar. Sie wird bereits heute von körperlich behinderten Personen als alternative Eingabemodalität verwendet, etwa zur Positionierung eines Cursors auf einer grafischen Bedienoberfläche (z.B. HEADMOUSE der Firma ORIGIN INSTRUMENTS). Da diese Art der Bedienung jedoch hochgradig unnatürlich ist, wird sie in dieser Arbeit nicht in Betracht gezogen. Das in Abb. 4.3 dargestellte Schema gibt einen Überblick über diejenigen Gestentypen, welche in dieser Arbeit letztlich für die Interaktion im Fahrzeug eingesetzt werden.

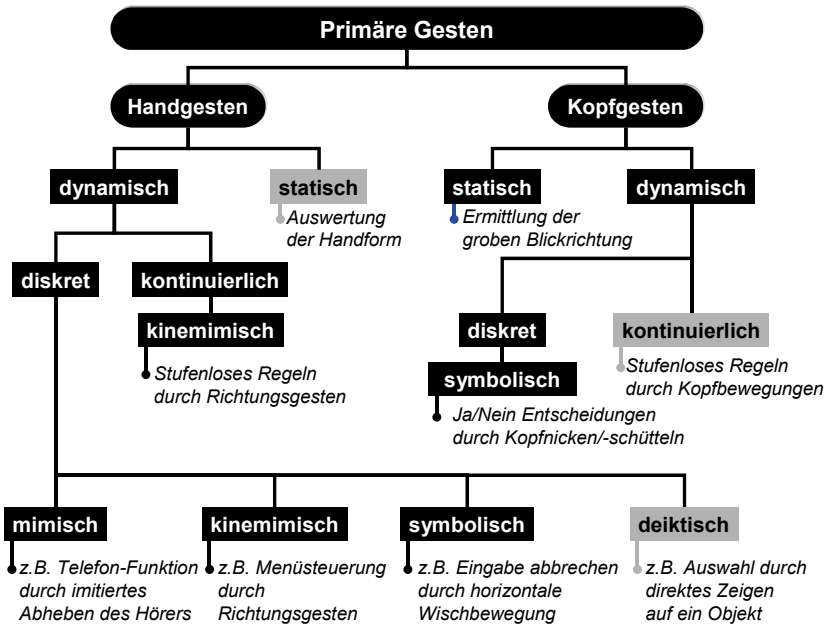


Abb. 4.3: Geeignete (schwarz hinterlegt) sowie ungeeignete (grau hinterlegt) Gestentypen für die Bedienung von Geräten im Fahrzeug (jeweils mit Ausführungsbeispiel).

[3] Definition nach Langenscheidts Fremdwörterbuch: „... 4. *(math.)* voneinander getrennt, einzeln, gesondert (Zahlen, Werte) 5. *(ling.)* abgrenzbar, voneinander trennbar ...“

4.2 Voruntersuchungen

Als Ausgangsbasis für die Nutzung des visuellen Kanals für Informationseingaben im Fahrzeug wurden umfangreiche Voruntersuchungen durchgeführt (siehe auch [ZOB01] und [ZOB02]). Mit diesen Studien sollte generell abgeklärt werden, ob sich die gestische Bedienung in dieser Domäne gewinnbringend, d.h. zum Nutzen des Fahrers, einsetzen lässt. Dabei wurde zunächst ausschließlich die Handgestik als isolierte Eingabemodalität betrachtet. Im Folgenden wird auf einige Ergebnisse dieser Voruntersuchungen eingegangen, da sie das Fundament der vorliegenden Arbeit bilden. Für detaillierte Angaben zur angewandten Methodik sei an dieser Stelle auf die oben genannten Quellen verwiesen.

Um herauszufinden, ob weiterer Forschungsaufwand zu dieser Thematik generell lohnenswert ist, war zunächst die Beantwortung folgender Fragestellungen von zentralem Interesse:

- Wie *intuitiv* ist die gestische Bedienung?
- Wie hoch ist die Benutzer-*Akzeptanz*?
- Inwiefern lässt sich die gestische Bedienung durch die MMI-Gestaltung (*Visualisierung*) beeinflussen?

Untersucht wurden sowohl fahrzeugbezogene Funktionen (z.B. Betätigung des Schiebedachs) als auch gerätetypische Funktionen von Infotainmentsystemen, wobei für diese Arbeit ausschließlich die Letztgenannten von Interesse sind. Dabei wurden den Versuchspersonen nahezu keine Randbedingungen für die Ausführung der Gesten auferlegt. Eine Ausnahme bestand in folgender Einschränkung: Es durften ausschließlich Einhand-Gesten eingesetzt werden. Unter Anbetracht der Einsatzdomäne *Fahrzeug* leuchtet diese Maßnahme ein.

4.2.1 Intuitivität

Die Intuitivität wurde allen anderen Aspekten als wichtigste Voraussetzung für den potenziellen Einsatz von Gestik im Fahrzeug vorangestellt. Es scheint plausibel, dass die Bereitstellung einer zusätzlichen Eingabemodalität wenig sinnvoll ist, wenn deren Anwendung abermals die Zuhilfenahme eines Benutzerhandbuchs voraussetzt. Es stellte sich also die Frage nach der Existenz eines intuitiv gestisch bedienbaren Funktionsumfangs. Ein Gesten-Funktions-Paar wurde in diesem Zusammenhang nur dann als intuitiv erachtet, wenn dessen Auftreten mit hoher intra- und interpersoneller Übereinstimmung zu beobachten war. Dies war in der Tat für zahlreiche Funktionalitäten der Fall. Als Beispiel sei hier die Telefon-Funktion genannt: Um einen Anruf einzuleiten, imitierten nahezu alle Versuchspersonen das Abheben eines virtuellen Telefonhörers.

Erwartungsgemäß erwies sich die gestische Bedienung für einige Funktionen jedoch als ungeeignet. In diesen Fällen waren die Versuchspersonen entweder nicht in der Lage, den verlangten Bedienschritt gestisch zu imitieren oder es zeichneten sich keine signifikanten interpersonellen Übereinstimmungen ab. Dies war beispielsweise bei der Aufgabe „Schalten Sie mit einer Geste die Klimaanlage ein“ der Fall. Aus bereits genannten Gründen kommen derartige Funktionen im Folgenden nicht für die gestische Bedienung im Fahrzeug in Frage.

4.2.2 Akzeptanz

Ein wichtiger Aspekt bei der Einführung neuartiger Techniken ist die Akzeptanz des Benutzers. Die Nutzung der Gestik kann nur dann als sinnvoll erachtet werden, wenn sie nicht nur objektive Vorteile mit sich bringt, sondern auch aus Sicht des Anwenders positiv bewertet wird.

Zum Sammeln subjektiver Eindrücke wurden daher nach allen Benutzertests entsprechende Befragungen durchgeführt. Hierbei wurde die Gestik größtenteils sehr positiv bewertet. So scheint die neue Eingabemodalität „weitgehend intuitiv“ und „angenehm“ zu sein sowie „Spaß“ zu machen. Das Resultat einer zentralen Frage, welche *nach* einem der Benutzertests gestellt wurde, ist in Abb. 4.4 dargestellt.

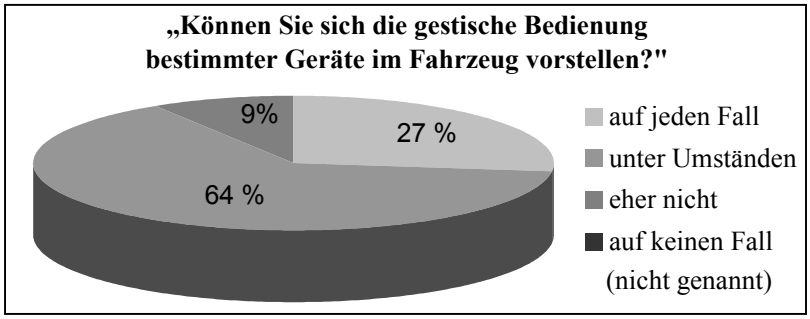


Abb. 4.4: Benutzer-Akzeptanz (elf Versuchsteilnehmer).

Eine Mehrheit der Befragten kann sich die gestische Bedienung im Fahrzeug - zumindest unter Umständen - vorstellen. Eine grundsätzliche Ablehnungshaltung wurde nicht beobachtet.

Ein weiteres Indiz für eine hohe Akzeptanz zeigte sich in jenen Fällen, bei denen dem Benutzer die Wahl der Eingabemodalität (haptisch bzw. gestisch) freigestellt war. Hierbei wurde der Großteil der gestellten Aufgaben von den Versuchspersonen unaufgefordert mittels gestischer Eingaben gelöst.

4.2.3 Visualisierung

Im Laufe der Untersuchungen wurde offensichtlich, dass die grafische Präsentation der Bedienoberfläche (z.B. Anordnung und Ausrichtung der grafischen Grundbausteine) starken Einfluss auf das Gestikulierverhalten des Benutzers übt. Mit dem Ziel, die Intuitivität der gestischen Interaktion durch eine optimierte Gestaltung der Bedienumgebung zusätzlich unterstützen zu können, wurden diese Zusammenhänge näher untersucht. Dazu wurden die Versuchspersonen mit unterschiedlichen grafischen Menüstrukturen (siehe Abb. 4.5) konfrontiert, welche gestisch bedient werden mussten.

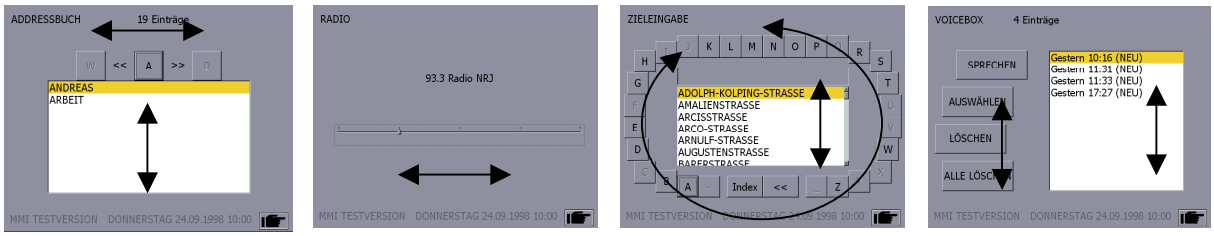


Abb. 4.5: Unterschiedliche Menüstrukturen und offensichtliche Hauptrichtungen (schwarze Pfeile).

Mit nahezu hundertprozentiger Übereinstimmung werden derartige Menüstrukturen entsprechend ihrer Visualisierung, d.h. ihrer Anordnung und Ausrichtung, durch kinemimische Gesten (Richtungsgesten, siehe 4.1) bedient. Der Benutzer orientiert sich also an den offensichtlichen Hauptrichtungen der grafischen Darstellung (schwarze Pfeile in Abb. 4.5) und führt Gesten aus, deren Trajektorien konsequent diesen Strukturen entsprechen.

So werden horizontal bzw. vertikal angeordnete Elemente nahezu ausschließlich durch horizontale bzw. vertikale Richtungsgesten (z.B. Zeigen oder Winken nach rechts/links bzw. oben/unten) bedient (siehe Abb. 4.6).

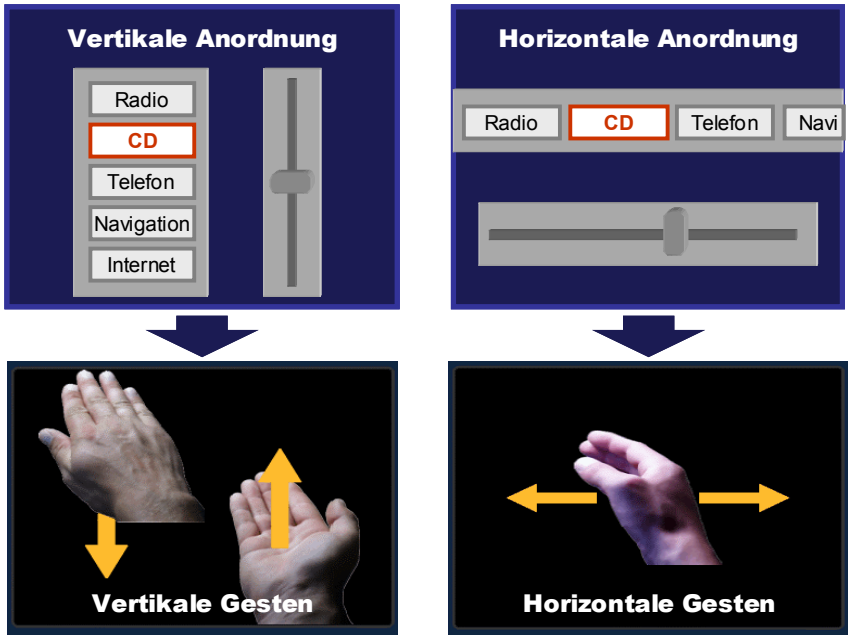


Abb. 4.6: Beeinflussung der Gestik durch die grafische Darstellung.

Darüber hinaus wurde die gestische Bedienung in einem abstrakten Szenario untersucht, bei dem auf die Vorgabe einer Bedienoberfläche gänzlich verzichtet wurde. Auch hier zeigen sich zahlreiche Übereinstimmungen in den beobachteten Gesten. So führen beispielsweise viele Versuchspersonen eine Rechtsbewegung (z.B. Winken nach rechts) im Sinne von „weiter zur nächsten Funktion“ aus bzw. eine Linksbewegung für „zurück zur vorigen Funktion“. Entsprechend werden Auf- bzw. Abwärtsbewegungen verwendet, um eine Regelgröße (z.B. die Musiklautstärke) zu erhöhen bzw. zu erniedrigen.

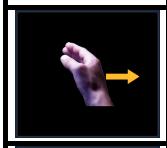
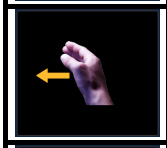
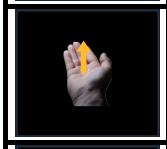
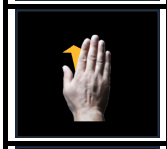
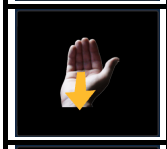
Anhand der gewonnenen Erkenntnisse wurden Gestaltungsregeln aufgestellt, welche bei der Entwicklung der gestenoptimierten Bedienumgebung GECOM berücksichtigt werden (siehe Kap. 6).

4.2.4 Inventar für diskrete Handgestik

Insgesamt wurden im Laufe der Voruntersuchungen sämtliche Kategorien der Handgestik beobachtet, welche in Abb. 4.3 aufgeführt sind. Dabei wurden von den Versuchspersonen nahezu ausschließlich *dynamische* Gesten verwendet, statische Gesten traten hingegen äußerst selten auf. Bei den wenigen beobachteten Ausnahmen handelte es sich meist um deiktisches Zeigen auf bestimmte Geräte bzw. grafisch dargestellte Funktionen. Die Implementierung dieses Gestentyps wird jedoch,

wie bereits in Kap. 4.1 erwähnt, nicht als sinnvoll erachtet, da die Ausführung deiktischer Gesten mit starken Ablenkungseffekten einhergeht.

Der Einsatz von Gesten erweist sich jedoch generell im Hinblick auf die multimodale Interaktion als vielversprechender Ansatz zur Optimierung der Mensch-Maschine-Schnittstelle im Fahrzeug. Aufgrund der beobachteten Gesten ist es nun möglich, ein begrenztes Gesteninventar zu extrahieren, welches für die Bedienung eines Fahrzeug-Infotainmentsystems geeignet erscheint. Das Hauptkriterium für diese Selektion ist die eingangs geforderte Eigenschaft einer hohen intra- und interpersonellen Übereinstimmung als Voraussetzung für die Intuitivität der visuellen Interaktion. Tab. 4.1 zeigt die Klasseneinteilung des Gesteninventars, welches letztlich für geeignet befunden wurde. Zusätzlich erfolgt eine exemplarische Auflistung einiger Systemfunktionen, welche mit den jeweiligen Gesten adressiert werden können.

	Gestenklasse	Kategorie	Zugeordnete Systemfunktion
	1) Winken nach rechts	kinemimisch	- Zum nächsten Menüpunkt wechseln - Objekt (Navigationskarte) nach rechts bewegen
	2) Winken nach links	kinemimisch	- Zum vorigen Menüpunkt wechseln - Objekt (Navigationskarte) nach links bewegen
	3) Winken nach oben	kinemimisch	- Lautstärke erhöhen - Objekt (Navigationskarte) nach oben bewegen
	4) Winken nach unten	kinemimisch	- Lautstärke verringern - Objekt (Navigationskarte) nach unten bewegen
	5) Winken nach vorne	kinemimisch	- Objekt (Navigationskarte) verkleinern (auszoomen)
	6) Winken nach hinten	kinemimisch	- Objekt (Navigationskarte) vergrößern (heranzoomen)
	7) Zeigen nach vorne[4]	symbolisch	- Aktuellen Menüpunkt auswählen

(Fortsetzung auf nächster Seite)

[4] Diese Geste könnte laut Definition (siehe Kap. 4.1) als *deiktisch* aufgefasst werden. Dennoch wird sie hier als *symbolische* Geste kategorisiert, da sie nicht zielgerichtet auf ein Objekt verweist, sondern vielmehr im Sinne einer bestätigenden Handbewegung (vergleichbar mit der Betätigung einer unsichtbaren Taste) zu verstehen ist.

	Gestenklasse	Kategorie	Zugeordnete Systemfunktion
	8) Horizontale Wischbewegung	symbolisch	- Aktuelle Funktion abbrechen - Lautstärke stumm schalten
	9) Ziehen an virtuellem Griff[5]	mimisch	- Hauptmenü aufrufen (Geräteauswahl)
	10) Virtuellen Telefonhörer abheben	mimisch	- Zum Telefonmenü wechseln - Teilnehmer anrufen - Eingehenden Anruf annehmen
	11) Virtuellen Telefonhörer auflegen	mimisch	- Telefonmenü verlassen - Telefongespräch beenden - Eingehenden Anruf ablehnen

Tab. 4.1: Inventar der intuitiven diskreten Handgesten.

4.3 Untersuchungen zur kontinuierlichen Handgestik

Die kontinuierliche Handgestik zur direkten Regelung von Systemparametern als Ergänzung zur bereits in Kap. 4.2 explizit untersuchten diskreten Handgestik wurde in gesonderten Versuchsreihen näher evaluiert (siehe auch [FLE01]). Folgende Fragestellungen standen hierbei im Vordergrund:

- Welches Potenzial bietet die kontinuierliche Gestik als Eingabemodalität hinsichtlich *Intuitivität* und *Akzeptanz*?
- Besteht ein Zusammenhang zwischen der grafischen MMI-Gestaltung (*Visualisierung*) und der bevorzugten Gestenart (diskret bzw. kontinuierlich)?
- Wie signalisiert der Benutzer seine Absicht, den gestischen *Bedienmodus* (diskret bzw. kontinuierlich) zu wechseln?

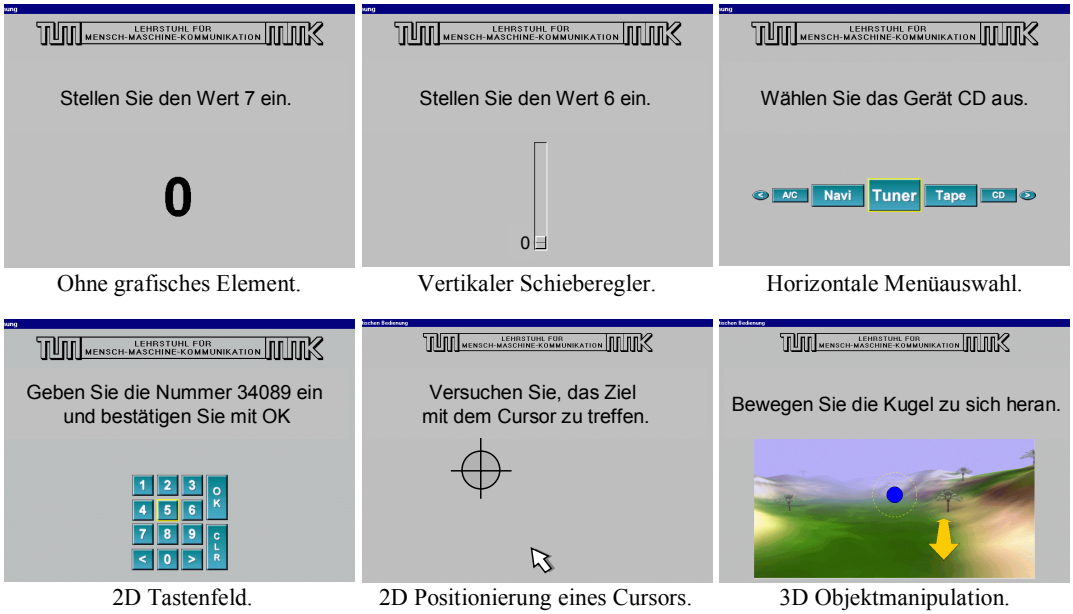
4.3.1 Versuchssetup

Die Untersuchung der kontinuierlichen Handgestik wurde im Rahmen zweier Versuchreihen durchgeführt, an denen insgesamt 22 Personen (14 in Versuchsreihe 1, 8 in Versuchsreihe 2) teilnahmen. Für Aussagen über die Benutzerakzeptanz der gestischen Bedienung wurden testbegleitende Fragebögen eingesetzt.

[5] Diese Geste stellt eine Ausnahme dar, da sie im Gegensatz zu den anderen Gesten nicht mit hoher interpersoneller Übereinstimmung beobachtet wurde. Sie wird aus Gründen, welche in Kap. 6 erläutert werden, dennoch in das Gesteninventar aufgenommen.

Bedienoberfläche

Um möglichst allgemeingültige Erkenntnisse über die Verwendung der betrachteten Gestentypen zu gewinnen, wurden die Versuchspersonen mit relativ abstrakten Szenarien konfrontiert. Dabei wurde die gestische Bedienung zahlreicher allgemein bekannter aktiver Bedienelemente in grafischen Bedienoberflächen isoliert untersucht, ein Auszug ist in Abb. 4.7 dargestellt. Die zur Erfüllung der jeweiligen Aufgabenziele eingesetzte Gestenart war den Versuchspersonen freigestellt.



Ohne grafisches Element. | Vertikaler Schieberegler. | Horizontale Menüauswahl.

2D Tastenfeld. | 2D Positionierung eines Cursors. | 3D Objektmanipulation.

Abb. 4.7: Einige Beispiele für untersuchte Grundbausteine grafischer Bedienoberflächen.

Versuchsumgebung

Die Versuchsreihe wurde in einem Usability-Labor (siehe Abb. 4.8) unter Verwendung der Wizard-of-Oz-Methodik durchgeführt, wobei zahlreiche Automatismen aus [GEI00] zum Einsatz kamen. Der Wizard befindet sich dabei im sogenannten Regieraum und beobachtet via Kamera die Versuchsperson, welche sich im abgetrennten Probandenraum aufhält.

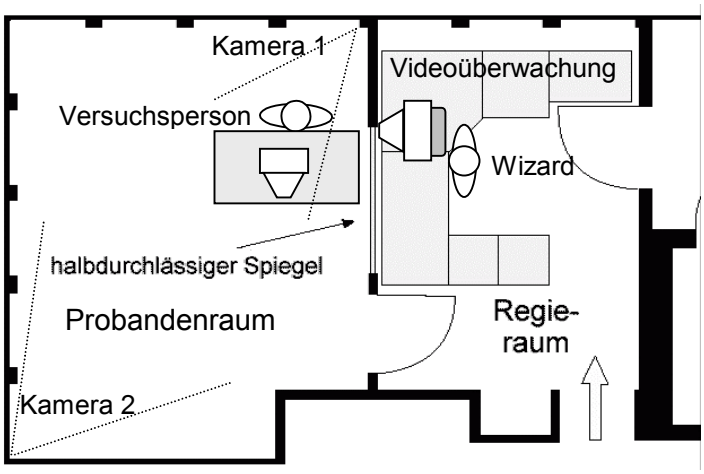


Abb. 4.8: Schematischer Aufbau des Usability-Labors.

Sowohl die Aufgabenstellungen als auch erklärende Erläuterungen zum Versuchsablauf werden durch automatisches Abspielen vorbereiteter Audio-Dateien dargeboten. Zusätzlich besteht eine Audioverbindung über die der Wizard bei Bedarf Rücksprache mit der Versuchsperson halten kann.

Simulation kontinuierlicher und diskreter Gesten

Die Hauptaufgabe des Wizards besteht neben der Steuerung des Versuchsfortschritts darin, die beobachteten Gesten der Versuchsperson umzusetzen. Er simuliert also das Gestenerkennungssystem und zwar sowohl für kontinuierliche als auch für diskrete Handgestik. Zur Gewährleistung echtzeitnaher Systemreaktionen aus Sicht der Versuchsperson, wurde das bereits in Kap. 3.4.1 erwähnte Eingabeverfahren für den Wizard entwickelt. Als Eingabegerät steht ihm eine Kombination aus Datenhandschuh und 3D-Tracker zur Verfügung (siehe Abb. 4.9).

Abb. 4.9: Datenhandschuh mit 3D-Tracker (am Handrücken befestigt).

Diese Ausstattung ermöglicht dem Wizard die Simulation beobachteter Gesten ohne Blickabwendungen vom Versuchsgeschehen, wie dies z.B. bei haptischen Eingaben der Fall wäre. Er beobachtet die Handbewegungen der Versuchsperson und ahmt diese zeit- und ausführungsgleich mit Datenhandschuh und 3D-Tracker nach. Dadurch steuert er die Bedienoberfläche der Versuchsperson, so dass diese den Eindruck gewinnt, die Bedienoberfläche über ein funktionierendes Gestenerkennungssystem selbst zu beeinflussen. Die Simulation der beiden Gestentypen - diskret bzw. kontinuierlich - erfolgt dabei auf jeweils unterschiedliche Weise. Über den Datenhandschuh wird zunächst die Fingerstellung des Wizards ausgewertet. Dies entspricht der Erkennung statischer Handgesten, wobei die in Abb. 4.10 dargestellten Handformen unterschieden werden.

Faust *Daumen abgespreizt* *Zeigefinger abgespreizt*

Abb. 4.10: Statische Handgesten zu Kodierung von Wizard-Eingaben.

Nachahmung diskreter Gesten:

Zur Simulation diskreter Richtungsgesten spreizt der Wizard den Daumen, wodurch der Eingabemodus für diskrete Gesten aktiviert wird. Mit dieser Handstellung führt der Wizard nun eine Bewegung in die gewünschte Richtung (links, rechts, hoch oder runter) aus. Diese wird durch den 3D-Tracker erfasst und zusammen mit der Handhaltung als entsprechende kinemimische Geste erkannt. Daraufhin wird die detektierte Richtungsgeste an die Bedienoberfläche der Versuchsperson übermittelt und es erfolgt eine entsprechende Systemreaktion. Zur Simulation beobachteter Bestätigungsgesten (z.B. Quittieren einer Eingabe) spreizt der Wizard den Zeigefinger ab.

Nachahmung kontinuierlicher Gesten:

Beobachtet der Wizard die Ausführung einer kontinuierlichen Geste, leitet er durch das Bilden einer Faust den kontinuierlichen Modus ein. In diesem Zustand werden die Positionsdaten des Trackers direkt an die Bedienoberfläche der Versuchsperson übertragen und bewirken dort z.B. die direkte Wertänderung einer Stellgröße. Während der kontinuierliche Modus aktiv ist, versucht der Wizard die beobachteten Handbewegungen der Versuchsperson möglichst authentisch zu imitieren, wodurch diese den Eindruck gewinnt, die beobachtete Systemreaktion selbst zu bewirken. Beendet die Versuchsperson das direkte Regeln (z.B. dann, wenn sie das Aufgabenziel erreicht hat), öffnet der Wizard die Hand, wodurch der kontinuierliche Modus beendet wird.

Zur besseren Veranschaulichung hierzu ein Beispielszenario: Die Versuchsperson hat die Aufgabe, mittels Gestik einen vorgegebenen Zahlenwert einzustellen. Dabei wird ein vertikaler Schieberegler auf der Bedienoberfläche dargestellt (siehe Abb. 4.7 oben Mitte). Die Versuchsperson führt eine Winkbewegung nach oben aus. Der Wizard erkennt diese diskrete Geste und simuliert sie indem er den Daumen abspreizt und seine Hand ebenfalls nach oben bewegt. Diese Wizard-Aktion wird detektiert und an die Bedienoberfläche der Versuchsperson übermittelt. Die dazugehörige diskrete Systemreaktion erfolgt, indem der aktuell dargestellte Wert um Eins erhöht wird und der Schieberegler in die entsprechende Position einrastet. Nun führt die Versuchsperson eine kontinuierliche Handbewegung nach oben aus. Der Wizard leitet also durch das Bilden einer Faust den kontinuierlichen Modus ein und vollzieht die Bewegungen der Versuchsperson mit seiner eigenen Hand nach. Der Schieberegler (und der angezeigte Zahlenwert) auf der Bedienoberfläche folgt sodann der Bewegungsrichtung des Wizards bzw. augenscheinlich der Hand der Versuchsperson. Diese direkte Systemmanipulation bleibt solange aktiv, bis der Wizard die Absicht der Versuchsperson erkennt, die Eingabe zu beenden und daher seine Hand wieder öffnet.

4.3.2 Ergebnisse

Intuitivität

Generell erfolgt die gestische Bedienung sehr spontan, wodurch sich positive Rückschlüsse auf deren Intuitivität ziehen lassen. Dies gilt sowohl für die Verwendung diskreter als auch kontinuierlicher Gesten. So konnten die gestellten Aufgaben von nahezu allen Versuchspersonen durch selbstständiges Experimentieren ohne zusätzliche Hilfestellungen gelöst werden.

Dabei erfolgt die diskrete Gestik zur Menüsteuerung (wie schon in den Voruntersuchungen, siehe Kap. 4.2) in erster Linie durch kinemimische Wink- oder Zeigebewegungen. Das Spektrum der be-

obachteten kontinuierlichen Gesten reicht von kleinen Auslenkungen eines Fingers bis hin zu großen Bewegungsamplituden des gesamten Arms.

Akzeptanz

Um ein generelles Meinungsbild über die gestische Bedienung zu erhalten, wurde den Versuchspersonen sowohl vor Beginn der Untersuchung als auch direkt danach die Frage gestellt, ob sie sich die Bedienung von Geräten mit Gesten prinzipiell vorstellen können. Das Resultat der Befragung geht aus Abb. 4.11 hervor.

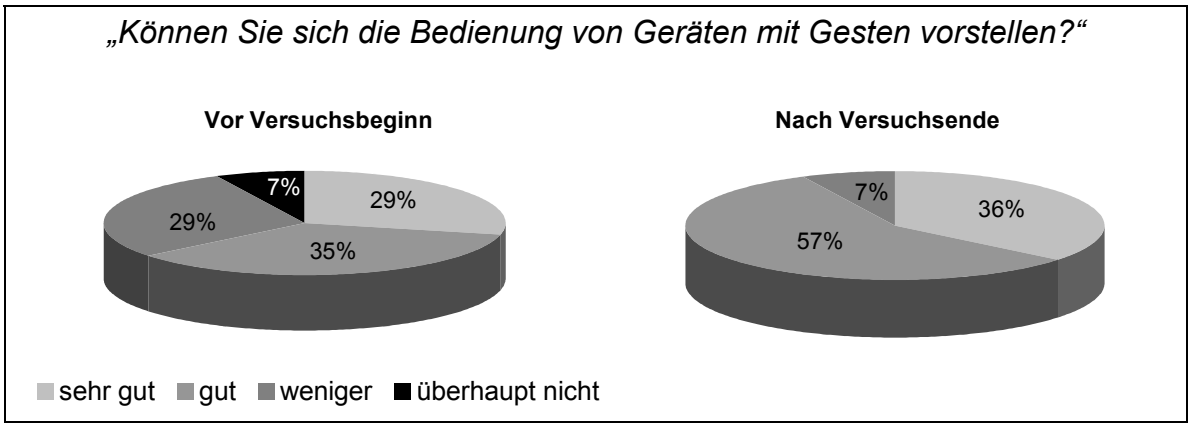


Abb. 4.11: Benutzerakzeptanz der gestischen Bedienung.

Es ist ersichtlich, dass sich bereits vor Versuchsbeginn (siehe linkes Diagramm) eine eher positive Grundhaltung zur gestischen Bedienung abzeichnet. Aufgrund der tatsächlich erfolgten gestischen Interaktion im Laufe der Untersuchung gewinnt die Gestik sogar noch an Akzeptanz hinzu (siehe rechtes Diagramm).

Auch bei der gezielten Beurteilung der Einzelfunktionen hinsichtlich ihrer gestischen Bedienbarkeit ergibt sich insgesamt ein bejahendes Meinungsbild (siehe Abb. 4.12). Besonders bemerkenswert ist hierbei der beobachtete Anteil an kontinuierlicher Gestik von 74 %. Daraus wird gefolgert, dass die kontinuierliche Gestik auf hohe Akzeptanz stößt, da der Eingabemodus während der Versuche freigestellt war.

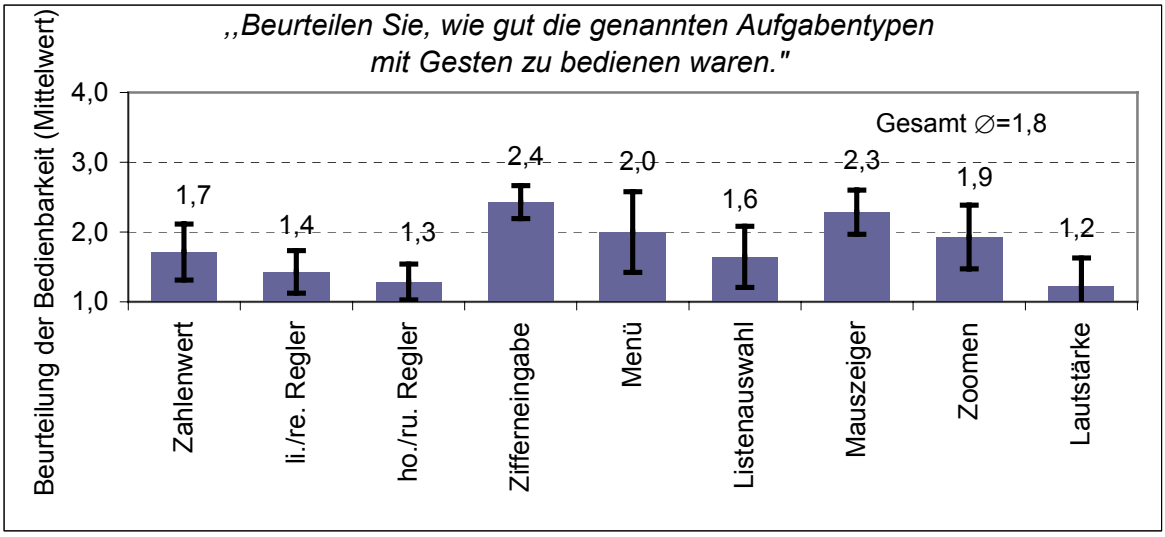


Abb. 4.12: Beurteilung der gestischen Bedienbarkeit durch die Versuchspersonen; Mittelwerte und Standardabweichungen; Bewertungsskala: 1 = sehr gut ... 4 = überhaupt nicht.

Bezüglich der Benutzerakzeptanz ist die Erweiterung der diskreten Handgestik um den kontinuierlichen Modus demnach ein vielversprechender Ansatz.

Visualisierung

Es zeigt sich, dass der bevorzugte Eingabemodus sehr stark vom Aufgabentyp bzw. von der jeweiligen Art der grafischen Darstellung abhängt. Wie aus Abb. 4.13 hervorgeht, überwiegt der beobachtete Anteil an kontinuierlicher Gestik bei zahlreichen Aufgabentypen den der diskreten Gestik.

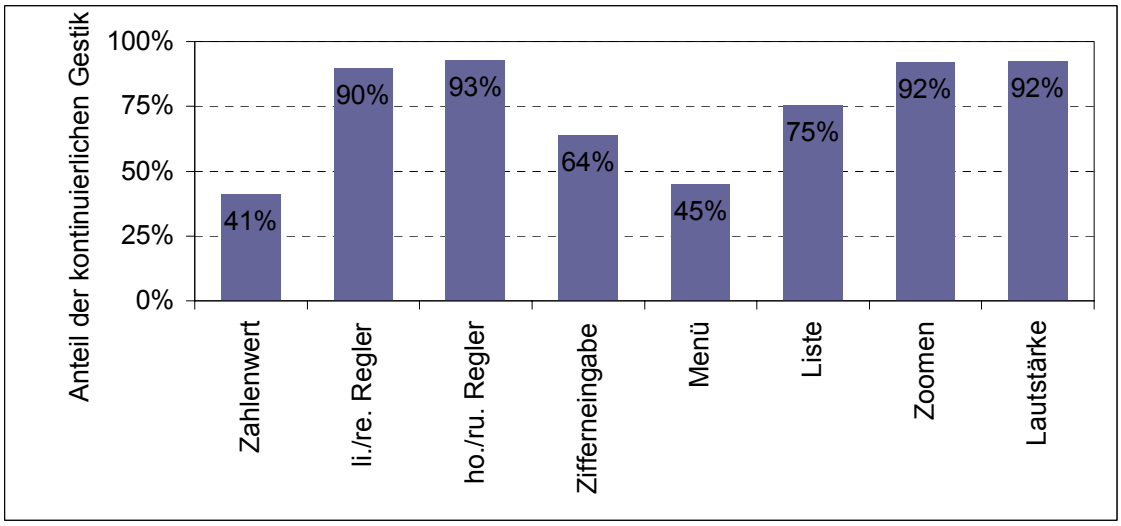


Abb. 4.13: Anteil der beobachteten kontinuierlichen Gestik aufgeschlüsselt nach Aufgabentypen.

Offensichtlich kann dem Benutzer allein durch die Visualisierung eines Eingabeelements der zur Bedienung bevorzugte Gestentyp nahegelegt werden. Dabei gilt folgender plausibler Zusammenhang: Wird dem Benutzer ein kontinuierliches Bedienelement (z.B. ein Schieberegler) dargeboten, so erfolgt auch die Bedienung vorzugsweise durch kontinuierliche Gestik. Dementsprechend steigt der Einsatz diskreter Gesten, wenn die Darstellung in diskreter Weise erfolgt, wie z.B. durch voneinander getrennte Menüpunkte. Diese Beobachtung deckt sich tendenziell weitgehend mit den subjektiven Beurteilungen der Versuchspersonen bei der abschließenden Befragung (siehe Abb. 4.14). Hierbei sollten für die verschiedenen Aufgabentypen persönliche Einschätzungen hinsichtlich des jeweils geeigneten Gestentyps abgegeben werden.

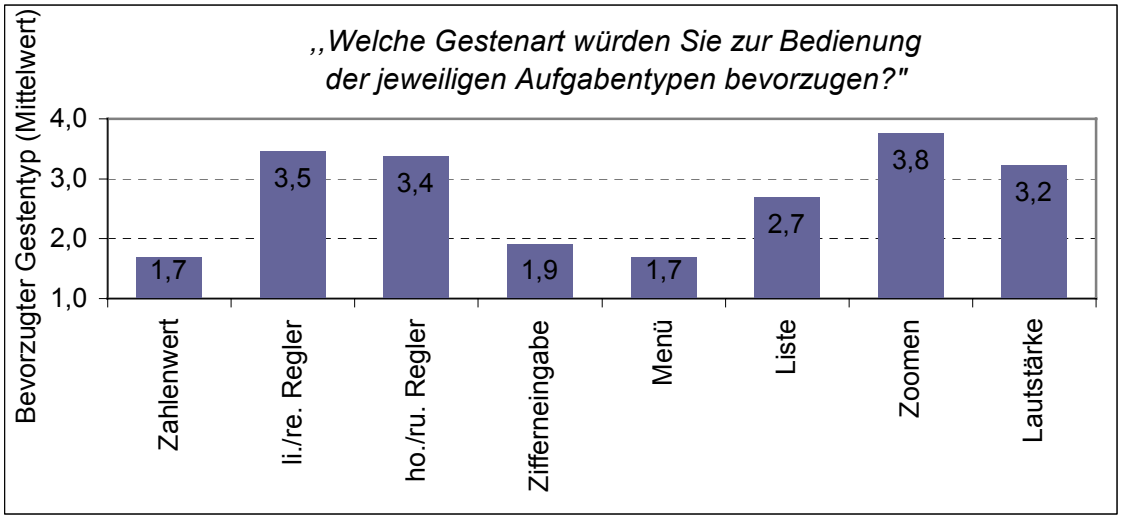


Abb. 4.14: Subjektive Beurteilung des bevorzugten Gestentyps abhängig vom Aufgabentyp; Optionen des Fragebogens: 1 = diskret, 2 = eher diskret, 3 = eher kontinuierlich, 4 = kontinuierlich.

Es ist also festzuhalten, dass die Verwendung des Gestentyps, der aus Sicht des MMI-Designers für eine bestimmte Bedienaufgabe optimal erscheint, durch eine entsprechende grafische Darbietung forciert werden kann.

Darüber hinaus bestätigt sich in dieser Versuchsreihe der bereits in den Voruntersuchungen (siehe Kap. 4.2) festgestellte starke Zusammenhang zwischen der Ausrichtung eines dargestellten Bedienelements und der Ausrichtung der zur Steuerung angewandten Gestik. Demgemäß wird auch hier ein horizontaler (bzw. vertikaler) Schieberegler ausschließlich durch horizontale (bzw. vertikale) Handbewegungen bedient. So banal dieser Zusammenhang auch erscheinen mag, ist er doch eine wichtige Erkenntnis, die bei der Gestaltung gestisch bedienbarer Oberflächen unbedingt berücksichtigt werden muss bzw. im Sinne einer möglichst intuitiven Bedienbarkeit ausgenutzt werden kann.

Moduswechsel

Im Hinblick auf die Realisierung eines automatischen Handgestenerkenners, der in der Lage sein soll, selbstständig zu detektieren, ob ein Benutzer kontinuierliche oder diskrete Gesten ausführt, wurden die diesbezüglichen Verhaltensweisen der Versuchspersonen analysiert. In einer ersten Versuchsreihe wird dem Probanden die Vorgehensweise zur Signalisierung seiner Bedienabsichten weitgehend freigestellt.

Die hierbei gesammelten Beobachtungen sind sehr unterschiedlich. In manchen Fällen ist selbst der menschliche Versuchsleiter (der Wizard) nicht in der Lage, die Absicht des Benutzers eindeutig zu erkennen, da der Übergang zwischen diskreter und kontinuierlicher Gestik nahezu fließend erfolgen kann. Lediglich die Ausführungsgeschwindigkeit erlaubt gewisse Rückschlüsse auf die Absicht des Benutzers: So werden kontinuierliche Gesten meist deutlich langsamer begonnen als diskrete, bei denen der gesamte Bewegungsablauf in der Regel sehr rasch erfolgt. Dahingegen signalisieren einige der Versuchspersonen ihre Absicht sehr deutlich, indem sie in der Luft nach dem dargestellten Bedienelement „greifen“, bevor sie mit der kontinuierlichen Bedienung beginnen.

Insgesamt werden jedoch starke interindividuelle Unterschiede beobachtet. Somit erweist es sich als notwendig, klare Vorgaben für die Ausführung kontinuierlicher Gesten zu definieren, um es einem

automatischen System zu ermöglichen, diese von den diskreten Gesten zu differenzieren. Dazu werden die folgenden beiden Methoden erwogen:

Zunächst wird der diskrete Modus als Standardzustand definiert. Dies bedeutet, dass jede detektierte Handgeste als diskrete Geste interpretiert wird, wenn kein expliziter Wechsel in den kontinuierlichen Modus erkennbar ist. Dieser Moduswechsel muss also vom Benutzer signalisiert werden, wofür im Folgenden zwei Vorgehensweisen festgelegt werden:

- *Zeitkodierter Moduswechsel:* Hierbei signalisiert der Benutzer die Absicht der kontinuierlichen Bedienung, indem er seine Hand zuvor für eine kurze Aktivierungszeit im Erkennungsbereich still hält. Der kontinuierliche Modus wird automatisch beendet, wenn die Hand den Erkennungsbereich verlässt.
- *Handformkodierter Moduswechsel:* In Anlehnung an die beobachteten Greifgesten aktiviert der Benutzer den kontinuierlichen Modus durch das Bilden einer Faust. Solange er diese Handform beibehält, verweilt das System in diesem Bedienmodus. Mit dem Öffnen der Hand kehrt der Benutzer zurück zum diskreten Modus.

In einer zweiten Versuchsreihe werden nun die beschriebenen Methoden zum Moduswechsel fest vorgegeben und evaluiert. Dabei wird einerseits untersucht, ob sich die eingeführten Randbedingungen zu Lasten der Benutzerakzeptanz der kontinuierlichen Gesten auswirken. Andererseits sollte ermittelt werden, welche der beiden Methoden vom Anwender bevorzugt wird.

Hierbei kann sich jedoch aufgrund der subjektiven Benutzerbewertungen kein klarer Trend herauskristallisieren. Dies wird insbesondere durch Abb. 4.15 verdeutlicht, wo eine nahezu ausgeglichene Beurteilung der angebotenen Methoden für den Moduswechsel vorliegt.

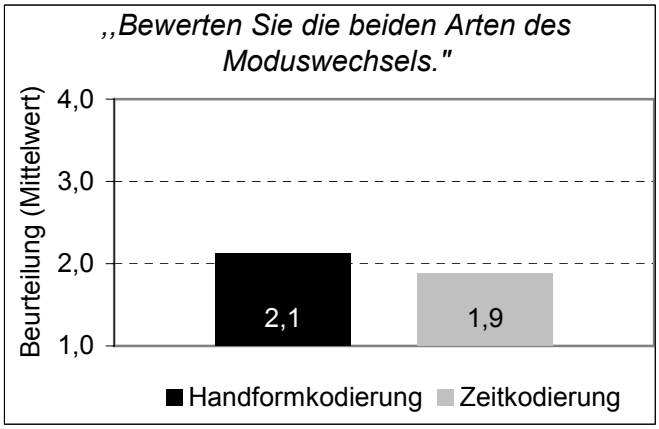
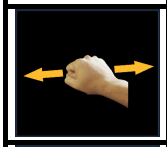


Abb. 4.15: Subjektive Beurteilungen der beiden Arten des Moduswechsels; Bewertungsskala: 1 = sehr gut ... 4 = sehr schlecht.

Ein Akzeptanzverlust gegenüber der ersten Versuchsreihe aufgrund der vorgegebenen Randbedingungen bei der Ausführung kontinuierlicher Gestik konnte nicht festgestellt werden. Hinsichtlich des Usability-Aspekts scheinen die beiden untersuchten Methoden des Moduswechsels einander ebenbürtig zu sein. In der späteren Realisierung (siehe Kap. 7) wurde der Ansatz des zeitkodierten Moduswechsels verfolgt, da sich dieser mit der dort eingesetzten Technologie besonders gut umsetzen lässt.

4.3.3 Inventar für kontinuierliche Handgestik

Für die Implementierung kontinuierlicher Handgestik zur direkten Regelung von Systemparametern der Bedienumgebung GECOM werden die beiden in Tab. 4.2 dargestellten Klassen erwogen.

	Gestenklasse	Kategorie	Zugeordnete Systemfunktion
	1) Horizontales Regeln	kinemimisch	- Cursor horizontal verschieben - Objekt (Navigationskarte) horizontal verschieben
	2) Vertikales Regeln	kinemimisch	- Musiklautstärke einstellen - Objekt (Navigationskarte) vertikal verschieben

Tab. 4.2: Inventar der kontinuierlichen Handgesten (handformkodierter Moduswechsel).

Diese zueinander orthogonalen Bewegungen erlauben die stufenlose, jeweils eindimensionale Beeinflussung von Stellgrößen. Ihre Verwendung wurde sowohl in den Untersuchungen sehr häufig beobachtet als auch seitens der Versuchspersonen besonders positiv bewertet.

4.4 Inventar für diskrete Kopfgestik

Für die Festlegung des Inventars zur Kopfgestenbedienung wurden vorab keine speziellen Untersuchungen durchgeführt. Da sich die natürliche Art der Kopfgestik im Wesentlichen auf Kopfnicken und Kopfschütteln[6] beschränkt, werden auch nur genau diese beiden Gesten in der vorliegenden Arbeit in Betracht gezogen. Es wird davon ausgegangen, dass sich Kopfgesten für die intuitive sowie wenig ablenkende Beantwortung von Systemrückfragen eignen. Die Evaluierung der Kopfgestik erfolgt im Rahmen der an GECOM durchgeführten Usability-Tests (siehe Kap. 6).

Gestenklasse	Kategorie	Zugeordnete Systemfunktion
1) Kopfnicken	symbolisch	- Systemrückfrage bejahen - Eingehenden Anruf annehmen
2) Kopfschütteln	symbolisch	- Systemrückfrage verneinen -Eingehenden Anruf ablehnen

Tab. 4.3: Kopfgesteninventar.

[6] Es sei an dieser Stelle erwähnt, dass die Interpretation dieser beiden Kopfgesten in seltenen Fällen interkulturelle Unterschiede aufweist (siehe auch [AXT98]).

5

Ablenkungseffekte

5.1 Vorüberlegungen

Der beabsichtigte Einsatz von Handgestik als Eingabemodalität im Fahrzeug war im Vorfeld dieser Arbeit umstritten. Ein Hauptdiskussionspunkt war hierbei die Vermutung, dass die gestische Bedienung mit unerwünschten Ablenkungseffekten einhergeht, da die ausführende Hand das Lenkrad zwangsläufig loslassen muss. Da das Anliegen dieser Arbeit jedoch insbesondere die Minimierung der durch die Gerätebedienung verursachten Fahrerablenkung beinhaltet, wurde eine detaillierte Untersuchung der tatsächlichen Ablenkungseffekte angestrengt. Hierbei wird ein objektiver Vergleich zwischen konventioneller Haptik und diskreter Handgestik als jeweilige Eingabemodalität gezogen. Sollte sich hierbei ergeben, dass gestische Bedienung stärker ablenkt als haptische, wäre von deren Einsatz im Fahrzeug sicherlich abzuraten.

In diesem Zusammenhang spielen die Forschungsarbeiten zur *Selektiven Wahrnehmung und Handlungssteuerung* von [DEU98] und [PAP99] eine wichtige Rolle. Untersucht wurden hierbei die Zusammenhänge zwischen der visuellen Aufmerksamkeit und der Selektion im Raum befindlicher Ziele durch Zeigebewegungen.

Die Autoren kommen zu dem Ergebnis, dass die visuelle Aufmerksamkeit des Menschen prinzipiell beschränkt ist auf ein intendiertes Objekt, welches sich im Umgebungsraum befindet. Die in der vorliegenden Arbeit betrachteten Ablenkungseffekte lassen sich sehr anschaulich unter Verwendung des Begriffs der visuellen Aufmerksamkeit definieren: So ist der Autofahrer genau dann abgelenkt, wenn die Zielposition seiner visuellen Aufmerksamkeit vom Verkehrsgeschehen abweicht, wobei sich die hier betrachteten Ablenkungsursachen auf die Bedienung von Geräten während der Fahrt beschränken. Bemerkenswert ist dabei der Sachverhalt, dass derartige Ablenkungseffekte selbst dann vorliegen können, wenn der Blick des Fahrers permanent nach vorne auf den Straßenverkehr gerichtet ist. Dies ist z.B. dann der Fall, wenn er mit einer Hand eine zielgerichtete Bewegung ausführt. Seine visuelle[7] Aufmerksamkeit ist in diesem Falle für einen bestimmten Zeitraum

[7] Der Begriff *visuell* ist in diesem Zusammenhang bezogen auf die verarbeitende Hirnregion.

ausschließlich an das Bewegungsziel gekoppelt. Während dieser Zeit sinkt seine Objekterkennungsleistung bezüglich des Straßenverkehrs drastisch ab. Zusammenfassend sind zunächst folgende Befunde nach [PAP99] festzuhalten:

- Zielgerichtete deiktische Handbewegungen sind gekoppelt mit erheblichen mentalen Ablenkungseffekten.
- Es ist nicht möglich, die visuelle Aufmerksamkeit auf ein Objekt zu richten und gleichzeitig auf ein anderes zu zeigen.
- Vor der Initiierung einer zielgerichteten Bewegung ist die Objekterkennungsleistung auf die Position des Bewegungsziels beschränkt.

Überträgt man diese Erkenntnisse auf die hier betrachtete Mensch-Maschine-Interaktion, so lassen sich folgende Behauptungen aufstellen:

- ➢ Die Betätigung eines haptischen Bedienelements (z.B. das Drücken einer Taste) entspricht im Allgemeinen einer zielgerichteten Bewegung.
- ➢ Bei der gestischen Bedienung handelt es sich hingegen um eher ungerichtete Bewegungen im Ausführungsraum.

Unter dieser Betrachtungsweise liegt also die Vermutung nahe, dass die gestische Bedienung entgegen den oben erwähnten Befürchtungen sogar mit geringeren Ablenkungseffekten einhergehen könnte, als die haptische. Diese Fragestellung wird daher wie nachfolgend beschrieben quantitativ untersucht.

5.2 Methodik

5.2.1 Versuchsumgebung

Um eine authentische Versuchsumgebung zu gewährleisten, wird diese Studie im institutseigenen Fahrsimulator (*Navigation-Lab*) durchgeführt. Das Versuchsfahrzeug befindet sich gemäß Abb. 5.1 vor einer Projektionsleinwand.

Abb. 5.1: Fahrsimulator und Projektion einer abstrakten Fahraufgabe (oberes Dreieck: Soll-Marke; unteres Dreieck: Ist-Marke).

Hauptaufgabe

Zunächst ist es erforderlich, die Versuchsperson mit einer Fahraufgabe zu konfrontieren, welche im Folgenden als *Hauptaufgabe* bezeichnet wird. Sie entspricht dem Steuern eines Fahrzeugs und ist diejenige Handlung, von der die Versuchsperson im Laufe der Untersuchung durch Bedienaufgaben (siehe *Nebenaufgaben*) abgelenkt wird.

Von der Verwendung einer möglichst realistischen Fahrsimulation wurde bewusst abgesehen, da hierbei mit unerwünschten Randeffekten - z.B. verursacht durch die Art der grafischen Darstellung, dem implementierten Fahrzeugmodell etc. - zu rechnen wäre. Die objektive Messung von Ablenkungseffekten gestaltet sich als extrem diffizil; störende Einflussgrößen jeglicher Art gilt es daher weitestgehend zu vermeiden. Aus diesem Grunde wurde ein abstraktes Szenario in Form einer einfachen Regelaufgabe entwickelt: Im primären Sichtfeld der Versuchsperson befinden sich zwei Objekte, die *Soll-Marke* und die *Ist-Marke* (siehe Abb. 5.1 und Abb. 5.3). Die Soll-Marke führt (zufällige) horizontale Bewegungen aus, die dem Verlauf einer Straße nachempfunden wurden (siehe auch Anh. A.1.2). Die Position der Ist-Marke kann über das Lenkrad gesteuert werden, wobei die horizontale Auslenkung linear vom Lenkwinkel abhängt. Die Hauptaufgabe der Versuchsperson besteht nun darin, die Ist-Marke permanent möglichst exakt mit der Soll-Marke in Deckung zu halten.

Nebenaufgaben

Während der Hauptaufgabenausführung wird die Versuchsperson mit zahlreichen Nebenaufgaben konfrontiert. Hierbei handelt es sich um Bedienaufgaben, welche mit der jeweils vorgegebenen Eingabemodalität (haptisch bzw. gestisch) ausgeführt werden müssen. Für die haptische Bedienung wird eine Konsole eingesetzt, welche mit zwei Tasten und einem Dreh-/Drücksteller bestückt ist (siehe Abb. 5.2). Die gestische Bedienung erfolgt innerhalb des natürlichen Greifraums über der Mittelkonsole unter Verwendung der rechten Hand. Die beiden Eingabemodalitäten werden paarweise verglichen, d.h. zu jeder haptischen Eingabeart gibt es ein gestisches Pendant. Eine Zusammenstellung der verwendeten Nebenaufgaben sowie der zugehörigen Eingabemethoden zeigt Tab. 5.1.

		Nebenaufgabe			
		Auswahl	*Telefon*	*Erhöhen*	*Erniedrigen*
Eingabe-Methode	*Haptik*	Auswahl-Knopf drücken	Telefon-Knopf drücken	Drehknopf im Uhrzeigersinn drehen	Drehknopf gegen den Uhrzeigersinn drehen
	Gestik[8]	Zeigegeste nach vorne	Telefonhörer abheben	Winken nach oben	Winken nach unten

Tab. 5.1: Nebenaufgaben und zugehörige Eingabemethoden.

[8] Die hier aufgeführten Gesten sind in Tab. 4.1 jeweils durch eine Grafik veranschaulicht.

Die einzelnen Aufgaben werden dabei sehr abstrakt gehalten: Beispielsweise kann die Aufgabe „Erhöhen“ im Sinne der Erhöhung einer Regelgröße wie etwa der Musiklautstärke verstanden werden. Abb. 5.2 zeigt exemplarisch die Ausführung der Bedienaufgabe „Telefon“ für beide Eingabemethoden.

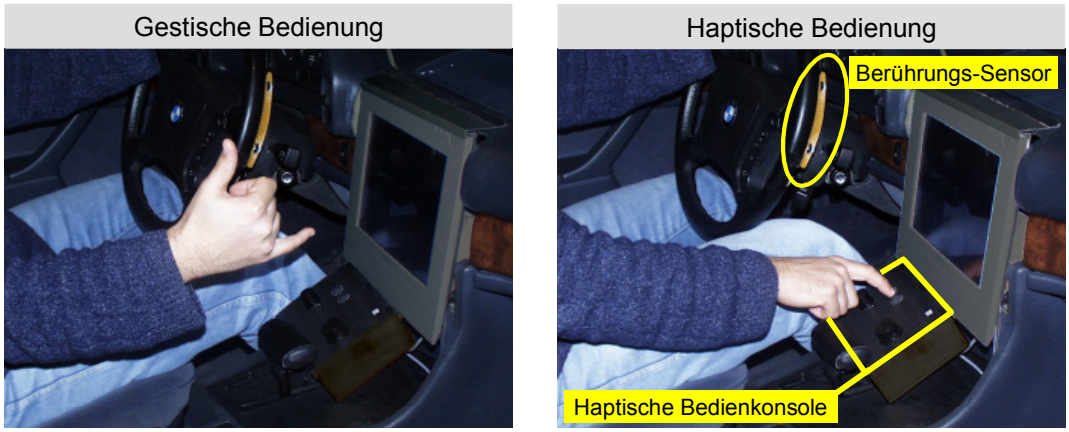


Abb. 5.2: Gestische (links) und haptische (rechts) Ausführung der Bedienaufgabe *Telefon*.

Um sicher zu stellen, dass die gewonnenen Daten versuchsübergreifend optimal verglichen werden können, wurde der gesamte Versuchsablauf vollständig automatisiert. Dadurch wird garantiert, dass alle Ereignisse für jede Versuchsperson zur exakt gleichen Zeit erfolgen. Die Nebenaufgaben werden vorgelesen, indem vorgefertigte Audio-Dateien eingespielt werden. Dabei wird jeweils zuerst die Bedienaufgabe und dann die zu verwendende Modalität genannt. Zur Vermeidung von störenden Nebeneffekten und mit der Absicht, die Vorlesezeit für alle Aufgaben nahezu identisch zu halten, erfolgen die Anweisungen in sehr knapper Form: z.B. „Auswahl mit Geste!“ oder „Telefon mit Knopf!“.

Versuchsablauf

Zunächst wird die Versuchsperson vom Leiter des Experiments mündlich instruiert. Dabei werden einerseits die verschiedenen Eingabemethoden und andererseits der prinzipielle Versuchsablauf erklärt. Sobald die Versuchsperson angibt, alles begriffen zu haben, wird der eigentliche (vollautomatische) Versuch gestartet. Abschließend erfolgt ein kurzes Interview, um die subjektiven Eindrücke der Versuchsperson zu ermitteln.

5.2.2 Messung von Ablenkungseffekten

Um später Rückschlüsse auf die Fahrerablenkung ziehen zu können, werden die folgenden objektiven Messgrößen während des Versuchs ermittelt:

- *Regelfehler:* Betrachtet wird hierbei, wie präzise der Fahrer dem Verlauf der Sollmarke zu folgen vermag. Ein schlechtes Führungsverhalten wird dabei auf eine entsprechend starke Ablenkung zurückgeführt.

- *Ausführungszeit:* Die erforderliche Dauer zur Erfüllung einer Bedienaufgabe ist einerseits ein Maß für die Effizienz der jeweiligen Eingabemodalität. Sie spielt andererseits auch eine Rolle bei der Bewertung der Ablenkung: Je schneller eine Bedienung erfolgt, desto kürzer

können sich die dabei auftretenden Ablenkungseffekte nachteilig auf den Straßenverkehr auswirken.

- *Objekterkennungsleistung:* Die Objekterkennungsleistung ist in dieser Untersuchung die wichtigste Größe zur Beurteilung von Ablenkungseffekten. Sie ist ein direktes Maß für die mentale Fähigkeit des Menschen, Objekte visuell zu erfassen und zudem zu identifizieren bzw. bewusst zu verarbeiten. Sie gibt in dieser Studie Aufschluss darüber, ob die visuelle Aufmerksamkeit nach vorne (auf das Verkehrsgeschehen) gerichtet ist, oder nicht.

Nachfolgend werden die hier angewandten Methoden zur Erfassung der genannten Messgrößen erläutert.

Regelfehler

Während des gesamten Versuchsablaufs werden die x-Positionen von Ist- und Soll-Marke mit einer Abtastfrequenz von $f_{abt} = 50$ Hz aufgezeichnet und anschließend in eine Datei gespeichert. Einen zeitlich zusammenhängenden Auszug der hierbei gewonnenen Daten zeigt Abb. 5.5.

Im Vorfeld des Versuchsdesigns wurde die durchschnittliche Bewegungsgeschwindigkeit der Soll-Marke so eingestellt, dass ein sehr präzises Nachregeln möglich ist, wenn sich der Proband ausschließlich auf diese Regelaufgabe konzentriert, d.h. wenn währenddessen keine Nebenaufgaben gestellt werden. Die Regelabweichung Δx (siehe Abb. 5.3) ergibt sich als Differenzbetrag der aufgezeichneten x-Positionen und wird in der späteren Auswertung zur Berechnung des Regelfehlers herangezogen.

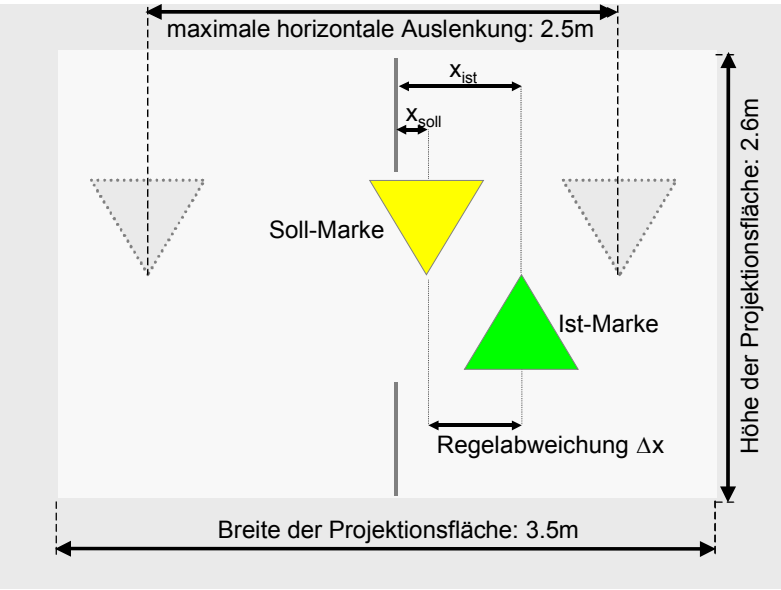


Abb. 5.3: Regelaufgabe und Regelabweichung.

Ausführungszeit

Die Ausführungszeit wird definiert als die Dauer des Zeitintervalls, in dem das Lenkrad von der ausführenden rechten Hand nicht berührt wird. Sie wird mittels eines am Lenkrad angebrachten Berührungssensors (siehe Abb. 5.2) gemessen. Vor Versuchsbeginn wird der Teilnehmer instruiert,

diesen Sensor permanent zu betätigen, d.h., er darf lediglich zur Ausführung einer Bedienaufgabe losgelassen werden. Darüber hinaus wird die Versuchsperson dazu angehalten, jede Bedienung schnellstmöglich sowie ohne Unterbrechung auszuführen. Der jeweilige Zustand des Berührungssensors wird dabei zeitsynchron zusammen mit den *x*-Positionsdaten (siehe *Regelfehler*) gespeichert.

Objekterkennungsleistung

Zur Erfassung der Objekterkennungsleistung erfolgt eine kurzzeitige Einblendung verschiedener grafischer Muster - die sogenannten Diskriminations-Objekte - innerhalb der Ist-Marke, d.h. im primären Sichtfeld der Versuchsperson. Es werden zwei verschiedene Diskriminations-Objekte verwendet: eine „2“ und eine „5“ (jeweils in Sieben-Segment-Darstellung, siehe Abb. 5.4).

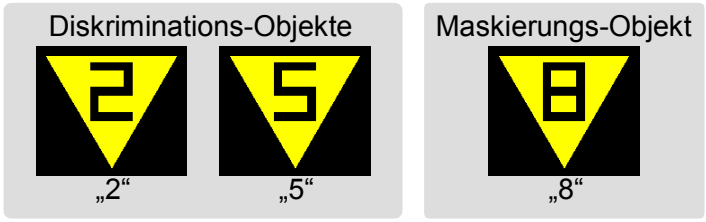


Abb. 5.4: Diskriminations-Objekte (links) und Maskierungs-Objekt (rechts).

Die Auswahl des jeweiligen Diskriminations-Objekts geschieht aus Sicht der Versuchsperson zufällig. Die Darbietungsdauer beträgt 70 ms; danach wird das Diskriminations-Objekt ersetzt durch das Maskierungs-Objekt „8“ (siehe Abb. 5.4). Diese Maßnahme verhindert das „Nachleuchten“[9] des Diskriminations-Objekts auf der Netzhaut der Versuchsperson, wodurch eine konstante effektive Darbietungsdauer sichergestellt wird. Das Maskierungs-Objekt wird nach etwa einer Sekunde wieder ausgeblendet.

Nach jeder Objektdarbietung wird die Versuchsperson aufgefordert, das erkannte Objekt zu nennen. Hierbei wird die *Two-Alternative-Forced-Choice-Methode* (2AFC) eingesetzt, d.h., die Versuchsperson *muss* sich selbst dann für eines der beiden Diskriminations-Objekte entscheiden, wenn sie das Objekt nicht erkannt hat. Bei der späteren Auswertung ist die vorliegende Ratewahrscheinlichkeit von 50 % zu berücksichtigen.

5.2.3 Zeitlicher Ablauf

Im Folgenden wird der zeitliche Ablauf der Einzelereignisse während einer Bedienaufgabe beschrieben. Wie oben erwähnt, werden die Nebenaufgaben eingespielt, *während* die Versuchsperson der Hauptaufgabe, dem Regeln der Soll-Marke, nachgeht. Die Zeitpunkte der Aufgabenstellungen folgen einem festgelegten Zeitschema (jedoch zufällig aus Sicht der Versuchsperson).

---

[9] Aufgrund der Trägheit des menschlichen Auges vergeht nach dem Eintreffen eines Lichtreizes eine gewisse Zeit, bis die Netzhaut den Reiz weiterleitet. Entsprechend überdauert diese Erregung den auslösenden Reiz für eine kurze Weile. Das entstehende „Nachbild“ hat zur Folge, dass ein Objekt länger sichtbar bleibt, als es tatsächlich dargeboten wird.

Zunächst wird eine Bedienaufgabe - z.B. „Telefon mit Knopf“ - eingespielt. Daraufhin wird die Versuchsperson das Lenkrad loslassen und den Telefonknopf betätigen. Mit dem Loslassen des Berührungssensors startet einerseits die Messung der Ausführungszeit. Andererseits triggert diese Aktion die Darbietung eines Diskriminations-Objekts, welches jedoch nicht instantan sondern erst nach einer variierenden Verzögerungszeit Δt_{obj} eingeblendet wird. Um sicher zu stellen, dass die Objektdarbietung *während* der Ausführung der Nebenaufgabe erfolgt, werden dabei folgende Grenzen für das Zeitintervall gewählt: 100 ms < Δt_{obj} < 500 ms (Erfahrungswerte). Nach der Ausführung der Bedienaufgabe betätigt die Versuchsperson den Berührungssensor, indem sie wieder an das Lenkrad greift, wodurch die Messung der Ausführungszeit endet. Nun wird die Versuchsperson aufgefordert, die Identität des eingeblendeten Diskriminations-Objekts zu nennen. Diese Angabe wird vom Versuchsleiter schriftlich festgehalten.

Abb. 5.5 zeigt den zeitlichen Auszug einer Protokoll-Datei, wobei die relevanten Einzelereignisse während einer Bedienaufgabe gekennzeichnet sind. Vergleicht man die Verläufe der x-Positionen von Soll- und Ist-Marke, so fällt in diesem Beispiel auf, dass die Versuchsperson während der Ausführung der Bedienaufgabe die Regelaufgabe gänzlich vernachlässigt. Des Weiteren ist ein interessanter Randeffekt erkennbar, welcher insgesamt sehr häufig beobachtet wurde und daher erwähnenswert erscheint, obwohl er nicht Gegenstand dieser Untersuchung ist: Es handelt sich hierbei um die Tatsache, dass offensichtlich allein das Vorlesen der Bedienaufgabe eine ablenkende Wirkung auf die Versuchsperson ausübt. Dies macht sich durch einen kurzen Abfall der Regelgüte bemerkbar, welcher sich in Form eines „Überschwingens“ äußert (siehe Abb. 5.5 zum Zeitpunkt 434 s).

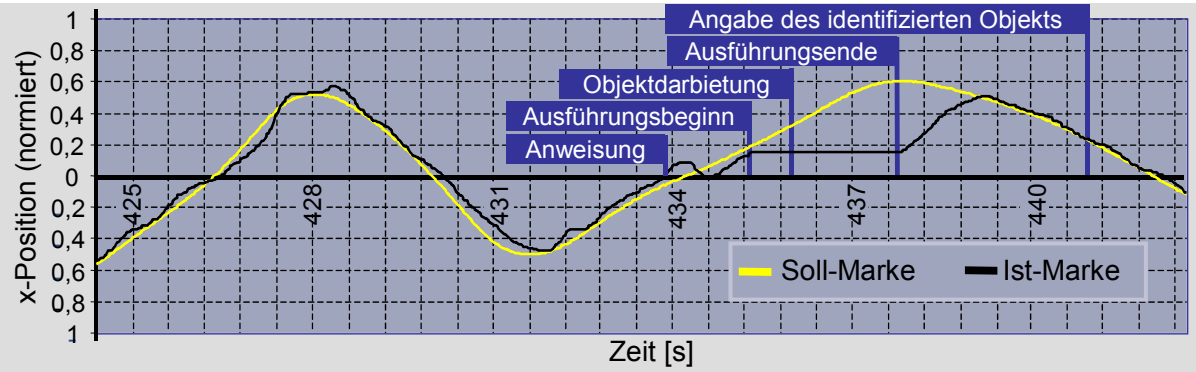


Abb. 5.5: Auszug aus einer aufgezeichneten Datensequenz: x-Positionen von Ist- und Soll-Marke sowie Ereignisse während einer Bedienaufgabe.

5.2.4 Problematik der Ausführungsverzögerung

Die Probanden wurden vor Versuchsbeginn dazu angehalten, die Bedienaufgaben schnellstmöglich und insbesondere ohne Unterbrechungen auszuführen. Während des Versuchsdesigns wurde jedoch eine Problematik offensichtlich, deren Auswirkungen sämtliche Messungen unbrauchbar macht:

Bereits nach kurzer Versuchsdauer wird von der Versuchsperson erkannt, dass einerseits jede Bedienaufgabe mit einer Objekteinblendung einhergeht sowie andererseits, dass dieses Ereignis durch das Loslassen des Berührungssensors getriggert wird. Die Versuchspersonen entwickeln nun anscheinend den Ehrgeiz, die eingeblendeten Diskriminations-Objekte um jeden Preis identifizieren

zu wollen. Dies führt dazu, dass der Bedienablauf nach dem Loslassen des Sensors - entgegen der eingangs erfolgten Anweisung - solange unterbrochen wird, bis die Objekteinblendung erfolgt (siehe Abb. 5.6), erst danach wird die Bedienaufgabe vollendet. Die nach den oben beschriebenen Methoden erfassten Daten verlieren somit jegliche Aussagekraft.

Abb. 5.6: Typischer zeitlicher Ablauf bei der verzögerten Ausführung einer Bedienaufgabe.

Zur Lösung dieser Problematik war die Einführung folgender Mechanismen notwendig:

- *Akustisches Feedback:* Um die Versuchsperson zu stimulieren, die Bedienaufgaben schnellst möglich auszuführen, wurde ein akustisches Feedback implementiert, welches nach jeder Aufgabenausführung in Form eines Sinus-Tons dargeboten wird. Dabei verhält sich die Frequenz des Tons (innerhalb gewisser Grenzen) proportional zur Ausführungsdauer, d.h. je mehr Zeit die Versuchsperson zur Ausführung benötigt, desto höher - und somit unangenehmer - ist die Frequenz des Sinus-Tons. Diese bewegt sich in einem Wertebereich von 280 Hz (sehr schnelle Ausführung) bis 2,8 kHz (sehr langsame Ausführung). Die Wiedergabedauer beträgt 1 s.
- *Unvorhersehbare Objekteinblendungen:* In der vorliegenden Studie ist die Messung der Objekterkennungsleistung *während* eines Bedienvorgangs von zentralem Interesse. Zur Vermeidung der oben aufgeführten Effekte gilt es jedoch, den Zusammenhang zwischen Bedienaufgabe und Objektdarbietung zu entkoppeln, d.h. für die Versuchsperson zu verschleiern. Dazu wurden weitere „unvorhersehbare“ Objekteinblendungen eingeführt: So können diese nun einerseits auch außerhalb der Bedienaufgaben auftreten[10]. Andererseits muss eine Bedienaufgabe nicht mehr zwangsläufig mit einer Objektdarbietung einhergehen. So wurde die ursprüngliche Anzahl an Nebenaufgaben verdoppelt, wobei jedoch lediglich die Hälfte der Aufgaben an eine Objektdarbietung gekoppelt ist.

Erst nach dem gemeinsamen Einsatz dieser Maßnahmen konnte das angestrebte Verhalten einer zügigen Aufgabenausführung der Versuchspersonen beobachtet werden. Das Versuchsszenario beinhaltet nun insgesamt 32 Nebenaufgaben: 16 haptische und 16 gestische Benutzereingaben, wobei jeweils die halbe Aufgabenanzahl mit einer Objektdarbietung einhergeht. Des Weiteren erfolgen im Laufe des Versuchs 16 Objekteinblendungen außerhalb von Bedienaufgaben. Daraus ergeben sich also 48 Ereignisse während einer Versuchsdurchführung, deren zeitliche Abfolge in Anh. A.1 aufgeführt wird. Insgesamt nahmen 25 Versuchspersonen an dieser Studie teil, woraus sich eine Da-

[10] Diese Maßnahme erlaubt in der späteren Auswertung zudem die Angabe von Referenzwerten bezüglich der Objekterkennungsleistung im Normalfall, also dann, wenn ausschließlich der Regelaufgabe nachgegangen wird.

tenbasis von 800 Benutzereingaben (bzw. 400 zur Auswertung der Objekterkennungsleistung) ergibt.

5.3 Ergebnisse

5.3.1 Statistische Datenauswertung

Ermittlung des Regelfehlers

Um objektive Aussagen über die Qualität machen zu können, mit der eine Versuchsperson die Regelaufgabe erfüllt, wird der Regelfehler über einen bestimmten Zeitraum berechnet. Dabei handelt es sich um den Effektivwert der Regelabweichung Δx (siehe auch Abb. 5.3), die zum Zeitpunkt n definiert wird als die Differenz der x-Positionen von Soll- und Ist-Marke:

$$\Delta x[n] = x_{soll}[n] - x_{ist}[n] \tag{5.1}$$

Somit ergibt sich für ein Zeitintervall, welches sich über N Abtastwerte erstreckt und zum Zeitpunkt $n = j$ beginnt folgender Regelfehler r :

$$r = \Delta x_{eff} = \sqrt{\frac{1}{N} \sum_{n=j}^{j+N-1} \Delta x^2[n]} \tag{5.2}$$

Für die Auswertung des Regelfehlers sind vorrangig diejenigen Zeitintervalle von Interesse, innerhalb derer Bedienaufgaben ausgeführt werden. Die jeweiligen Start- und Endzeitpunkte sind in der automatisch generierten Protokolldatei enthalten.

Die gespeicherten x-Positionsverläufe - und somit auch der Regelfehler - beziehen sich auf Bildschirmkoordinaten, d.h. sie liegen in der „Einheit“ *Pixel* vor. Möchte man sich den nach Gl. 5.2 ermittelten Regelfehler bildlich veranschaulichen, so müssen dabei die speziellen geometrischen Randbedingungen (gewählte Darstellungsauflösung, Ausmaße der Projektionsfläche, Abstand zwischen Versuchsperson und Leinwand etc.) des Versuch-Setups berücksichtigt werden. Um den Regelfehler davon unabhängig und anschaulich darstellen zu können, wird dieser nachfolgend in Form eines Winkels ρ ausgedrückt. Es handelt sich hierbei um jenen Winkel, unter welchem ein Betrachter die beiden Marken[11] aus $d_{ref} = 1$ m Entfernung sehen würde, wenn sie einen horizontalen Abstand zueinander aufweisen, welcher dem Regelfehler r entspricht (siehe Abb. 5.7).

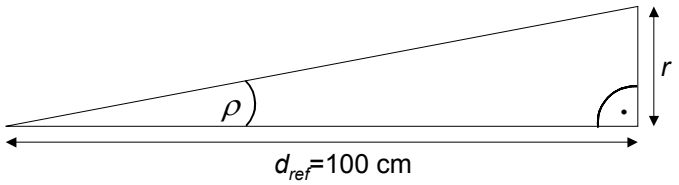


Abb. 5.7: Veranschaulichung des Regelfehlers r durch den Winkel ρ.

Der Abstand p zwischen zwei Pixeln auf der Projektionsleinwand wurde gemessen; er beträgt p = 0,086 cm. Somit gilt für den Winkel ρ folgender Zusammenhang:

[11] Eine der Marken befinde sich dabei gemäß Abb. 5.7 in der horizontalen Bildmitte (siehe auch Abb. 5.3).

$$\rho = \arctan\frac{r \cdot p}{d_{ref}} = \arctan(r \cdot 8,6 \cdot 10^{-4}) \tag{5.3}$$

Der Vorteil dieser Winkeldarstellung des Regelfehlers liegt nun darin, dass sich für beliebige Abstände zwischen dem Betrachter und einer gedachten Abbildungsebene sehr einfach ein hypothetisches Längenmaß Δx_{hyp} berechnen lässt, welches dem horizontalen Versatz der beiden dort projizierten Regelmarken entspräche. Zur Veranschaulichung ein Beispiel:

Für einen Regelfehler von $\rho = 2°$ ergibt sich nach Gl. 5.4 ein horizontaler Versatz von Δx_{hyp} = 3,5 cm bei d = 100 cm Betrachtungsabstand und entsprechend 3,5 m bei d' = 100 m.

$$\Delta x_{hyp} = d \cdot \tan\rho \tag{5.4}$$

Ermittlung der Ausführungszeit

Die Ausführung einer Bedienaufgabe erfolgt innerhalb des Zeitintervalls Δt. Mit der Abtastfrequenz f_{abt} erhält man N Abtastwerte und für die Ausführungszeit z ergibt sich somit:

$$z = \Delta t = N \cdot f_{abt} \quad \text{mit} \quad f_{abt} = 50 \text{ Hz} \tag{5.5}$$

Zur Ausgrenzung von unerwünschten Randeffekten werden für die Beurteilung der Ausführungszeiten nur diejenigen Bedienaufgaben betrachtet, bei denen *keine* Objektdarbietung stattfand.

Ermittlung der Objekterkennungsleistung

Die Bewertung der Objekterkennungsleistung erfolgt anhand der beobachteten Fehlerkennungsraten bei Objektdarbietungen innerhalb von Bedienaufgaben. Als Fehlerrate bzw. Fehlerhäufigkeit e wird hierbei wie üblich das Verhältnis der Anzahl o_{err} falsch erkannter Objekte zu der Gesamtzahl o_{ges} der dargebotenen Objekte definiert:

$$e = \frac{o_{err}}{o_{ges}} \tag{5.6}$$

Da die 2AFC-Methode eingesetzt wird, ist hierbei eine Ratewahrscheinlichkeit von 50 % zu berücksichtigen.

t-Test

Für den objektiven Vergleich der Eingabemodalitäten Haptik und Gestik wurde wie vorangehend beschieben ein Versuchsszenario erstellt, in welchem die interessierenden Einzelszenarien jeweils paarweise enthalten sind. Jedes dieser Paare besteht aus zwei Stichproben, nämlich einer für den haptischen sowie einer für den gestischen Bedienfall. Als optimaler Test für den Vergleich derart gepaarter Beobachtungen bietet sich der sogenannte t-Test (siehe [SAC02]) an. Um den Test für die drei betrachteten Einflussgrößen unter der Verwendung einheitlicher Hypothesen durchführen zu können, wurden diese so definiert, dass jeweils folgende Aussage gültig ist:

Je größer der Betrag der Beobachtungsgröße X ist, desto nachteiliger wirkt sich deren Eigenschaft auf den Fahrzeug-Einsatz der betrachteten Modalität aus.

Die Beobachtungsgröße X kann also in den folgenden drei t-Tests gleichgesetzt werden, mit der jeweiligen Einflussgröße:

$$X \triangleq \begin{cases} \text{Regelfehler } \rho \\ \text{Ausführungszeit } z \\ \text{Objekt-Fehlerkennungsrate } e \end{cases}$$

Für jede Beobachtungsgröße existieren pro Versuch acht Aufgabenpaare, von denen nun die Differenzen Ξ_p gebildet werden:

$$\Xi_p = X_p^{hapt} - X_p^{gest} \quad \text{mit} \quad p = \{1, \ldots, 8\} \tag{5.7}$$

Trägt man diese Differenzen für jede der drei Beobachtungsgrößen aus den 25 Versuchsdurchführungen in einer Verteilung an, so ergibt sich jeweils (augenscheinlich) eine Normalverteilung. Dies ist im Übrigen eine Voraussetzung für die Zulässigkeit der Verwendung des t-Tests[12]. Entsprechend der eingangs geäußerten Vermutung, dass die gestische Bedienung mit geringeren Ablenkungseffekten einhergeht als die haptische, werden basierend auf dieser einseitigen Fragestellung nun folgende Testhypothesen aufgestellt:

$$\begin{aligned} \text{Nullhypothese } H_0: &\quad \mu_\Xi = 0 \quad (\text{d.h.: } \mu_{X^{hapt}} = \mu_{X^{gest}}) \\ \text{Alternative } H_A: &\quad \mu_\Xi > 0 \quad (\text{d.h.: } \mu_{X^{hapt}} > \mu_{X^{gest}}) \end{aligned} \tag{5.8}$$

Die Nullhypothese H_0 lässt sich also folgendermaßen formulieren: „Der Mittelwert μ_Ξ der Verteilung der Differenzen aus haptischen und gestischen Beobachtungspaaren beträgt Null“ bzw. „Die Mittelwerte der Verteilungen der Beobachtungsgrößen für haptische und gestische Bedienung sind identisch“. Die Nullhypothese H_0 wird gegen die Alternative H_A getestet. Wird die Nullhypothese widerlegt, so gilt automatisch die Alternative H_A, welche besagt, dass aus der haptischen Bedienung im Mittel höhere Werte $|X|$ resultieren als bei der gestischen. Die Gültigkeit der Alternative wird dabei durch die Angabe einer Irrtumswahrscheinlichkeit, dem sogenannten Signifikanzniveau α_S, abgesichert.

5.3.2 Befunde

Die Resultate für alle drei Beobachtungsgrößen werden im Folgenden präsentiert und bewertet.

Regelfehler

Abb. 5.8 zeigt die Verteilungen des Regelfehlers ρ für haptische und gestische Bedienaufgaben. Sie weisen qualitativ ähnliche Formen auf, wobei jedoch die Verteilung für haptische Eingaben hin zu etwas höheren Werten verschoben ist.

---

[12] Neben der hier beschriebenen t-Test-Auswertung wurde auch ein *verteilungsfreies* Testverfahren angewandt, welches die vorliegenden Befunde in nahezu identischer Weise absichert.

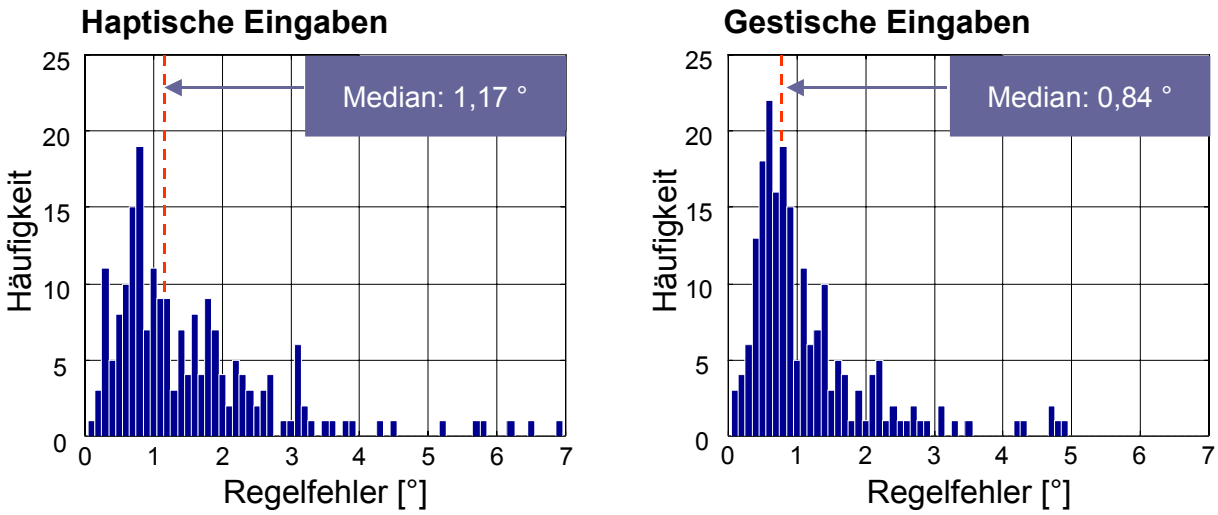


Abb. 5.8: Verteilung des Regelfehlers ρ für haptische (links) und gestische Eingaben (rechts).

Aufgrund des t-Tests kann die Nullhypothese H_0 mit einem Signifikanzniveau von $\alpha_S < 0{,}01$ verworfen werden und somit gilt die Alternative H_A. Dies deckt sich mit der augenscheinlichen Begutachtung der beiden Verteilungen: Der Median des Regelfehlers bei haptischen Eingaben liegt um etwa 28 % höher als der für gestische Eingaben. Für das $(1-\alpha_S)$-Intervall, in welchem sich der wahre Mittelwert μ_Ξ der Verteilung der Differenzen Ξ_p aufhält, ergibt sich: $\mu_\Xi \in [0{,}18°;\ 0{,}58°]$.

Zur Veranschaulichung: Ein Regelfehler von 0,58° entspricht nach Gl. 5.4 einem horizontalen Versatz Δx_{hyp} von etwa einem Meter bei einem Betrachtungsabstand von 100 Metern.

Um das Ausmaß der Ablenkung während der Benutzereingaben zu verdeutlichen, wird der mittlere Regelfehler, welcher sich *außerhalb* der Bedienaufgaben ergibt, betrachtet. Er beträgt etwa 0,3° und liegt damit deutlich unter den ermittelten Werten *während* der Bedienaufgaben (vgl. Abb. 5.8) sowohl für haptische als auch für gestische Eingaben.

Besonders interessante Erkenntnisse ergeben sich aus der Betrachtung des Regelfehlers innerhalb verschiedener zeitlicher Phasen der Bedienaufgabe. Dazu wird der Regelfehler berechnet, welcher sich in einem kurzen Zeitraum Δt_E (im Sinne einer „Momentaufnahme“) von jeweils 200 ms (d.h. 10 Abtastwerte bei f_{abt}=50 Hz) *vor* dem Zeitpunkt t_E des Eintretens folgender Ereignisse ergibt:

E1) t_1: Vorlesen der Nebenaufgabe; $\Delta t_1 = [t_1-200\text{ ms}, t_1-180\text{ ms}, ..., t_1]$

E2) t_2: Ausführungs-Beginn (Loslassen des Lenkrades); $\Delta t_2 = [t_2-200\text{ ms}, t_2-180\text{ ms}, ..., t_2]$

E3) t_3: Ausführungs-Ende (Greifen des Lenkrades); $\Delta t_3 = [t_3-200\text{ ms}, t_3-180\text{ ms}, ..., t_3]$

Dazu werden die Daten aller (800) Bedienaufgaben nach der jeweiligen Eingabemodalität getrennt ausgewertet. Die Resultate sind in Abb. 5.9 dargestellt. Bei den hier angegebenen Werten handelt es sich um die Mittelwerte des modalitätenspezifischen Regelfehlers innerhalb der betrachteten Zeitintervalle Δt_1, Δt_2 und Δt_3.

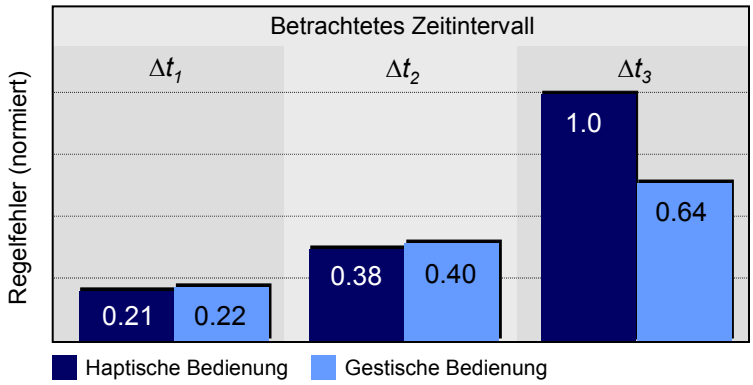


Abb. 5.9: Mittelwert des Regelfehlers ρ innerhalb verschiedener zeitlicher Ausführungsphasen der Bedienaufgaben für haptische und gestische Bedienung (alle Werte wurden auf das auftretende Maximum $\rho_{hapt}[\Delta t_3] = 1{,}7°$ normiert).

Erwartungsgemäß ergeben sich für den Regelfehler vor dem Beginn des Vorlesens der Bedienaufgabe (Δt_1) unabhängig von der Eingabemodalität (nahezu) identische Werte.

Diese Aussage gilt auch für das Zeitintervall Δt_2 kurz vor Ausführungsbeginn. Bemerkenswert ist hierbei jedoch die Tatsache, dass offensichtlich allein die mentale Verarbeitung der Aufgabenstellung deutliche Ablenkungseffekte bewirkt: Im Vergleich zum vorangehenden Zeitintervall Δt_1 verdoppelt sich hier der Regelfehler beinahe.

Der Regelfehler innerhalb des Zeitraums Δt_3 gibt Auskunft über das Ausmaß der Regelabweichung, mit der die Versuchsperson eine Bedienaufgabe „verlässt“. Die Aussage der obigen Testbefunde wird hier abermals bestätigt: Der Regelfehler bei haptischer Bedienung übertrifft den der gestischen Bedienung um 36 % (vgl. Abb. 5.9).

Die hier gewählte Definition des Regelfehlers in Form eines horizontalen Versatzes bzw. eines zugehörigen Winkels erfolgte mit dem Ziel, ein objektives Maß für die Regelgüte insbesondere während der Tätigung von Benutzereingaben zu erhalten. Zwar ist es nicht möglich bzw. zulässig, dieses Maß auf reale Parameter einer Autofahrt (wie z.B. Spurhaltung) abzubilden. Es ist jedoch festzuhalten, dass die Qualität, mit der ein Benutzer generell eine Regelaufgabe erfüllt, während haptischer Eingaben deutlich stärker abfällt als bei gestischen.

Ausführungszeit

Die modalitätenspezifischen Verteilungen der Ausführungszeiten sind in Abb. 5.10 dargestellt. Aus den angegebenen Medianen geht hervor, dass haptische Eingaben im Mittel etwa um den Faktor 1,4 länger dauern als gestische. Dies bestätigt auch der t-Test, welcher die Alternative H_A mit einem Signifikanzniveau von $\alpha_S < 0{,}01$ stützt. Für das $(1-\alpha_S)$-Intervall ergibt sich $\mu_\Xi \in [0{,}36\text{ s};\ 0{,}57\text{ s}]$, d.h. haptische Eingaben dauern im Mittel um mindestens 0,36 s länger als gestische.

Somit erweist sich die Gestik hier als die effizientere Eingabemethode. Bezüglich der Fahrerablenkung ist sie der Haptik vorzuziehen, da die Gesamtzeit möglicher Ablenkung geringer ist.

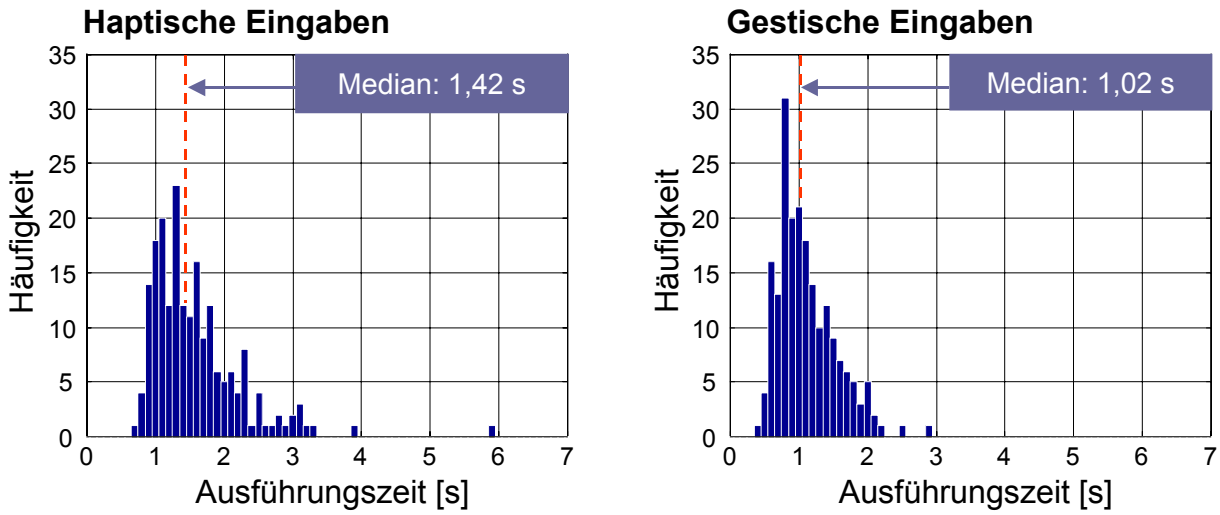


Abb. 5.10: Verteilung der Ausführungszeit *z* für haptische (links) und gestische Eingaben (rechts).

Objekterkennungsleistung

Auch bei der Betrachtung der Objekt-Fehlerkennungsraten (siehe Abb. 5.11) zeigt sich die gestische Eingabemethode der haptischen überlegen. Vergleicht man die Mediane, so wird ersichtlich, dass die haptische Bedienung deutlich stärker zu Lasten der Objekterkennungsleistung erfolgt als die gestische. Diese Aussage stimmt mit dem Resultat des t-Tests überein: Die Gültigkeit der Alternative H_A wird auch hier durch ein Signifikanzniveau von $\alpha_S < 0{,}01$ abgesichert; für das $(1-\alpha_S)$-Intervall ergibt sich $\mu_\Xi \in [0{,}08;\ 0{,}32]$.

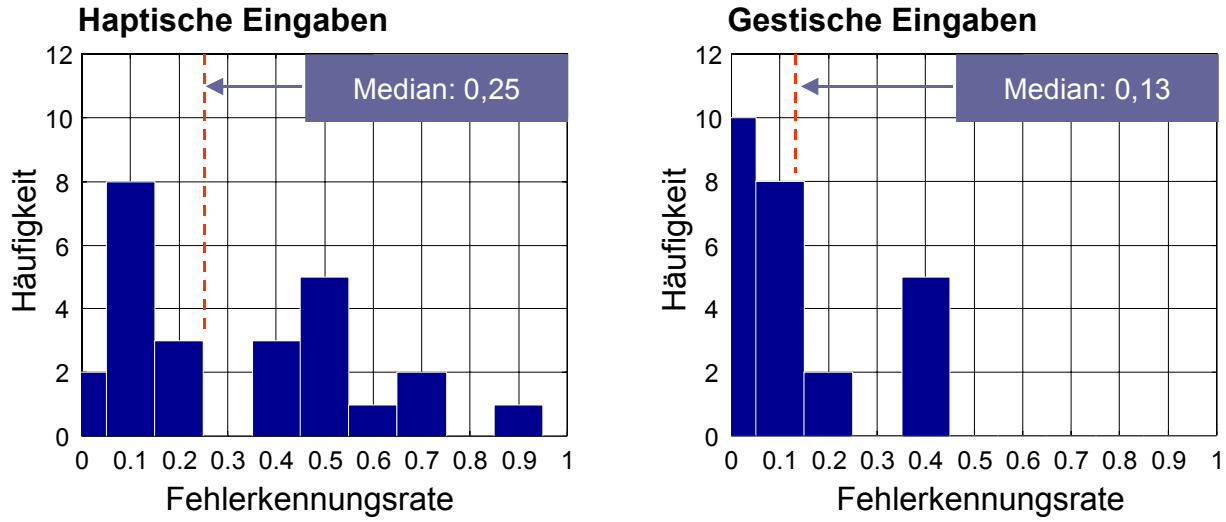


Abb. 5.11: Verteilung der Fehlerkennungsrate *e* für haptische (links) und gestische Eingaben (rechts).

Wie bereits erwähnt, muss aufgrund der Verwendung der 2AFC-Methode berücksichtigt werden, dass eine Fehlerkennungsrate von 0,5 gleichbedeutend ist mit dem Resultat reinen Ratens. Dieser Wert wird bei der haptischen Bedienung von einigen Versuchspersonen sogar deutlich überschritten (siehe Abb. 5.11). Darüber hinaus muss angemerkt werden, dass die Objekterkennungsleistung bei gestischen Eingaben nahezu mit jener übereinstimmt, welche sich bei Objektdarbietungen *außer-*

halb von Bedienaufgaben ergibt: Für diese Fälle wurde eine mittlere Fehlerkennungsrate von 0,12 (vgl. e_{med} = 0,13 bei gestischen Eingaben) ermittelt.

Die Position der visuellen Aufmerksamkeit des Fahrers weicht also während der Tätigung haptischer Eingaben offensichtlich stark vom primären Sichtfeld - d.h. vom Verkehrsgeschehen - ab. Dieser schwerwiegende Nachteil konnte bei der gestischen Bedienung nicht beobachtet werden.

Subjektive Beurteilung

Die subjektive Beurteilung der untersuchten Eingabemodalitäten erfolgt in hoher Übereinstimmung mit den objektiven Befunden. Wie aus Abb. 5.12 entnommen werden kann, wird die gestische Bedienung sowohl deutlich weniger ablenkend als auch angenehmer empfunden.

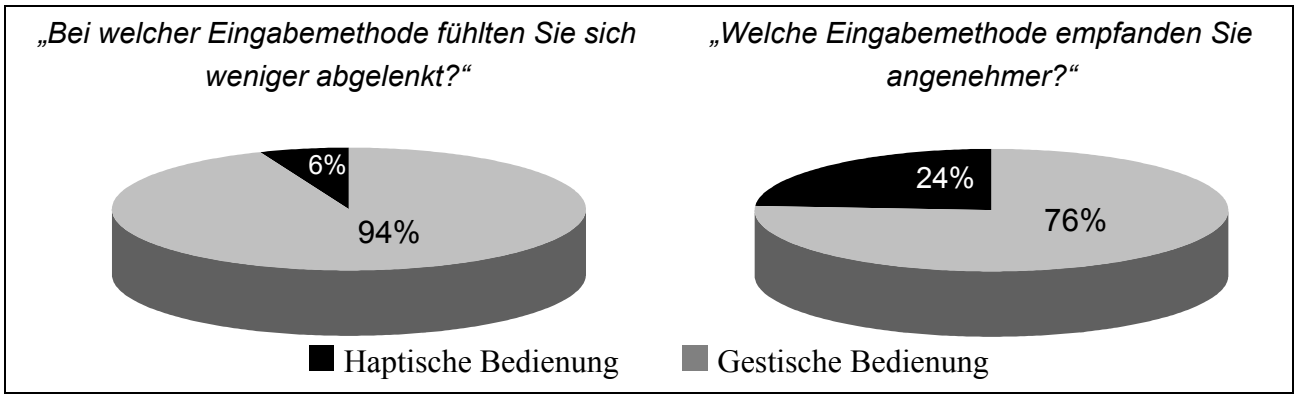


Abb. 5.12: Subjektives Urteil der Versuchspersonen.

5.3.3 Schlussfolgerungen

Die Befunde der vorliegenden Studie sprechen sehr deutlich für den Einsatz der Gestik als Eingabemodalität im Fahrzeug. Hinsichtlich der untersuchten Einflussgrößen zeigten sich signifikant geringere Ablenkungseffekte im Vergleich zu haptischen Eingaben.

Darüber hinaus kann aufgrund der positiven Beurteilung der Gestik auf einen entsprechenden Gewinn an Benutzerakzeptanz gehofft werden, wenn diese Eingabemodalität tatsächlich für die Bedienung bestimmter Funktionen im Fahrzeug eingesetzt wird. Die entsprechende Realisierung durch die Bedienumgebung GECOM wird im nachfolgenden Kapitel detailliert beschrieben.

6

Gestische Bedienumgebung GECOM

6.1 Motivation

Die Befunde grundlegender Untersuchungen (siehe Kap. 4.2.4) deuten darauf hin, dass die Nutzung der Gestik als Eingabemodalität den Mensch-Maschine-Dialog im Fahrzeug entscheidend verbessern kann. So lässt sich für die Bedienung zahlreicher Funktionen ein „Vokabular“ aus diskreten Handgesten definieren, welches mit großer interpersoneller Übereinstimmung intuitiv angewandt wird. Dabei zeigt sich eine starke Abhängigkeit der beobachteten Gestentypen vom strukturellen Aufbau der verwendeten Bedienoberfläche. Diese Erkenntnis kann zur Unterstützung der gestischen Bedienung durch bestimmte Gestaltungsmethoden gezielt genutzt werden. Darüber hinaus weist die gestische Interaktion in mehrfacher Hinsicht Vorteile gegenüber anderen Eingabemethoden wie Sprache und Haptik auf: So können Eingaben durch Gestik sehr direkt und intuitiv erfolgen - hierzu zwei Beispiele:

1) Ein eingehender Anruf kann entweder durch eine horizontale Wischbewegung mit der Hand oder durch einfaches Kopfschütteln abgelehnt werden.
2) Der Aufbau eines Telefongesprächs kann durch das imitierte Abheben eines Telefonhörers eingeleitet werden.

Schließlich konnte gezeigt werden, dass die gestische Bedienung mit signifikant geringeren Ablenkungseffekten einhergeht als die haptische und zudem breite Akzeptanz findet (siehe Kap. 5).

Diese Ergebnisse motivieren die Entwicklung einer speziell für die gestische Bedienung optimierten Anwendungsumgebung, um das Potenzial dieser Modalität im Fahrzeug anhand eines realen Systems detailliert untersuchen zu können. Hauptziel ist hierbei die Bereitstellung eines fahrzeugtypischen Infotainment-Systems, dessen gesamter Funktionsumfang allein durch gestische Interaktion verfügbar sein soll. Dies geschieht nicht mit der Absicht, ein derartiges System später in exakt dieser Form im Fahrzeug einzusetzen. Es soll zunächst lediglich erforscht werden, in welchen Funktionsbereichen diese Eingabemodalität besondere Vorzüge gegenüber herkömmlichen Bedienmethoden aufweist. Die dabei gewonnenen Erkenntnisse sollen die optimalen Einsatzmöglichkeiten der Gestik in *multimodalen* Infotainment-Systemen aufzeigen. In diesem Rahmen entstand die pro-

totypische Bedienumgebung GECOM (*Ge*sture *Co*ntrolled *M*MI), welche in diesem Kapitel näher vorgestellt wird. Dieses gestenoptimierte Fahrzeug-MMI stellt durch die nachfolgende Implementierung einer sprachverstehenden Komponente einen ersten Ansatz für ein derartiges multimodales Bedienkonzept dar.

6.2 Entwicklungsstadien

Im Zuge einer schrittweisen Erweiterung der bereitgestellten Eingabekanäle durchlief GECOM mehrere Entwicklungsstadien:

1) *Diskrete Handgestik*
   Ursprünglich wurde GECOM für die Bedienung durch diskrete Handgesten ausgelegt. Hierfür wurde das experimentell gewonnene intuitive Gesteninventar (siehe Tab. 4.1) integriert. Die Bedienoberfläche wurde sodann im Laufe mehrerer Benutzertests für diese Eingabemodalität optimiert. Die Anbindung einer haptischen Bedienkonsole erfolgte zu Vergleichszwecken, um die neuartige gestische Bedienung und die konventionelle haptische Bedienung objektiv gegeneinander evaluieren zu können.

2) *Kontinuierliche Handgestik und Kopfgestik*
   Durch die Implementierung der kontinuierlichen Handgestik sollten die speziellen Vorteile dieses Gestentyps (siehe Kap. 4.3) zur direkten Beeinflussung von Systemparametern genutzt werden. Dadurch gelingt z.B. das stufenlose Einstellen der Musiklautstärke, welches weder durch diskrete Handgesten noch durch Spracheingaben möglich ist.

   Zusätzlich bot es sich an, das intuitive Beantworten von Systemrückfragen - z.B. „Soll das eingegebene Ziel in den Zielspeicher aufgenommen werden?“ - durch Kopfgestik bereitzustellen.

3) *Natürliche Sprachbedienung*
   Obwohl GECOM zunächst für die gestische Bedienung optimiert wurde, geht das nachträgliche „Aufsetzen“ einer sprachverstehenden Komponente aus folgenden Gründen nicht zu Lasten der intuitiven Bedienbarkeit: Während die gestische Bedienung in hohem Maße von der Art der grafischen Darstellung unterstützt werden kann, ist dieser Aspekt bei der Sprachbedienung von sehr untergeordneter Bedeutung. Es müssen jedoch auch hierbei nach [Neu00] bestimmte Grundregeln beachtet werden. Demnach spielen z.B. die Beschriftungen von Schaltflächen eine sehr wichtige Rolle, da sie dem Benutzer einen einfachen Einstieg in den Umgang mit dem implementierten Sprachwortschatz ermöglichen. Wie sich herausstellt, sind die Gestaltungskriterien für die sprachliche Bedienung im vorliegenden Falle durchgehend erfüllbar, ohne mit denen für die gestische Interaktion im Widerspruch zu stehen. Durch die Bereitstellung der Sprachsteuerung wird dem Benutzer die Anwendung von *Shortcuts* ermöglicht, d.h., er kann bestimmte Systemzustände durch gezielte Äußerungen direkter erreichen als mit Gestik. Hierzu ein Beispiel: Mit der Spracheingabe *„Bring mich nach München zum Hauptbahnhof!“* gelangt der Benutzer direkt in das entsprechende Menü des Navigationssystems, wobei das angegebene Ziel bereits angezeigt wird.

6.3 Funktionsumfang

In Anlehnung an integrierte Fahrzeug-MMIs, wie sie bereits heute kommerziell verfügbar sind (siehe Kap. 2), vereint GECOM fahrzeugtypische Komponenten der Informations- und Unterhaltungselektronik. Abb. 6.1 gibt einen groben Überblick über die implementierten Hauptfunktionen sowie deren Einbettung in eine hierarchische Menüstruktur.

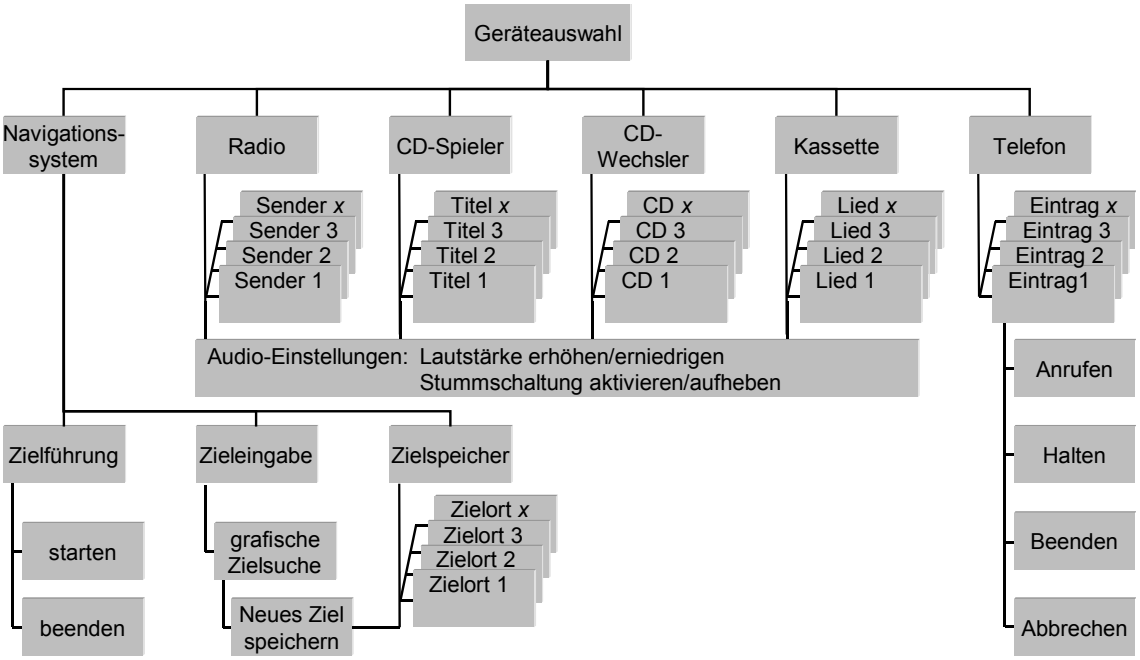


Abb. 6.1: Schematische Darstellung des Funktionsumfangs und der Menüstruktur von GECOM.

Die Hierarchie wurde hierbei bewusst möglichst flach angelegt, woraus eine entsprechend breite Struktur resultiert. Diese parallele Bereitstellung vieler Funktionen kann sich im schlimmsten Falle nachteilig auf die Übersichtlichkeit auswirken. Sie wurde dennoch einer tiefen und schmalen Hierarchie vorgezogen, da der Benutzer aufgrund der geringen Menütiefe schnell lernt, auf welcher Ebene sich eine bestimmte Funktion befindet.

6.4 Gestaltungskriterien für die gestische Bedienung

6.4.1 Visualisierung

Zur Visualisierung der grafischen Bedienoberfläche dient ein TFT-Farbdisplay (etwa 10 Zoll bzw. 25 cm Bilddiagonale), welches gemäß Abb. 6.2 an der Mittelkonsole angebracht ist. Die nachfolgend beschriebenen Visualisierungsmerkmale dienen der Unterstützung des *ungeübten* Benutzers - etwa beim Erstkontakt mit GECOM. Dies bedeutet keineswegs, dass die gestische Bedienung nur unter Blickkontakt zum Display erfolgen kann. Der Benutzer wird dadurch vielmehr sehr effizient in die Lage versetzt, sich in kurzer Lernzeit ein korrektes System-Modell zu verschaffen und zu verinnerlichen - eine wichtige Voraussetzung für die im Fahrzeug angestrebte Blindbedienbarkeit (Bedienung ohne Blickabwendung von der Straße).

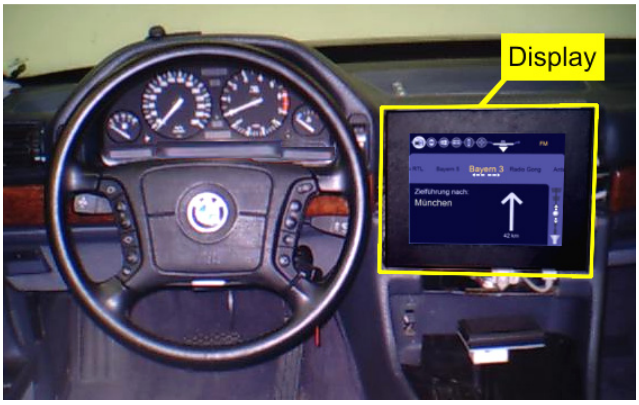


Abb. 6.2: Anordnung des Displays im Fahrzeug-Cockpit.

Grundlayout

Wie in Kap. 4.2.3 beschrieben, lässt sich die gestische Bedienung durch die grafische Darbietung der Bedienelemente stark beeinflussen. Diese Erkenntnis wird bei der Entwicklung von GeCoM ausgenutzt, indem einerseits die Verwendung bestimmter Gestentypen durch die Orientierung der Interaktionselemente, d.h. deren Ausrichtung auf dem Display, nahegelegt wird. Andererseits wird bei der Strukturierung des Layouts darauf geachtet, Mehrdeutigkeiten zu vermeiden. Dies bedeutet beispielsweise, dass niemals mehrere gestisch bedienbare Elemente mit derselben Ausrichtung gleichzeitig dargestellt werden. Für die Festlegung eines Grundlayouts werden folgende Befunde aus den Voruntersuchungen (siehe Kap. 4.2.3) berücksichtigt:

(1) Kinemimische Gesten zur Menübedienung orientieren sich nahezu ausschließlich an der Ausrichtung der sichtbaren Eingabeelemente.

(2) Wird keine Bedienoberfläche vorgegeben, stimmen die beobachteten Gesten in folgenden Punkten weitgehend überein:

- Die Ausführung einer Handbewegung nach rechts (z.B. Winken oder Zeigen nach rechts) erfolgt im Sinne von „weiter zur nächsten Funktion“; entsprechend werden Linksbewegungen für „zurück zur vorigen Funktion“ eingesetzt.

- Um eine Regelgröße zu erhöhen bzw. zu erniedrigen werden Auf- bzw. Abwärtsbewegungen verwendet.

Kombiniert man diese Erkenntnisse, so liegt es im Sinne einer möglichst intuitiven Bedienbarkeit nahe, sequentielle Auswahlmenüs horizontal auszurichten: Die Bedienung durch horizontale, kinemimische Gesten wird somit einerseits durch die offensichtliche Darstellung gestützt (1) und deckt sich darüber hinaus mit der gestischen Bedienung, welche sich auch im abstrakten Sinne als intuitiv erweist (2). Aufgrund derselben Überlegung werden Elemente, bei denen Wertänderungen vorgenommen werden können, vertikal ausgerichtet. Im vorliegenden Einsatzbereich findet dies etwa zur Einstellung der Musiklautstärke Anwendung.

Unter Berücksichtigung dieser grundlegenden Gestaltungskriterien entstand im Laufe mehrerer Test- und Redesign-Zyklen das in Abb. 6.3 dargestellte Layout. Es existieren zwei primäre Bereiche

für die gestische Interaktion, welche orthogonal zueinander angeordnet sind: Eine horizontale Fläche für die sequentielle Menü-Navigation sowie ein vertikaler Bereich zur Bedienung der Lautstärke.

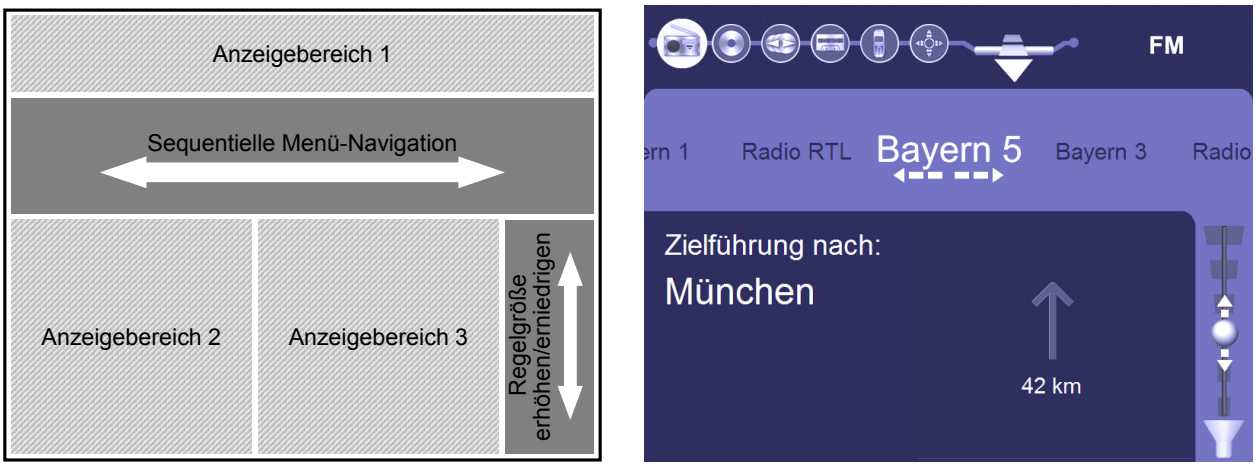


Abb. 6.3: Grundlayout der GECOM-Bedienoberfläche (links) und grafische Umsetzung (rechts) am Beispiel *Radio* bei aktiver Zielführung (Navigationssystem).

Darüber hinaus sind drei zusätzliche Bereiche für die Ausgabe gerätetypischer Informationen vorgesehen. Hierbei wird generell darauf geachtet, einerseits die Quantität zeitgleich dargestellter Informationen so gering wie möglich zu halten sowie diese andererseits thematisch konsistent auf den reservierten Flächen darzubieten. Es werden hier folgende Inhalte angezeigt:

- *Anzeigebereich 1:* Statuszeile für aktives Gerät (farblich hervorgehobenes Piktogramm) und dessen Zustand (z.B. Radio, FM) sowie die weiteren verfügbaren Geräte.
- *Anzeigebereich 2:* Textinformation über im Hintergrund laufende Prozesse (z.B. aktueller Zielort des Navigationssystems bei laufender Zielführung) und Anzeige von grafischer Hilfeinformation (siehe Kap. 6.4.4).
- *Anzeigebereich 3:* Darstellung von Navigationshinweisen bei laufender Zielführung bzw. einer Landkarte für die Zielprogrammierung.

Generell werden alle gestisch manipulierbaren Elemente durch eine einheitliche Farbgebung (weiß[13]) hervorgehoben. Wie bereits erwähnt, entspricht das implementierte Gesteninventar jenem, welches sich aufgrund zahlreicher Untersuchungen als intuitiv erwies. Eine Ausnahme stellt hierbei die Geste zur Aktivierung des Hauptmenüs dar. Da einerseits für eine derartige Funktion keine intuitive Geste beobachtet werden konnte, andererseits jedoch die rein gestische Bedienbarkeit des gesamten Systems angestrebt wurde, musste für diese Funktion eine „künstliche“ Geste eingeführt werden. Es handelt sich hierbei um das virtuelle Ziehen an einem Griff, welcher gemäß Abb. 6.3 im *Anzeigebereich 1* räumlich dargestellt wird.

[13] Manche der in Abb. 6.3 dargestellten Schriftzüge (z.B. *Zielführung nach: München*) tragen hier lediglich zum Zweck der besseren Lesbarkeit in der Graustufen-Darstellung die Textfarbe *weiß*.

Kontinuierliche und diskrete Visualisierung

Wie sich in weiteren Untersuchungen zeigte (siehe Kap. 4.3), spielt neben der Ausrichtung der Bedienelemente auch die Art ihrer Darstellung eine wichtige Rolle hinsichtlich der zur Bedienung beobachteten Gesten: So erfolgt die gestische Interaktion mit *diskreten* Strukturen, wie z.B. einzelnen Menüpunkten mit Auswahlcursor, nahezu ausschließlich durch *diskrete* Gesten. Entsprechend wird bei einer *kontinuierlichen* Darstellung, wie z.B. in Form eines Schiebereglers, überwiegend *kontinuierliche* Gestik beobachtet. Dieser Befund wird bei GeCoM berücksichtigt, indem die Darstellungsart auf den jeweils als optimal erachteten Gestentyp angepasst ist. Zur sequentiellen Navigation durch relativ kurzer Menüs bietet sich die Bedienung mittels diskreter Gesten an. Die einzelnen Einträge werden hierbei in Textform dargestellt, wobei die jeweilige Auswahl durch einen Cursor markiert ist und zudem farblich sowie durch Vergrößerung hervorgehoben wird (siehe Abb. 6.4 und Abb. 6.5).

Abb. 6.4: Darstellungsart für die sequentielle Menü-Navigation mittels diskreter Gesten.

Die einzelnen Menüpunkte bilden einen „geschlossenen Ring“, d.h., bei wiederholtem Wechsel zum nächsten Menüpunkt gelangt man nach dem Erreichen des letzten Eintrages wieder zum Startpunkt. Innerhalb der horizontalen Interaktionsfläche werden maximal fünf Menüpunkte gleichzeitig dargestellt. Dies bedeutet, dass manche Einträge bei längeren Menüs den sichtbaren Displaybereich verlassen. Ihr Vorhandensein wird dem Benutzer bewusst gemacht, indem sie entsprechend „abgeschnitten“ an den Displayrändern erscheinen (siehe Abb. 6.4).

Um dem Benutzer den Zusammenhang zwischen Ursache (Geste) und Wirkung (Systemreaktion) zu verdeutlichen, werden grafisch animierte Zustandswechsel eingesetzt. Betrachtet man den in Abb. 6.4 dargestellten Systemzustand, so bedeutet dies beispielsweise, dass der Cursorpfeil des Interaktionsfeldes nach einer Rechtsgeste nicht von *Bayern 5* zu *Bayern 3* „springt“, sondern durch eine angedeutete Bewegung in die entsprechende Richtung animiert ist. Der neue Menüpunkt wird daraufhin vergrößert und farblich hervorgehoben, während die vorige Auswahl entsprechend in den Hintergrund rückt. Die Menüleiste zentriert sich nach erfolgter Auswahl automatisch in Form einer kontinuierlichen Animation, so dass sich der aktuelle Menüpunkt im Ruhezustand stets in der Mitte befindet.

Abb. 6.5: Unterstützung des optimalen Gestentyps durch grafische Darstellung.

Ein Nachteil der diskreten Handgestik wird bei der Navigation durch lange Listen offensichtlich. Dies ist beispielsweise bei der Auswahl eines Namens aus dem implementierten Telefonbuch der

Fall: Hierbei ergeben sich vorwiegend negative Benutzerkritiken, da die Auswahl eines Eintrages über eine große Distanz zwischen dem gegebenen Anfangsbuchstaben („A“) und dem gewünschten Namen durch sehr viele Einzelgesten sehr umständlich ist.

Daher wurde in der zweiten Ausbauphase die Möglichkeit der kontinuierlichen Gestenbedienung implementiert. Diese kann bei langen Listen in Kombination mit diskreter Gestik angewendet werden. Dem Benutzer wird dies durch ein entsprechendes kontinuierliches Bedienelement (siehe Abb. 6.5) visualisiert: Es handelt sich um eine Art Schieberegler, dessen aktueller Zustand durch eine plastische Kugel angezeigt wird. Diese kann durch eine Greifgeste „angefasst“ werden, welche den kontinuierlichen Bedienmodus einleitet. Die Kugel wird in diesem Systemzustand vergrößert dargestellt (siehe Abb. 6.6), während sowohl die Hervorhebung des diskreten Auswahlmenüs als auch der zugehörige Cursorpfeil ausgeblendet werden. Mittels kontinuierlicher Handbewegungen längs der Reglerschiene kann der Benutzer nun den gewünschten Anfangsbuchstaben anwählen. Der kontinuierliche Modus wird beendet, wenn die Hand wieder geöffnet wird bzw. wenn sie den Erkennungsbereich verlässt, woraufhin sich das System wieder im diskreten Modus befindet.

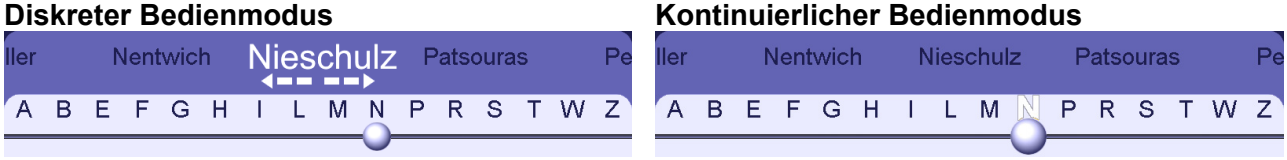


Abb. 6.6: Kombination von diskreter und kontinuierlicher Gestenbedienung: Diskreter Modus (links) und kontinuierlicher Modus (rechts).

Darüber hinaus zeigte sich die kontinuierliche Gestik als sehr geeignet für das stufenlose Einstellen einer Regelgröße und fand hierbei hohe Akzeptanz (siehe Kap. 4.3). Diese Bedienmöglichkeit wird in der zweiten Ausbauphase von GECOM für das Regeln der Musiklautstärke implementiert. Während derartige Einstellungen unter Verwendung diskreter Gestik oder Sprachbedienung nur schrittweise mit diskreten Änderungsintervallen erfolgen kann, erlaubt die kontinuierliche Gestik das exakte Einregeln des gewünschten Zielwerts. Auch in diesem Fall werden die beiden Bedienmodi kombiniert angeboten. Die entsprechende grafische Umsetzung geht aus Abb. 6.7 hervor.

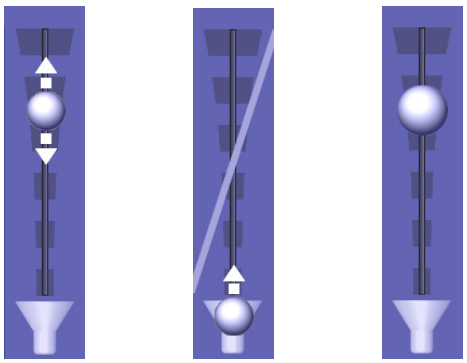


Abb. 6.7: Kombinierte Bedienung der Lautstärke. Im Grundzustand (links) und bei aktiver Stummschaltung (Mitte) ist sowohl die diskrete als auch die kontinuierliche Bedienung möglich. Rechts: Aktiver kontinuierlicher Modus mit vergrößerter Kugeldarstellung.

Ferner kann auch die Eingabe eines Navigationszieles in der grafisch dargestellten Landkarte durch kombinierte Anwendung der beiden Gestentypen erfolgen. Dies wird visualisiert durch diskrete Pfeile und ein Fadenkreuz, welches (wie die Kugel) „gegriffen“ und kontinuierlich verschoben werden kann.

Abb. 6.8: Grafische Auswahl eines Navigationsziels durch diskrete und kontinuierliche Gesten. GECOM (links) und Fadenkreuz-Cursor im diskreten (rechts oben) bzw. vergrößert im kontinuierlichen Modus (rechts unten).

Die Auswahl eines Navigationsziels erfolgt, indem der Fadenkreuz-Cursor auf den gewünschten Zielort bewegt wird. Der Kartenmaßstab kann durch diskrete Gesten (Winken nach vorne/hinten = Karte verkleinern/vergrößern, vgl. Tab. 4.1) stufenweise verändert werden, wobei sich der Umfang der dargestellten Details automatisch an die jeweilige Stufe anpasst.

Visualisierung bei Kopfgestik

Die Möglichkeit der Eingabe mittels Kopfgesten wird durch ein Kopfsymbol (siehe Abb. 6.9 rechts) angezeigt. In der vorliegenden Anwendung können Kopfgesten zur Beantwortung von Systemrückfragen eingesetzt werden. Ein typisches Beispiel hierfür zeigt Abb. 6.9 (links); nach der Auswahl eines Zielorts wird der Benutzer gefragt, ob dieses Ziel in den Zielspeicher aufgenommen werden soll. Alternativ zur Kopfgestik kann diese Frage auch mittels Handgestik durch die Auswahl des entsprechenden Menüpunktes erfolgen.

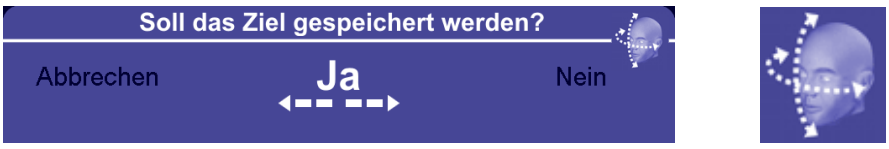


Abb. 6.9: Darstellung bei Kopfgestik. Systemrückfrage (links) und Kopfgesten-Symbol (rechts).

Neben der Beantwortung von Systemrückfragen kann die Kopfgestik auch zum Annehmen bzw. Ablehnen eines eingehenden Telefonanrufes angewendet werden.

6.4.2 Akustisches Feedback

Der Einsatz akustischer Informationsausgaben erweist sich als wichtige Unterstützung der Blindbedienbarkeit. Hierbei wird beabsichtigt, dem Benutzer Aufschluss über den aktuellen Systemzustand - insbesondere bei Zustandsübergängen während der Bedienung - zu geben, ohne dass dieser den Blick von der Strasse abwenden muss. Als akustisches Feedback werden einerseits die ausgewählten Menüpunkte bzw. Systemrückfragen stets in gesprochener Form dargeboten. Andererseits werden einfache Signaltöne („Earcons“; z.B. Klickgeräusch = Gestenerkenner hat den Beginn einer Eingabe detektiert) eingesetzt. Derartige Rückmeldungen stellen aus folgendem Grund sehr wichtige Maßnahmen zur Vermeidung von Blickabwendungen dar: Erhält der Benutzer unmittelbar nach Beginn bzw. Ende einer gestischen Eingabe kein Systemfeedback, so reagiert dieser üblicherweise innerhalb weniger hundert Millisekunden mit Kontrollblicken in Richtung des Displays. Das Ausbleiben derartiger Benutzerreaktionen aufgrund akustischer Feedbacks konnte in mehreren Untersuchungen im institutseigenen Fahrsimulator belegt werden. Es ist jedoch insbesondere bei häufig auftretenden Klangereignissen unbedingt darauf zu achten, kurze bzw. dezente Signalgeräusche einzusetzen, da diese beim Benutzer ansonsten wenig Akzeptanz finden. Als Negativbeispiel seien hier stark tonhaltige Klänge wie z.B. Sinustöne genannt.

6.4.3 Konsistenz

Wie weiter oben angesprochen, ging die Entwicklung von GECOM mit schritthaltenden Test- und Redesign-Zyklen einher. Dabei wurden zahlreiche verschiedene Teilkonzepte getestet, indem jeweils relativ kleine (< 10) Versuchspersonengruppen mit der Bedienumgebung konfrontiert wurden. Diese Vorgehensweise ist gerechtfertigt, da die Erfahrung zeigt, dass sich die gravierendsten Mängel einer Bedienoberfläche bereits bei einer sehr geringen Personenanzahl aufdecken lassen (siehe auch [NEU00]).

Hierbei stellte sich heraus, dass gerätespezifische Um- und Neubelegungen der gestisch bedienbaren Funktionen zu vermeiden sind, da sie häufig zur Verwirrung des Benutzers führen. Dazu ein Beispiel: Zur Änderung der Lautstärke werden bei GECOM vertikale Handgesten eingesetzt. In jenen Funktionsgruppen, in denen keine Musiklautstärke eingestellt werden kann (z.B. Telefon) könnte man nun die vertikale Gestik etwa zur Menüsteuerung zusätzlich zur horizontalen Gestik verwenden. Dies führt jedoch aus Sicht des Benutzers zu einer deutlich erschwerten Bedienbarkeit. Um die gestische Bedienung des Systems möglichst transparent zu gestalten, wurde bei GECOM insbesondere auf Konsistenz geachtet. Das bedeutet neben der eindeutigen Zuordnung von Geste zu Systemreaktion, dass *jede* implementierte Geste *jederzeit* - soweit sinnvoll - verwendbar ist. Darüber hinaus enthält das Bedienkonzept eine gewisse Redundanz, d.h. in manchen Fällen führen verschiedene Bedienabfolgen (bestehend aus Einzelgesten) zum selben Ziel.

6.4.4 Audiovisuelles Hilfesystem

Da es sich bei der gestischen Bedienung um eine neue und somit ungewohnte Eingabemethode handelt, liegt es nahe, dem Benutzer bei Bedarf Hilfeinformationen bereit zu stellen. Dabei hat der Benutzer derzeit zwei Möglichkeiten, Hilfe anzufordern, nämlich einerseits durch das Betätigen der Hilfe-Taste auf der haptischen Bedienkonsole sowie andererseits per Spracheingabe (über das integrierte sprachverstehende System INSENSE; siehe Kap. 3.3.2). Im letztgenannten Fall können natürliche Äußerungen, wie etwa „Ich brauche Hilfe“ oder „Gib mir doch mal einen Tipp“, eingesetzt werden. Wie bereits in Kap. 3.5 erörtert, existieren laufende Forschungsarbeiten mit dem Bestreben, die Hilfebedürftigkeit des Benutzers automatisch sowie adaptiv zu erfassen, wodurch die aktive Hilfeanforderung durch den Benutzer im Idealfall entfallen kann.

Nach Aktivierung des audiovisuellen Hilfesystems wird auf dem Display im *Anzeigebereich 2* (siehe Abb. 6.3) die Einzelbildaufnahme einer Geste angezeigt, wobei ein räumlicher Pfeil die Bewegungsrichtung angibt. Durch Sprachausgabe wird sowohl die Ausführung der Bewegung als auch die Auswirkung der Geste auf das System erklärt (siehe Abb. 6.10). Der Informationsinhalt bezieht sich dabei auf den aktuellen Kontext innerhalb der Bedienumgebung GECOM.

Abb. 6.10: Beispiel für audiovisuelle Hilfe: Navigation durch das Menü *Geräteauswahl*.

Die zur Verfügung stehenden „Hilfekonserven“ zur gestischen Bedienung wurden an den jeweiligen Entwicklungsstand (siehe Kap. 6.2) von GECOM angepasst. Ein detaillierter Überblick befindet sich im Anhang A.2.

6.5 Multimodalität

Aufgrund der Nutzung mehrerer Kommunikationskanäle (siehe Abb. 6.11) und insbesondere der Bereitstellung mehrerer Eingabekanäle handelt es sich bei GECOM um ein sogenanntes *multimodales* Bedienkonzept. Die in der Literatur vertretenen Abhandlungen zur Thematik *Multimodalität* bieten diesbezüglich ein breites Spektrum sowohl an Definitionen als auch an verschiedenen Taxonomien an (siehe z.B. [NIG93]). Zur groben Einteilung lassen sich zunächst zwei grundsätzliche Ausprägungsformen multimodaler Systeme unterscheiden: die *sequentielle* und die *parallele* Multimodalität. Während Eingaben bei der sequentiellen Multimodalität zeitlich aufeinanderfolgend

(und somit getrennt) stattfinden, können bei der parallelen Multimodalität (bewusste) zeitliche Überlappungen auftreten. In der extremsten Ausprägung ergibt sich hierbei die vom Benutzer verfolgte Intention nur dann, wenn die verwendeten Modalitäten in zeitlich korrektem Bezug zueinander ausgewertet werden. Als Beispiel sei an dieser Stelle das „Put-That-There"-Konzept nach [BOL80] genannt, welches Positionsänderungen von Objekten durch kombinierte Anwendung von Sprache und Zeigegesten vorsieht.

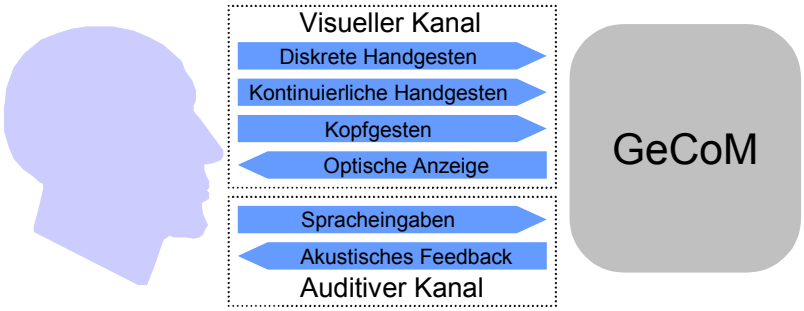


Abb. 6.11: Nutzung mehrerer Kommunikationskanäle bei GECOM.

Bei der Bedienumgebung GECOM wird auf die Auswertung des zeitlichen Bezuges bei überlappendem Einsatz unterschiedlicher Modalitäten bewusst verzichtet, da derartige Bedienabfolgen bei gezielten Untersuchungen nahezu nie beobachten wurden (siehe auch [NEU00]). Dies erlaubt den Schluss, dass die parallele Multimodalität zur Bedienung von Infotainmentsystemen im Fahrzeug bezüglich einer intuitiven Bedienbarkeit keine Vorteile gegenüber der sequentiellen Form bietet.

Die hier angewandte Form der sequentiellen Multimodalität vertritt folgende Grundanschauung: Die gleichzeitige Bereitstellung mehrerer Eingabekanäle soll dem Benutzer die Möglichkeit bieten, seine Eingaben unter Ausnutzung der jeweiligen Vorteile einer Modalität zu tätigen. Dabei besteht jedoch kein Zwang, vorgegebenen Strategien zu folgen, d.h. der Benutzer kann persönlich bevorzugte Eingabemethoden weitgehend beliebig kombinieren.

Die Stärke der Spracheingabe liegt z.B. in der kompakten Übermittlung von komplexen Anweisungen. Auch die gestische Eingabe weist spezifische Vorteile auf. Sie kann einerseits dann eingesetzt werden, wenn eine Spracherkennung aufgrund lauter Störgeräusche nur unzureichend funktioniert. Andererseits gestaltet sich die gestische Eingabe bei vielen kurzen Kommandos, wie z.B. „nächsten Menüpunkt auswählen" oder „Ton stummschalten", als schneller und somit effizienter als die Spracheingabe. Systemrückfragen lassen sich durch einfaches Kopfnicken oder -schütteln sehr intuitiv beantworten. Analoge Einstellungen, wie z.B. die Regelung der Musiklautstärke, können komfortabel mit kontinuierlichen Handbewegungen getätigt werden. Hier liegt wiederum eine Schwäche der Sprache, welche sich im Allgemeinen nur für absolute Angaben eignet.

Das Management der eingehenden Daten verschiedener Modalitäten übernimmt bei GECOM ein Regelwerk (siehe Kap. 3.6). Folgendes Szenario gibt ein Beispiel für einen multimodalen Bedienablauf: Der Benutzer sagt "Ich möchte jetzt Radio hören", woraufhin die Radiofunktion aktiviert wird. Da der aktuelle Sender nicht dem Geschmack des Benutzers entspricht, wechselt dieser mit einer Winkbewegung nach rechts zum nächsten Sender. Das neue Musikprogramm gefällt ihm, weshalb er mit einer kontinuierlichen Handgeste die Lautstärke auf ein angenehmes Niveau anhebt.

6.6 Evaluierung

Nachfolgend werden zwei Usability-Studien vorgestellt, die in verschiedenen Entwicklungsstadien der Bedienumgebung GECOM durchgeführt wurden: 1) *Diskrete Handgestik* und 2) *Kontinuierliche Handgestik und Kopfgestik* (siehe Kap. 6.2)[14].

Zur Vermeidung von störenden Nebeneffekten eines automatischen Gestenerkennungssystems (z.B. Fehlerkennungen) wurde in beiden Untersuchungen die in Kap. 3.4.1 beschriebene Wizard-of-Oz-Methode unter Einsatz der dort aufgeführten Mechanismen zur automatischen Versuchssteuerung und Datenerfassung sowie zur Eingabe kontinuierlicher Handbewegungen verwendet. Die Benutzerstudien wurden im lehrstuhleigenen Fahrsimulator (siehe Abb. 6.12) durchgeführt und waren jeweils in drei Versuchsblöcke unterteilt:

1) Selbstständige Exploration ohne Fahraufgabe
2) Geführter Versuch ohne Fahraufgabe
3) Geführter Versuch mit Fahraufgabe

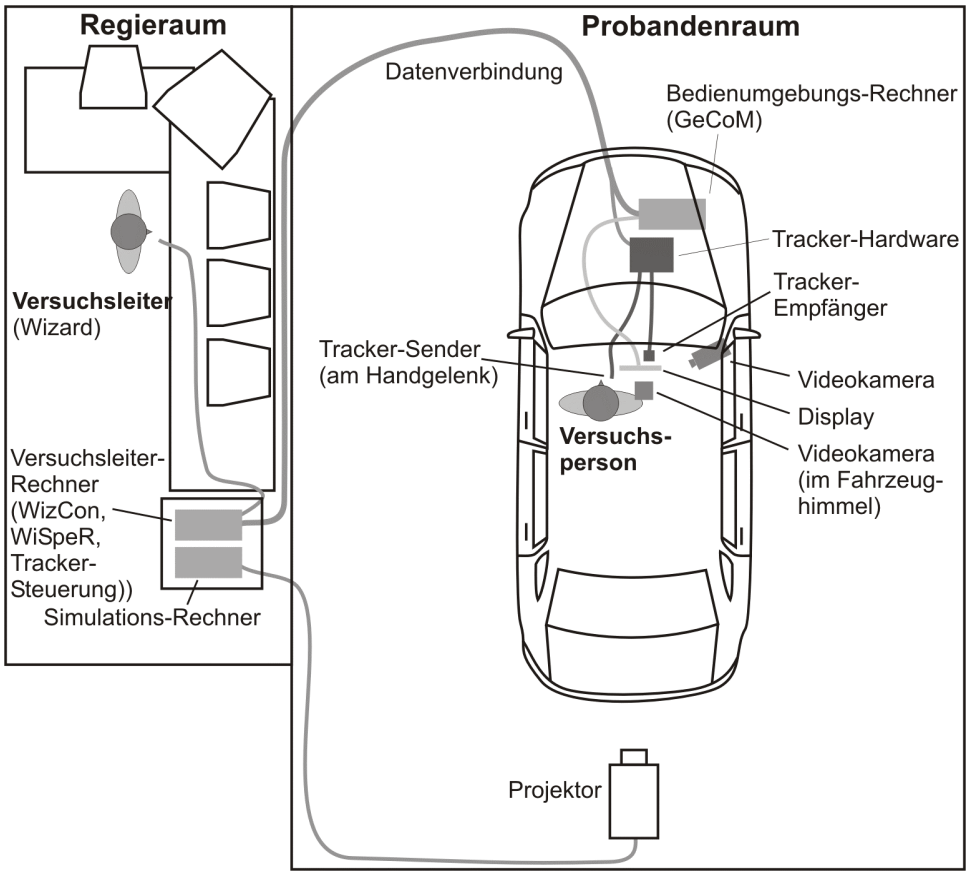


Abb. 6.12: Usability-Labor mit Fahrsimulator; schematische Darstellung der eingesetzten Komponenten. Für weitere Informationen zu den einzelnen Modulen (WIZCON, WISPER etc.) siehe Kap. 3.4.

[14] Die Bedienung unter zusätzlicher Verwendung natürlicher Sprache (Entwicklungsstadium 3) wird derzeit noch in laufenden Forschungsarbeiten von [NIE03] untersucht.

Zur Sicherstellung unverfälschter Aussagen über die Intuitivität waren ausschließlich Probanden ohne Vorkenntnisse zur gestischen Bedienung involviert. Die Versuchspersonen erhielten weder eine Einweisung in die Bedienumgebung GECOM noch erklärende Hinweise zur gestischen Bedienung. Als Informationsquelle wurde lediglich das audiovisuelle Hilfesystem (siehe Kap. 6.4.4) zur Verfügung gestellt.

6.6.1 Usability-Untersuchung 1: *Diskrete Handgestik*

Zur Evaluierung des Einsatzes von GECOM unter Verwendung von diskreten Handgesten diente ein Benutzertest mit 19 Versuchspersonen im Alter zwischen 22 und 60 Jahren. Insgesamt wurden 64 Aufgaben gestellt, die ausschließlich mittels diskreter Handgestik erfüllt werden mussten. Ziel dieser Untersuchung war die Klärung folgender Fragestellungen:

- Wie *intuitiv* erfolgt die Bedienung von GECOM?
- Bestätigt sich das implementierte *Gestenvokabular*?
- Wie hoch ist die *Benutzerakzeptanz*?

Intuitivität

Die zur Erfüllung aller 64 Bedienaufgaben benötigte Anzahl an Bedienschritten variiert zwischen 239 und 323. Diese starke Schwankung resultiert einerseits aus der angesprochenen Redundanz von GECOM, welche das Erreichen eines Bedienziels über verschiedene Bedienabfolgen zulässt. Andererseits wird jedoch auch beobachtet, dass die Anzahl der benötigten Einzelgesten mit zunehmender Unsicherheit des Benutzers ansteigt, da in diesen Fällen auch Eingaben erfolgen, welche nicht zur Erfüllung der gestellten Aufgabe beitragen. Als objektives Kriterium für die Intuitivität der Bedienung wird die Anzahl der Hilfeaufrufe in verschiedenen Kontexten betrachtet (siehe Abb. 6.13).

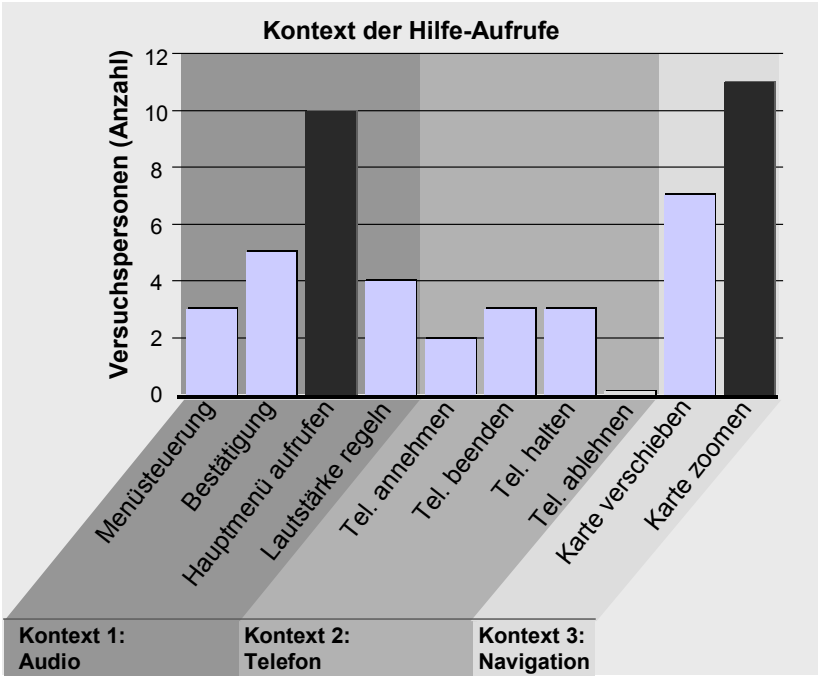


Abb. 6.13: Anzahl der Hilfeaufrufe in verschiedenen Kontexten.

Verglichen mit der Anzahl der gestellten Aufgaben (64 Aufgaben, 19 Versuchsdurchführungen) hält sich die Häufigkeit der Hilfeaufrufe bei nahezu allen Bedienkategorien in Grenzen. Lediglich zwei Funktionsgruppen weisen hier einen deutlich höheren Hilfebedarf auf: *Hauptmenü (Geräteauswahl) aufrufen* und *Karte zoomen*. Die Unsicherheit bei der gestischen Bedienung dieser Funktionen lässt sich wie folgt erklären:

Die Geste zum Aufruf des Hauptmenüs (Ziehen an einem virtuellen Griff; siehe Tab. 4.1) ergab sich im Laufe der Voruntersuchungen im Gegensatz zum übrigen Gestenvokabular nicht als intuitiv (siehe Kap. 6.4.1). Die gewählte grafische Darbietung des plastischen Bügels, an dem die verschiedenen Geräte des Hauptmenüs angeordnet sind (siehe Abb. 6.3, *Anzeigebereich 1*), sollte dem Benutzer die Verwendung dieser „Ziehen"-Geste nahe legen. Offensichtlich wird dieser Hinweis jedoch von den meisten Benutzern nicht wahrgenommen bzw. nicht verstanden, weswegen bei der entsprechenden Aufgabenstellung gezwungenermaßen Hilfe angefordert wird.

Die Unsicherheiten bei der Veränderung der Kartengröße (Karte zoomen) resultieren vermutlich aus der Tatsache, dass die Bedienung von GeCoM bis zu dieser Aufgabenstellung ausschließlich zweidimensional erfolgte. Der Benutzer glaubt offensichtlich, alle anwendbaren Methoden der gestischen Interaktion bereits zu kennen, und rechnet nicht mehr mit der Möglichkeit einer dreidimensionalen Bedienung. Zwar wird diese bei der Navigationslandkarte durch ein entsprechendes Gestaltungsmerkmal (räumlicher Pfeil mit Lupe; siehe Abb. 6.8) visualisiert, dieses wird jedoch in vielen Fällen übersehen oder falsch interpretiert.

Insgesamt erfolgt die Bedienung weitgehend unter Verwendung des implementierten Gestenvokabulars, wodurch sich dieses als intuitiv bestätigt. Zwei der 19 Versuchspersonen waren in der Lage, den gesamten Funktionsumfang ohne Einweisung und Hilfestellungen gestisch zu bedienen. Für die restlichen Teilnehmer erwies sich das audiovisuelle Hilfesystem als nützliche Informationsquelle: Die hier dargebotene Hilfe ermöglichte ihnen die vollständige Bedienung gemäß den gestellten Aufgaben.

Akzeptanz

Das eingesetzte Audiofeedback wurde positiv aufgenommen, d.h., früher beobachtete häufige Blickabwendungen traten - speziell im Versuchsteil mit simulierter Autofahrt - kaum mehr auf. Insbesondere der Einsatz von Sprachausgabe zur Orientierung in Menülisten wurde von vielen Versuchspersonen lobend erwähnt. Das in Abb. 6.14 dargestellte Meinungsbild erlaubt Rückschlüsse auf die Akzeptanz der gestischen Bedienung seitens der Versuchspersonen. Hier sind die gesammelten Äußerungen der abschließenden Befragung vollständig aufgelistet und in vier Meinungsgruppen unterteilt. Die Versuchsteilnehmer sollten in möglichst spontaner Weise qualitativ ausdrücken, wie sie die berührungslose Bedienung empfanden.

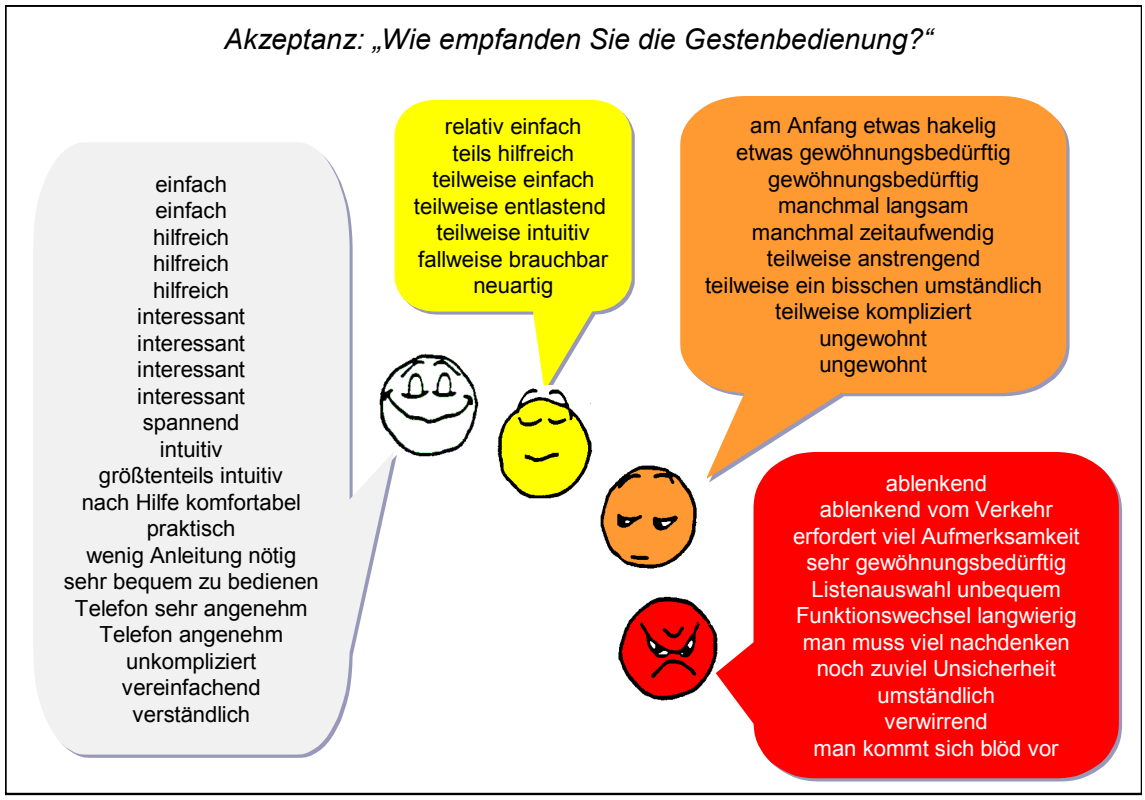


Abb. 6.14: Benutzeräußerungen zur Bewertung der Gestenbedienung.

Insgesamt ergeben sich zwar überwiegend positive Benutzeräußerungen, jedoch zeigt sich ein offensichtlicher Schwachpunkt bei der gestischen Navigation durch lange Listen (z.B. Telefonbuch), welche als „langwierig“ und „umständlich“ bewertet wird. Zudem wird die Gestenbedienung teilweise als „ablenkend vom Verkehr“ bezeichnet. Dabei ist jedoch zu bedenken, dass den Versuchspersonen in dieser Studie keine Alternative zur gestischen Eingabe bereitgestellt wurde. Es fehlt also die Vergleichsmöglichkeit z.B. zur haptischen Bedienung, welche in vorangehenden Untersuchungen (siehe Kap. 5) als deutlich ablenkender empfunden wurde als die gestische und auch objektiv mit signifikant stärkeren Ablenkungseffekten einhergeht.

Detaillierte Aussagen über Informationsbedarf und Lernverhalten bei der Gestenbedienung, welche sich aufgrund der beschriebenen Studie treffen lassen, werden in [NIE02B] beschrieben.

6.6.2 Usability-Untersuchung 2: *Kontinuierliche Handgestik und Kopfgestik*

Nach der Einführung der neuen Gestentypen (siehe auch *Entwicklungsstadium 2*; Kap. 6.2), wurde deren Gebrauchstauglichkeit anhand einer Benutzerstudie mit 21 Versuchspersonen im Alter von 22 bis 58 Jahren evaluiert. Dabei wurden insgesamt 80 Bedienaufgaben gestellt (siehe Anh. A.3), zu deren Erfüllung zusätzlich zur diskreten auch kontinuierliche Handgestik sowie Kopfgestik eingesetzt werden durften [FRA02].

Um den Einsatz kontinuierlicher Gesten zu forcieren, wurden an manchen Stellen sogenannte *Aufgaben-Loops* integriert: Der Benutzer wird beispielsweise aufgefordert, einen Eintrag aus dem Telefonbuch zu wählen, welcher sich in hoher Distanz zum aktuellen Eintrag befindet - etwa „Wählen

Sie den Eintrag ***G**orrst* aus“ bei aktuellem Eintrag ***A**DAC*[15]. Das Aufgabenziel lässt sich sowohl durch die Verwendung vieler (14) diskreter Einzelgesten als auch durch eine einzige kontinuierliche Geste erreichen. Falls die Versuchsperson die erste (umständlichere) Variante anwendet, beginnt der Aufgaben-Loop: Nach dem Erreichen des Eintrages ***G**orrst* wird der Benutzer angewiesen, wieder den Eintrag ***A**DAC* auszuwählen. Diese Schleife wird solange durchlaufen, bis der Benutzer entweder die kontinuierliche Gestik anwendet, oder diese Bedienmöglichkeit nach einem Hilfeaufruf vom audiovisuellen Hilfesystem erfährt. Derartige Aufgaben-Loops werden auch bei der Lautstärkeeinstellung und zur Kartenbedienung eingesetzt.

Folgende Fragestellungen standen bei dieser Untersuchung im Vordergrund:

- Wie *intuitiv* wird die kontinuierliche Gestik angewandt?
- Wie *intuitiv* wird die Kopfgestik angewandt?
- Wie hoch ist die *Akzeptanz* der neuen Gestentypen?
- Inwiefern profitiert die *Gebrauchstauglichkeit* von den neuen Gestentypen?

Intuitivität

Während die Anwendung diskreter Handgesten auch in dieser Untersuchung weitgehend intuitiv erfolgt, wird die Möglichkeit kontinuierlicher Eingaben von den meisten Versuchspersonen nicht selbstständig erkannt - lediglich zwei Probanden bilden hier die Ausnahme. Die übrigen Teilnehmer durchlaufen den ersten Aufgaben-Loop (Lautstärkeeinstellung, Aufgabe 13; siehe Anh. A.3) einige Male unter Verwendung diskreter Gesten, bis sie schließlich Hilfe anfordern. Daraufhin präsentiert das audiovisuelle Hilfesystem die Möglichkeit der kontinuierlichen Bedienung. Im folgenden Versuchsablauf ist nun zu beobachten, dass nahezu alle Versuchspersonen diese Gestenart spontan auch in anderen Kontexten anwenden - etwa zur Navigation durch lange Listen (Telefonbuch) und zur Kartenbedienung. Die gewählte visuelle Darbietung für kontinuierliche Gesten (siehe z.B. Abb. 6.7) ist demnach nicht genügend selbsterklärend, um diesen Gestentyp nahezulegen. Die Darstellung der plastischen Kugel erweist sich als zu abstrakt, da sie nur entfernt an einen klassischen Schieberegler erinnert.

Ähnliche Befunde ergeben sich für die Kopfgestik. Diese wird lediglich von einer einzigen Versuchsperson intuitiv zur Beantwortung von Systemrückfragen verwendet, während die übrigen Teilnehmer explizit auf diese Bedienmöglichkeit hingewiesen werden müssen. Aus der Befragung der Versuchspersonen geht hervor, dass das entsprechende Visualisierungsmerkmal (Kopfgestensymbol, siehe Abb. 6.9) meist schlicht übersehen wurde. Hier ist somit eine auffälligere Darstellung angebracht, um diese äußerst ungewohnte Methode der Mensch-Maschine-Interaktion zu veranschaulichen.

Ungefähr zwei Drittel der Probanden versuchen, kontinuierliche Gestik auch dort einzusetzen, wo dieser Gestentyp nicht zur Verfügung steht, wie etwa in der *Geräteauswahl*. Hier wurde lediglich

[15] Die Einträge sind in alphabetischer Reihenfolge im Telefonbuch abgelegt: An erster Stelle befindet sich der Eintrag *ADAC* (Teilnehmer *Gorrst* ist der fünfzehnte Eintrag).

die Verwendung diskreter Gesten vorgesehen, da es sich um eine sehr kurze Liste handelt, welche ohne Weiteres durch Einzelgesten bedient werden kann. Die Wahrung der Konsistenz erweist sich also auch in dieser Hinsicht als wichtiges Gestaltungskriterium: Wird neben der diskreten auch die kontinuierliche Bedienung vorgesehen, so muss diese auch in *jedem* Systemzustand anwendbar sein, selbst wenn dies aus Expertensicht in bestimmten Fällen nicht notwendig erscheint.

Akzeptanz

Obwohl die neuen Gestentypen, wie oben beschrieben, weniger intuitiv eingesetzt werden als die diskrete Handgestik, finden diese ähnlich hohe Akzeptanz (siehe Abb. 6.15). Hinsichtlich der etwas schlechteren Bewertung der kontinuierlichen Gestik muss an dieser Stelle angemerkt werden, dass dieser Gestentyp störenden Einflussgrößen unterlag:

Die elektromagnetischen Signale der Tracker-Hardware wurden durch Reflexionen an der Fahrzeugkarosserie beeinflusst, sodass der zulässige Aktionsradius auf etwa 0,5 m um den Empfänger (siehe Abb. 3.4) eingeschränkt werden musste. Die ausbleibende Systemreaktion beim Verlassen dieses Aktionsbereichs war für die Versuchspersonen verständlicherweise nur schwer nachvollziehbar.

Darüber hinaus erfolgte das dynamische Systemverhalten bei der kontinuierlichen Gestik teilweise nicht erwartungskonform. In Anlehnung an das typischerweise bei der Computer-Mouse eingesetzte *ballistische Tracking* wurden die Systemreaktionen dynamisch an die Geschwindigkeit der Handbewegungen angepasst. Die dabei gewählten Parameter erwiesen sich jedoch laut Aussage vieler Versuchspersonen nicht als optimal.

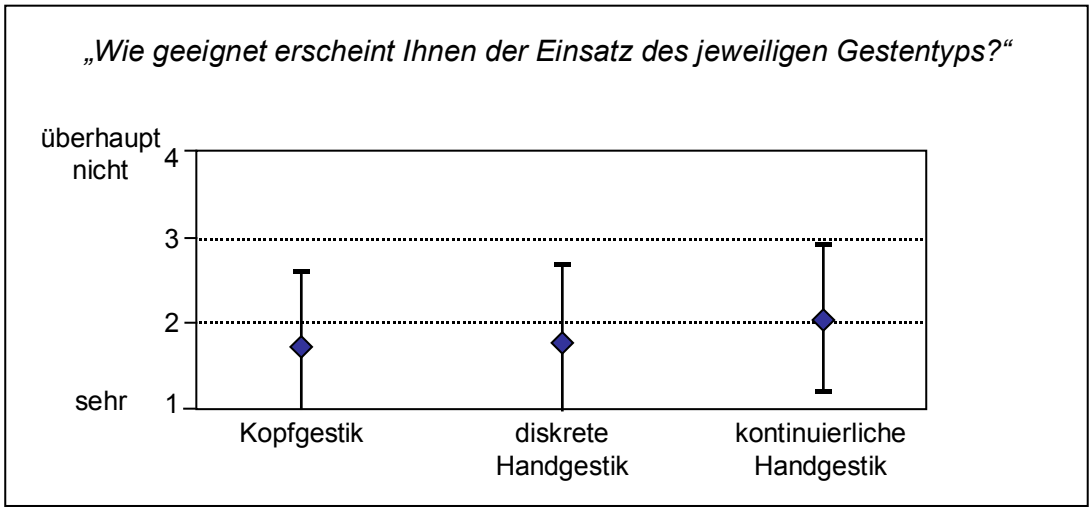


Abb. 6.15: Subjektive Beurteilung der implementierten Gestentypen (Mittelwerte und Standardabweichungen).

Ein objektives Indiz für die Akzeptanz der neuen Gestentypen ist der prozentuale Anteil ihrer Verwendung, da der jeweilige Gestentyp frei gewählt werden konnte.

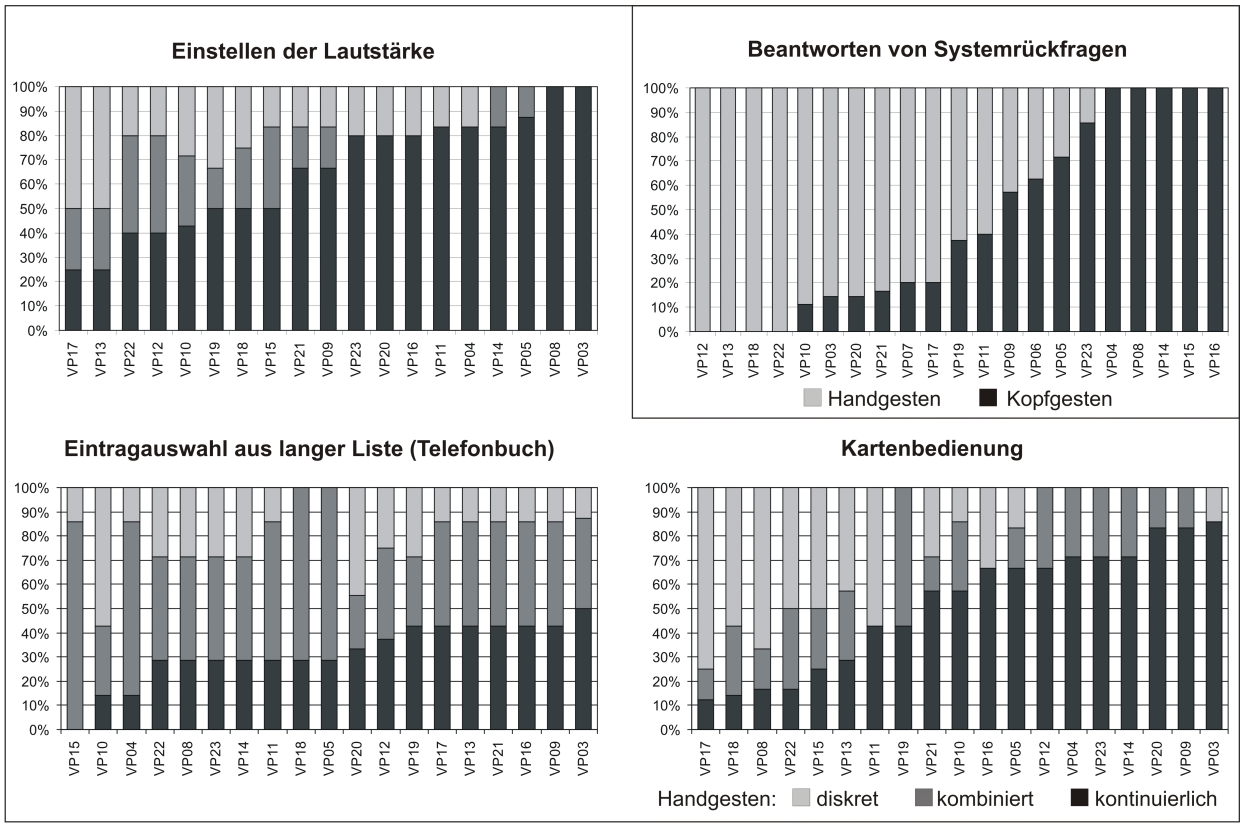


Abb. 6.16: Prozentuale Anteile der verwendeten Gestenarten nach Aufgabentypen und Versuchspersonen (jeweils akkumulierte Darstellungen). Die Bezeichnungen *kontinuierlich*, *diskret* und *kombiniert* beziehen sich jeweils auf die Bearbeitung *einer* Bedienaufgabe.

Wie aus Abb. 6.16 hervorgeht, wird die kontinuierliche Gestik sehr häufig sowohl ausschließlich als auch in Kombination mit diskreter Gestik eingesetzt. Der Anteil rein diskret bearbeiteter Aufgaben ist hingegen vergleichsweise gering. Besonders deutlich zeigt sich dies bei der Lautstärkeeinstellung, welche von zwei Teilnehmern sogar durchgängig kontinuierlich bedient wurde. Die offensichtlich hohe Akzeptanz der kontinuierlichen Lautstärkeeinstellung deckt sich zudem mit entsprechend positiven Aussagen der Versuchspersonen. Es muss hier zusätzlich beachtet werden, dass die Möglichkeit der kontinuierlichen Bedienung dem Hauptteil der Versuchspersonen nicht von Beginn an bewusst war (siehe oben: *Intuitivität*).

Der Einsatz von Kopfgestik zur Beantwortung von Systemrückfragen hält sich mit dem von Handgestik nahezu die Waage. Verglichen mit der Gesamtanzahl der gestellten Aufgaben ist die Möglichkeit, Kopfgesten einzusetzen, jedoch sehr selten gegeben, sodass hier keine gesicherte objektive Aussage bezüglich der Akzeptanz getroffen werden kann.

Gebrauchstauglichkeit

Abschließend wird nun kurz diskutiert, inwiefern die Gebrauchstauglichkeit (Usability) von GeCoM durch die Implementierung der neuen Gestentypen, kontinuierliche Handgestik und Kopfgestik, profitiert.

Die Kopfgestik zur Beantwortung von Systemrückfragen wird von den Versuchsteilnehmern stark befürwortet und auch gerne angewandt. Aufgrund der Natürlichkeit dieser Bewegungsmuster ist mit

hoher Sicherheit davon auszugehen, dass die einhergehende kognitive Belastung äußerst gering ist. Wird die zu beantwortende Frage wie im vorliegenden Fall via Sprachausgabe vorgelesen, so kann der gesamte Interaktionsprozess ohne Blickabwendungen vom Straßenverkehr erfolgen. Dies bestätigen auch die Beobachtungen im Versuchsteil während der simulierten Fahrt. Aufgrund dieser positiven Befunde wird die automatische Kopfgestenerkennung in Form einer realen und fahrzeugtauglichen Systemkomponente umgesetzt (siehe Kap. 7).

Die kontinuierliche Handgestik ist der diskreten in zwei Punkten überlegen: Sie ermöglicht einerseits das stufenlose Einstellen von Parametern wie etwa der Musiklautstärke. Andererseits können große Änderungsintervalle durch eine einzige kontinuierliche Geste anstelle vieler diskreter Einzelgesten erzielt werden. Somit kann der Benutzer das gewünschte Bedienziel in bestimmten Fällen (z.B. Navigation durch eine lange Liste) schneller erreichen als durch rein diskrete Bedienung. Dies verdeutlicht Abb. 6.17, die einen Zeitvergleich für die alternative Verwendung der Gestentypen kontinuierlich und diskret zum Erreichen desselben Aufgabenziels darstellt. Bei den benötigten Zeiten für rein diskrete Bedienung handelt es sich um Schätzwerte. Sie ergeben sich aus den gemessenen probandenspezifischen Durchschnittswerten für horizontale Winkgesten multipliziert mit der erforderlichen Anzahl an Einzelschritten, die zur Lösung der hier betrachteten Aufgabe (Auswahl eines bestimmten Eintrags im Telefonbuch) erforderlich sind.

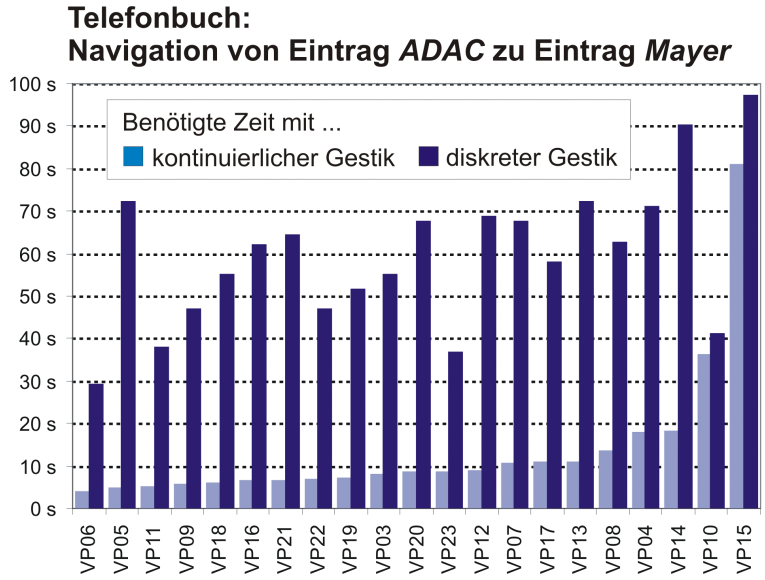


Abb. 6.17: Zeitvergleich für die Auswahl eines bestimmten Telefonbucheintrages mit kontinuierlichen bzw. diskreten Gesten (siehe Aufgabe 32 im Anh. A.3).

Der hier offensichtliche Zeitvorteil relativiert sich jedoch bei Betrachtung der damit verbundenen Ablenkungseffekte: So ist eine kontinuierliche Bedienung ohne permanenten Blickkontakt zum Display nahezu unmöglich und daher während der Fahrt kaum in Betracht zu ziehen. Die höhere Effizienz gegenüber der diskreten Gestik kann somit nur im stehenden Fahrzeug genutzt werden. Die einzige Ausnahme stellt hier die Lautstärkeregelung dar, die mit einem direkten akustischen Feedback einhergeht und daher - im Einklang mit den Versuchsbefunden - „blind“ bedienbar ist. Bei der im nachfolgenden Kapitel beschriebenen Realisierung eines Gestenerkennungssystems wird die kontinuierliche Gestik aus diesem Grunde ausschließlich für das stufenlose Regeln der Musiklautstärke implementiert.

7

Fahrzeugtaugliche Gestenerkennung

7.1 Spezielle Anforderungen

Für den praktischen Einsatz von Gestenerkennung im Fahrzeug ergeben sich domänenspezifische Randbedingungen, denen ein solches System unbedingt genügen muss.

Eine entscheidende Rolle spielt dabei der Wirtschaftlichkeitsaspekt, welcher den Einsatz von kostengünstiger Hardware fordert. Dies hat zur Folge, dass für einen Gestenerkenner nur geringe Ressourcen an Rechenleistung und Speicherkapazität bereit gestellt werden können. Darüber hinaus sind die aus Sicht des Fahrzeugherstellers akzeptablen Kosten für zusätzliche Komponenten, die zur Erfassung von Gesten benötigt werden, sehr eng bemessen.

Als weitere Randbedingung ist das sogenannte „Package“ zu berücksichtigen: Das durch die Vielzahl der bereits vorhandenen elektronischen Komponenten stark begrenzte Raumangebot im Fahrzeug stellt auch an die Hardware eines Gestenerkenners die Anforderung, nur minimalen Platz beanspruchen zu dürfen.

Bereits auf dem Gebiet der Spracherkennung hat sich die Fahrzeugumgebung aufgrund der vorhandenen Störgeräusche als schwer beherrschbares Terrain erwiesen. Vergleichbare Schwierigkeiten ergeben sich für die Gestenerkennung in Form von extremen Beleuchtungsverhältnissen. Dies bezieht sich sowohl auf die hohe Intensität - z.B. durch direkte Sonneneinstrahlung - als auch auf die starke Variabilität des Lichteinfalls. Dieser ändert sich permanent mit der Fahrtrichtung und es ist darüber hinaus mit abrupten Hell-/Dunkelübergängen (z.B. durch Einfahrt in einen Tunnel) zu rechnen. Um eine robuste Gestenerkennung dennoch zu gewährleisten, muss das System weitgehend fremdlichtunempfindlich sein.

Schließlich sei die Systemreaktionszeit als weiterer wichtiger Aspekt genannt. Während sich diese in anderen Domänen in erster Linie auf die Benutzerakzeptanz auswirkt, stellt sie im Fahrzeug zusätzlich eine wichtige Voraussetzung für eine verkehrssichere Bedienung dar. In Untersuchungen (siehe [GEI02A]) konnte gezeigt werden, dass das Ausbleiben einer Systemreaktion bereits nach

wenigen 100 ms zu Kontrollblicken des Fahrers und somit zur Ablenkung vom Verkehrsgeschehen führt. Die Verarbeitung von gestischen Eingaben muss somit möglichst echtzeitnah (d.h. idealerweise < 100 ms) erfolgen, was in Anbetracht der bereits erwähnten Einschränkungen hinsichtlich der zur Verfügung stehenden Rechenleistung eine besonders schwer handhabbare Zielvorgabe darstellt.

7.2 Motivation zum Einsatz von Infrarot-Sensoren

7.2.1 Videobasierte Gestenerkennung

Auf dem heutigen Stand der Forschung wird an die automatische Gestenerkennung vorwiegend mit Methoden der computergestützten Bildverarbeitung herangegangen. Dabei werden die Merkmale[16], die man für die Mustererkennung heranzieht, aus den Bilddaten von Kameras gewonnen. Bevor eine Merkmalberechnung stattfinden kann, ist es zunächst erforderlich, das beobachtete Körperteil - z.B. eine gestikulierende Hand - vom Bildhintergrund zu trennen (örtliche Segmentierung). Zur maschinenverständlichen Beschreibung der segmentierten Handform können z.B. flächenbeschreibende Momente (siehe auch [MOR00], S.61 ff.) berechnet werden. Die Merkmalgewinnung bei videobasierten Verfahren stellt im Wesentlichen eine Datenreduktion dar: Aus einer großen Menge von Pixeldaten soll eine handhabbare Anzahl an charakteristischen Informationen gewonnen werden. Sowohl die Vorverarbeitung der Eingangsdaten und die Berechnung der Merkmale als auch die eigentliche Mustererkennung gestalten sich im Allgemeinen sehr rechenaufwändig. Veränderliche Bildhintergründe und wechselnde Lichtverhältnisse, wie sie im Fahrzeug auftreten, bereiten bislang große Probleme, zu deren Lösung zusätzlicher Rechen- und Hardwareaufwand betrieben werden muss. So kann z.B. der störende Einfluss durch Umgebungslicht mit starker Beleuchtung der Bildszene verringert werden. Hierfür werden üblicherweise spezielle Leuchtmittel im nicht-sichtbaren Bereich eingesetzt (siehe z.B. [ZOB03A]). Insgesamt muss zur Gewährleistung einer zufriedenstellenden Robustheit ein Aufwand betrieben werden, der die im Fahrzeug vorhandenen bzw. die unter oben genannten Umständen realisierbaren Mittel weit übersteigt. Mit dem Einsatz der videobasierten Gestenerkennung im Fahrzeug ist somit in naher Zukunft kaum zu rechnen.

7.2.2 Grundidee der sensorbasierten Gestenerkennung

Hauptziel des hier verfolgten Ansatzes zur Gestenerkennung ist es, eine robuste visuelle Interaktion unter Einhaltung der beschriebenen Anforderungen in der Fahrzeugdomäne (siehe Kap. 7.1) zu ermöglichen. An Stelle einer Kamera werden kostengünstige Infrarot-Distanz-Mess-Sensoren (im Folgenden: IR-Sensoren) für die Merkmalgewinnung verwendet. Die Kombination mehrerer Sensoren zu einem zweidimensionalen Array ermöglicht die Erfassung von 3D-Information über ein beobachtetes Szenario. Dabei werden von den Sensoren die Abstände zum ausführenden Körperteil (Hand bzw. Kopf) gemessen. Sowohl die Anzahl *n* der eingesetzten Sensoren als auch deren Platzierung und Ausrichtung bestimmen dabei den Informationsgehalt. Zieht man den Vergleich zur

[16] Hierbei handelt es sich um eine charakteristische Datensequenz, welche die Information des zu erkennenden Musters beinhaltet.

bildbasierten Informationsaufnahme, so liefert ein derartiges Sensor-Array ein aus *n* diskreten Pixeln bestehendes „Bild“, wobei jedes Pixel quasikontinuierliche Abstandsinformation trägt.

Die örtliche Segmentierung gestaltet sich im Gegensatz zu videobasierten Verfahren unkompliziert: Der Szenenhintergrund besitzt einen deutlich größeren Abstand zum Sensor-Array als z.B. eine im erfassten Bereich befindliche Hand und kann daher einfach ausgeblendet werden. Des Weiteren sind für die Merkmalgewinnung keine aufwändigen Berechnungen zur Datenreduktion nötig, da die Gesten-Information ohnehin bereits in kompakter Form vorliegt. Jedes Abtastintervall liefert einen Messwert pro Sensor. Diese Distanzmesswerte können fast direkt als Merkmale verwendet werden, die sich mit konventionellen Mustererkennungsmethoden bei verhältnismäßig geringem Rechenaufwand verarbeiten lassen. Die verwendeten Sensortypen sind unempfindlich gegen die im Fahrzeug vorhandenen Fremdlichteinflüsse. Darüber hinaus werden die Messwerte nur sehr geringfügig von der Farbe des erfassten Objekts beeinflusst. Durch ihre geringen Abmessungen lassen sich die Sensoren problemlos in den Fahrzeuginnenraum integrieren. Insgesamt erfüllt das hier vorgestellte Gestenerkennungssystem die geforderten Kriterien für den realen Einsatz im Fahrzeug. Im Rahmen der bayerischen Hochschul-Patentinitiative BAYERNPATENT wurde das hier beschriebene Verfahren am 16.09.2002 von der Technischen Universität München zum Patent angemeldet [GEI02B].

7.3 IR-Sensor-Arrays zur Erfassung von 3D-Information

7.3.1 Funktionsweise der IR-Sensoren

Zur Erfassung von Objektbewegungen (Kopf oder Hand) werden IR-Sensoren der Typen SHARP GP2D12 und GP2D120 (im Folgenden: GP2D12 = *DS1*; GP2D120 = *DS2*) eingesetzt. Die Distanzmessung erfolgt nach dem *Triangulationsprinzip* (siehe Abb. 7.1).

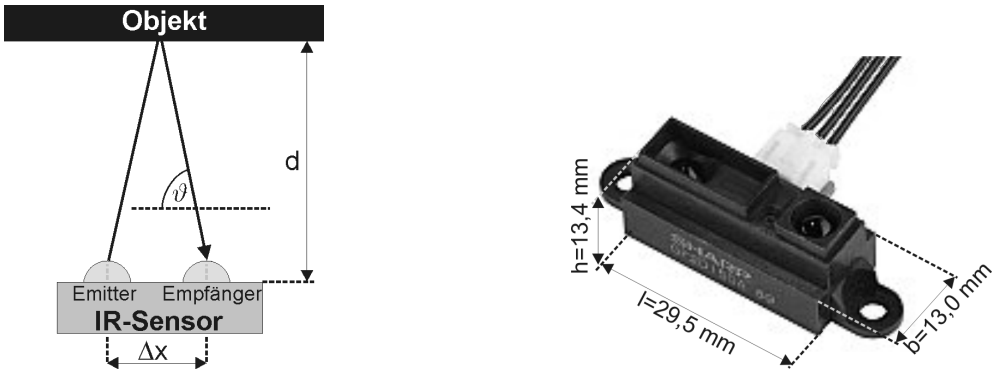


Abb. 7.1: Triangulationsprinzip (links), IR-Sensor SHARP GP2D12 (rechts).

Der Emitter sendet einen IR-Lichtimpuls aus, welcher bei Anwesenheit eines Objekts im Erfassungsbereich von diesem reflektiert wird und unter dem Winkel ϑ auf den Empfänger trifft. Dieser Winkel wird erfasst, indem das einfallende Licht mittels einer Präzisionsoptik auf ein integriertes CCD-Array gelenkt wird. Die Distanz d zum Objekt ist mit dem Winkel ϑ und dem Abstand Δx zwischen Emitter und Empfänger (etwa 20 mm) über den geometrischen Zusammenhang nach Gl. 7.1 festgelegt.

$$d = \frac{\Delta x \cdot \tan \vartheta}{2} \tag{7.1}$$

Der gemessene Abstand wird durch ein analoges Spannungssignal ausgegeben. Da die Sensoren über keine interne Linearisierung verfügen, verhält sich die Ausgangsspannung U_{out} aufgrund der vorliegenden geometrischen Zusammenhänge nicht linear zur Distanz *d;* sie ist jedoch nahezu unabhängig von der Farbe des reflektierenden Objekts (siehe Abb. 7.2).

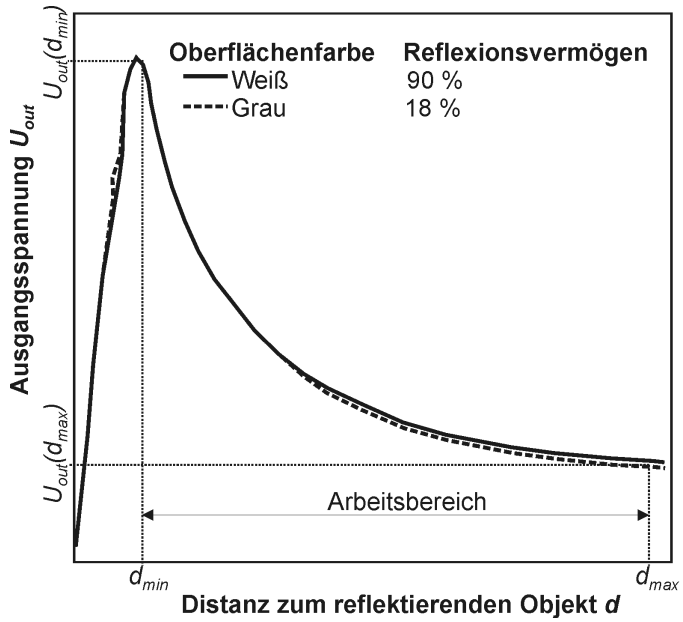


Abb. 7.2: Kennlinien des Sensortyps *DS2* für weißes und graues Papier.

Das Kennlinienmaximum legt die minimal erfassbare Distanz d_{min} fest; rechtsseitig davon befindet sich der vorgesehene Arbeitsbereich. Die qualitativen Kennlinienverläufe der IR-Sensoren *DS1* und *DS2* sind nahezu identisch. Sie unterscheiden sich lediglich in ihren Distanzmessbereichen und den zugehörigen Ausgangsspannungen (siehe Tab. 7.1). Somit handelt es sich bei *DS2* um einen Nahbereichsensor, während sich *DS1* zu Messung größerer Distanzen eignet (Fernbereichsensor).

	d_{min} [mm]	d_{max} [mm]	$U_{out}(d_{min})$ [V]	$U_{out}(d_{max})$ [V]
DS1 (GP2D12)	100	800	2,6	0,4
DS2 (GP2D120)	40	300	3,1	0,4

Tab. 7.1: Parameter der Sensortypen *DS1* und *DS2* (Versorgungsspannung U_{cc} = 5,0 V).

Bei der Verwendung der IR-Sensoren zur Gestenerkennung muss sichergestellt sein, dass die jeweilige Distanz zwischen bewegtem Körperteil (Kopf bzw. Hand) und Sensor im Arbeitsbereich liegt. Insbesondere bei der Unterschreitung der minimalen Distanz ist keine eindeutige Zuordnung der Ausgangsspannung zur Objektdistanz möglich, da $U_{out}(d)$ nur abschnittweise umkehrbar ist.

Für die Kopfgestenerkennung, bei der sich das Sensorfeld in unmittelbarer Kopfnähe befindet, wird der Sensortyp *DS2* eingesetzt. Die Bedingung $d \geq d_{min}$ wird aufgrund der räumlichen Anordnung der IR-Sensoren eingehalten (siehe *Sensor-Array für die Kopfgestenerkennung*, Kap. 7.3.2). Die

Handgestenerkennung erfolgt mit einer Kombination beider Sensortypen, wodurch ein Erfassungsbereich zwischen 40 mm und 800 mm abgedeckt werden kann.

Die Wandlung des vom Sensor erfassten Lichteinfallwinkels ϑ in die Ausgangsspannung U_{out} geschieht durch einen integrierten Mikroprozessor. Wie Abb. 7.3 zeigt, beansprucht jede Distanzmessung laut Hersteller eine Zeit von 38,3 ms ± 9,6 ms. Erfahrungsgemäß erfolgt die Aktualisierung der Messwerte im Mittel nach 40 ms. Die maximale Abtastrate der IR-Sensoren wird daher auf $f_{abt,max}$=25 Hz festgelegt (siehe auch 7.4.1).

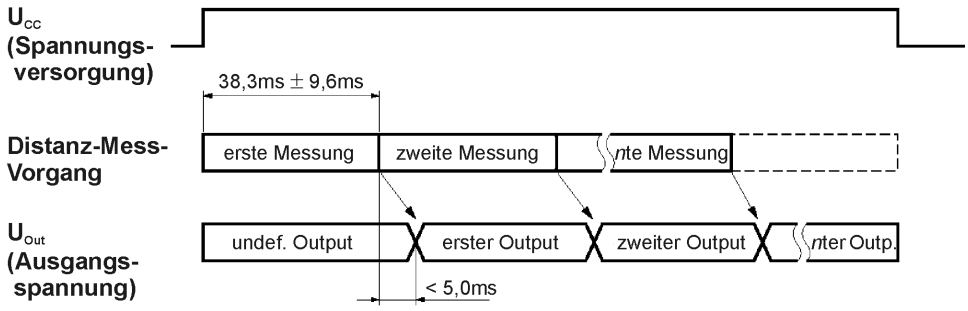


Abb. 7.3: Zeitdiagramm der Sensortypen *DS1 und DS2* (siehe auch Anh. A.4.1).

7.3.2 Anordnung der IR-Sensoren zu einem Array

Bei der Platzierung der Sensoren muss sichergestellt werden, dass die Ausgangssignale für die zu unterscheidenden Gesten charakteristische Musterverläufe liefern.

Sensor Array für die Handgestenerkennung

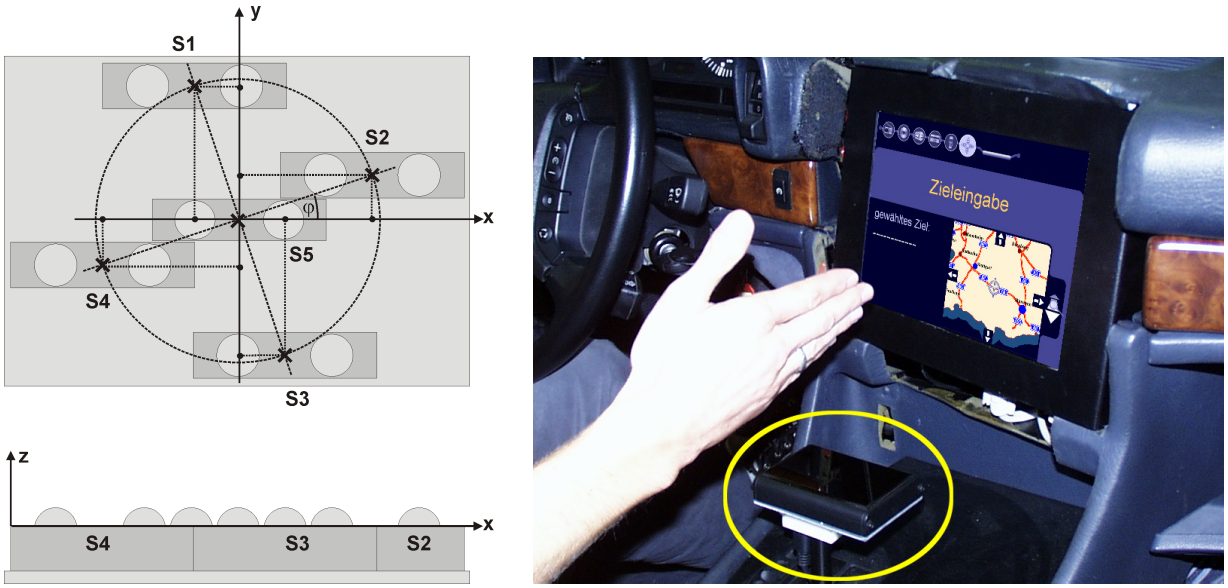


Abb. 7.4: Sensoranordnung für Handgestenerkennung (links), die in den Betätigungsgriff der Gangschaltung integriert wurde (rechts).

Für die Handgestenerkennung werden fünf IR-Sensoren auf einer ebenen Fläche (x/y-Ebene) platziert (siehe Abb. 7.4), so dass die Distanzmessung in z-Richtung erfolgt. Die hier gewählte Anord-

nung optimiert für die Sensoren *S1, ..., S4* die Anforderung, dass Handbewegungen in der horizontalen Ebene längs der x-Achse bzw. y-Achse unter Verwendung einer möglichst geringen Anzahl an Sensoren örtlich optimal aufgelöst werden. Die Mittelpunkte der Sensoren werden als Referenzpunkte der Abstandsmessung betrachtet und liegen auf einem quadratischen Raster. Somit weisen sowohl die senkrechten Projektionen der Sensormittelpunkte auf die x-Achse als auch auf die y-Achse äquidistanten Abstand zueinander auf. Dazu werden die Sensoren *S1, ..., S4* äquidistant auf einer Kreislinie verteilt, wobei diese orthogonale Anordnung um den Winkel $\varphi = atan(1/3) \approx 18{,}4°$ gegen die Koordinatenachsen gedreht ist.

Für geradlinige Handbewegungen längs der x- bzw. y-Achse liefert diese Sensoranordnung deutlich unterscheidbare Distanzverläufe (siehe Abb. 7.5), wobei keiner der Sensoren redundante Information liefert. Dies erlaubt den Schluss, dass die Anordnung auch zur Klassifikation komplexerer Gesten gut geeignet ist.

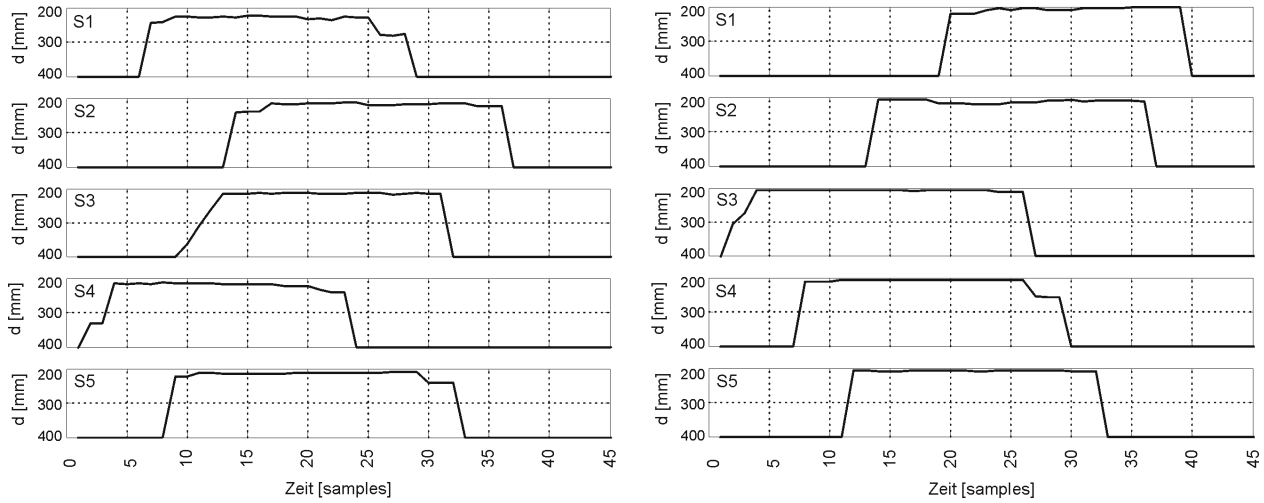


Abb. 7.5: Distanzverläufe[17] für eine horizontale Bewegung mit flacher Hand über das Sensorfeld in positiver x-Richtung (links) und in positiver y-Richtung (rechts).

Zusätzlich wird der Sensor *S5* im Kreiszentrum platziert, um sicher zu stellen, dass auch vertikale Handbewegungen, welche meist mittig über der Sensoranordnung ausgeführt werden, optimal erfasst werden.

Um den intuitiven Ausführungsraum (siehe [ZOB02]) von Handgestik im Fahrzeug zu erfassen, bietet es sich an, das Sensor-Array im Bereich der Mittelkonsole anzubringen. In der vorliegenden Arbeit wurden die IR-Sensoren in den Schaltknauf integriert (siehe Abb. 7.4). Es kann davon ausgegangen werden, dass die Handbewegung zur Betätigung der Gangschaltung hochgradig überlernt ist, woraus folgt, dass die Position des Schaltknaufs - und somit die der Gestensensorik - intuitiv bekannt ist. Der systembedingten Vorgabe, dass gestische Eingaben im Erfassungsbereich der Sensorik ausgeführt werden müssen, kann daher ohne nennenswerte kognitive Belastung genügt werden.

[17] Objekte, die vom Handsensor-Array weiter als d_{HG} = 400 mm entfernt sind, werden als Hintergrund betrachtet (siehe Kap. 7.4.1).

Sensor-Array für die Kopfgestenerkennung

Zur Erfassung von Kopfbewegungen liegt es nahe, die IR-Sensoren an der Kopfstütze des Fahrersitzes anzubringen (siehe Abb. 7.6). Daraus ergibt sich relativ zum Sensorfeld eine gut definierte Kopfposition, welche bei Personen mit unterschiedlicher Körpergröße im Wesentlichen in vertikaler Richtung variiert. Durch eine höhenverstellbare Kopfstütze kann das Sensorfeld also individuell in eine Position gebracht werden, die optimale Distanzmessungen zum Kopf erlaubt[18].

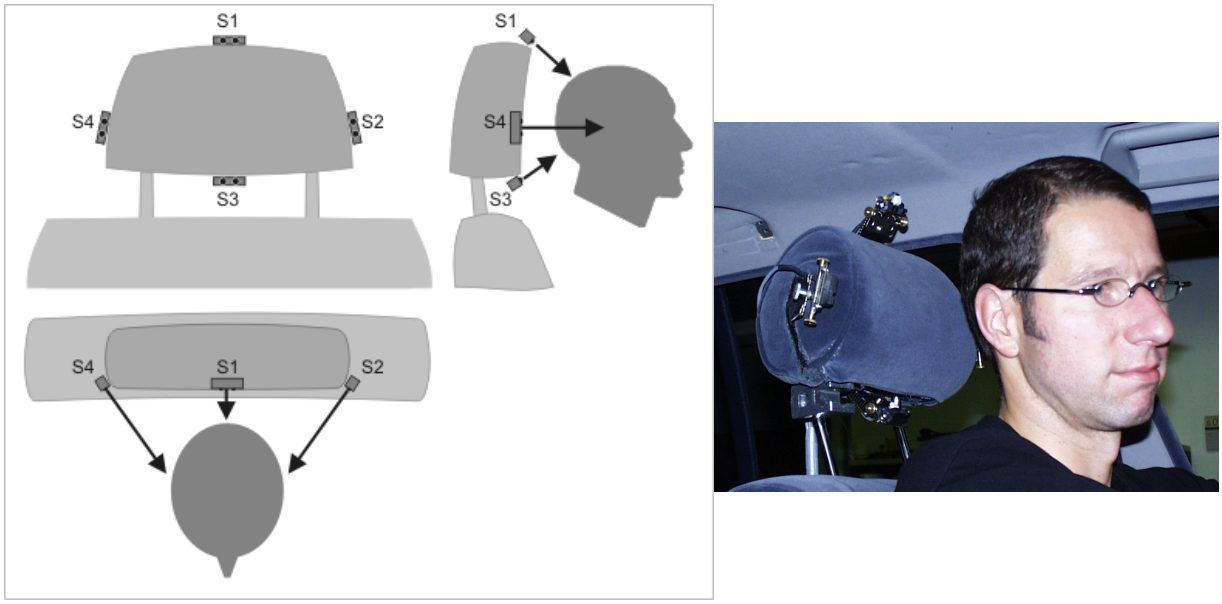


Abb. 7.6: Sensoranordnung für Kopfgestenerkennung; Schematische Darstellung (links) und realer Aufbau mit Benutzer (rechts).

Das Sensorfeld für die Kopfgestenerkennung besteht aus vier IR-Sensoren. Die hier gewählte Anordnung ist optimiert auf die Unterscheidung von Kopfnicken und Kopfschütteln: Bei einer „Ja"-Geste ergeben sich vor allem für das Sensorpaar *S1, S3* ausgeprägte Distanzschwankungen. Entsprechendes gilt für das Sensorpaar *S2, S4* bezüglich einer „Nein"-Geste. Das Sensorfeld liefert also für die beiden Kopfgesten deutlich unterscheidbare bzw. charakteristische Signalverläufe, welche sich durch Mustererkennungsverfahren klassifizieren lassen (siehe auch Abb. 7.15).

7.4 Verfahren

Abb. 7.7: Grobstruktur des Gestenerkenners.

Der prinzipielle Aufbau des hier vorgestellten Gestenerkenners (siehe Abb. 7.7) entspricht dem eines klassischen Mustererkennungssystems: Zunächst wird Information über die zu klassifizierende Geste aufgenommen - hier durch *Distanzmessung* - und einer *Vorverarbeitung* unterzogen. Bei der an-

[18] Die Sensoren sind dabei so justiert, dass die Höheneinstellung der Kopfstütze für optimale Distanzmessung mit der ergonomisch korrekten Position übereinstimmt.

schließenden *Segmentierung* werden Beginn und Ende eines Bewegungsablaufs detektiert. Die innerhalb dieses Zeitintervalls erfassten Daten werden durch *Merkmalgewinnung* aufbereitet und einer *Klassifikation* zugeführt. Hierbei wird die beste Zuordnung des ermittelten Musterverlaufs zu einer im Trainingsmaterial enthaltenen Geste bestimmt.

Im Folgenden wird detailliert auf die einzelnen Komponenten des Gestenerkennungssystems (siehe Abb. 7.8) und deren Zusammenspiel eingegangen. Zur besseren Übersicht, werden die beschriebenen Teilkomponenten jeweils zu Beginn jedes nachfolgenden Unterkapitels zusätzlich isoliert dargestellt.

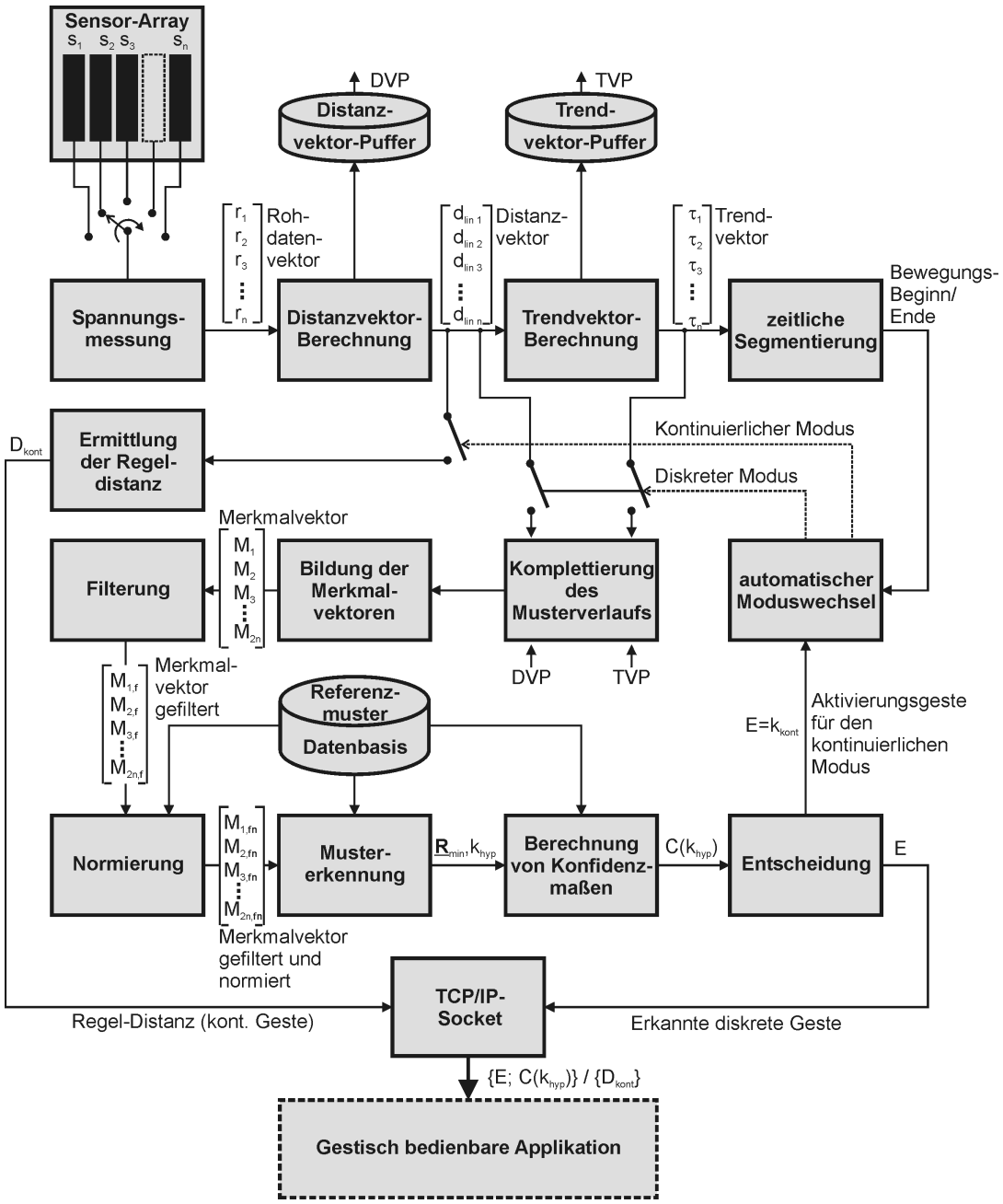


Abb. 7.8: Blockschaltbild des Gesamtsystems.

7.4.1 Spannungsmessung und Distanzberechnung

Die Ausgangsspannungen der IR-Sensoren eines Arrays werden von einer PC-Messkarte (12 bit A/D Wandler, siehe Datenblatt Anh. A.4.2) eingelesen. Die Abtastung erfolgt im Multiplexverfahren, wobei jede Kanalumschaltung nur etwa 4 µs beansprucht. Somit kann die Erfassung der aktuellen Distanzwerte *aller* Sensoren eines Arrays - im Vergleich zur verwendeten Abtastfrequenz - als quasi gleichzeitig betrachtet werden.

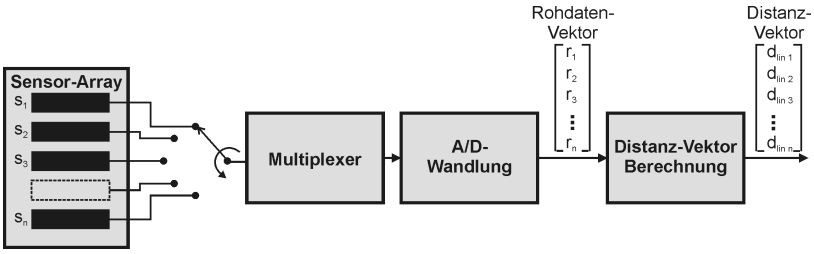


Abb. 7.9: Blockdiagramm Distanzmessung.

Bei der Handgestenerkennung werden die Distanzmesswerte des Sensorarrays mit der maximal möglichen Abtastfrequenz $f_{abt,Hand} = f_{abt,max} = 25$ Hz erfasst. Da Kopfbewegungen im Allgemeinen langsamer ausgeführt werden als Handgesten, hat sich hier eine Abtastfrequenz von $f_{abt,Kopf} = 10$ Hz als sinnvoll erwiesen.

Die Ausgangssignale der Sensoren werden mit einer Auflösung von 12bit A/D-gewandelt. Jedes Abtastintervall liefert einen *n*-dimensionalen Rohdatenvektor $\underline{r}$ (siehe Abb. 7.9), dessen Komponenten den aktuellen Ausgangsspannungen der *n* Sensoren entsprechen.

Aufgrund der nichtlinearen Sensorkennlinie (siehe Abb. 7.2) verhalten sich diese Rohdaten nicht proportional zur tatsächlichen Distanz. Speziell bei der kontinuierlichen Gestik, bei der eine Handbewegung direkt auf eine Regelgröße einwirkt, ist jedoch ein linearer Zusammenhang die Voraussetzung für einen gut funktionierenden Regelkreis zwischen Mensch und System. Daher werden bei der anschließenden Distanz-Vektor-Berechnung die tatsächlichen Distanzen in [mm] ermittelt und im Distanzvektor $\underline{d}_{lin}$ abgelegt.

Um die zur Linearisierung benötigten Parameter zu bestimmen, wurden zunächst die (gemittelten) Kennlinien der beiden Sensortypen *DS1* und *DS2* durch Messreihen ermittelt. Wie Abb. 7.10 zeigt, lassen sich diese in den jeweiligen Arbeitsbereichen (siehe Tab. 7.1) sehr gut durch die reziproke Funktion $U_{app}(d)$ unter Verwendung der in Tab. 7.2 angegebenen Parameter a_1 und a_0 approximieren.

$$\begin{aligned} &U_{out}(d) \cong U_{app}(d) \text{ für } d_{\min} \le d \le d_{\max} \\ &U_{app}(d) = a_1 \cdot \frac{1}{d} + a_0 \end{aligned} \tag{7.2}$$

	a_1 [Vmm]	a_0 [V]
DS1 (GP2D12)	226,0	0,19
DS2 (GP2D120)	123,0	0,0

Tab. 7.2: Parameter der Approximationsfunktion $U_{app}(d)$.

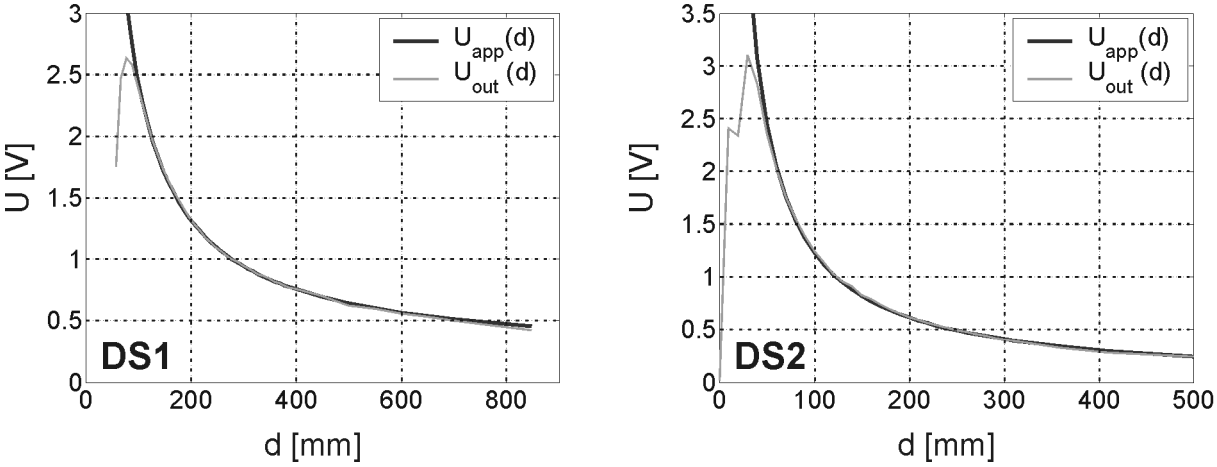


Abb. 7.10: Kennlinien $U_{out}(d)$ und Approximationsfunktionen $U_{app}(d)$ von *DS1* (links) und *DS2* (rechts).

Aus Gl. (7.2) ergibt sich für den linearen Distanzwert d_{lin}:

$$d_{lin}(U_{out}) = \frac{a_1}{U_{out} - a_0} \cong d(U_{out}) \tag{7.3}$$

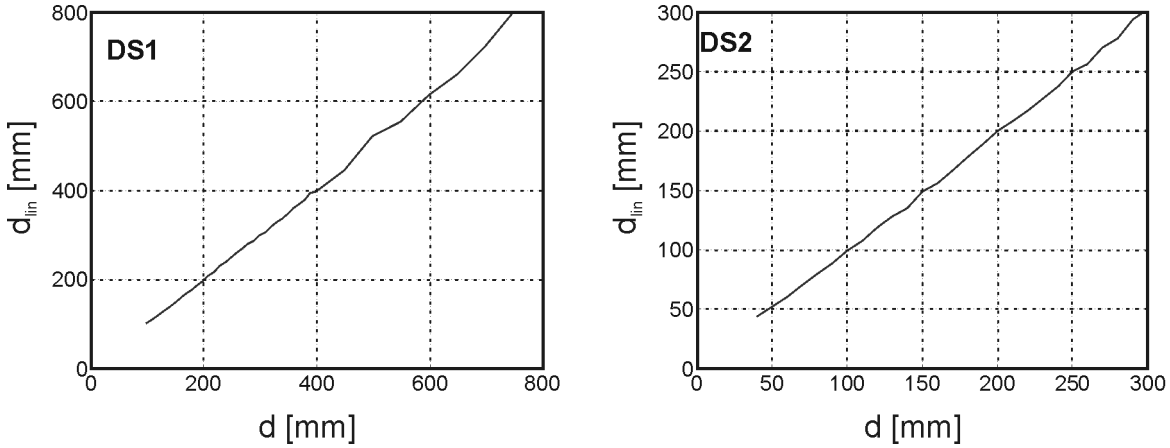


Abb. 7.11: Zusammenhang zwischen realer Distanz d und approximierter Distanz d_{lin} für *DS1* (links) und *DS2* (rechts).

Wie Abb. 7.11 zeigt, besteht nun ein annähernd linearer Zusammenhang zwischen der realen Distanz d und der vom System ermittelten Distanz d_{lin}. Während dies für *DS2* (Kopfgestenerkennung) im gesamten Arbeitsbereich gilt, steigt d_{lin} bei *DS1* (Handgestenerkennung) für Distanzen $d > 500$ mm etwas überproportional an. Dies spielt jedoch hier keine Rolle, da für die Gestenerken-

nung nicht der gesamte Arbeitsbereich ausgeschöpft wird: Übersteigt die ermittelte Objektdistanz d_{lin} einen bestimmten Schwellwert d_{HG}, so wird die Messung dem Szenenhintergrund zugeordnet. Die Hintergrundausblendung geschieht also, indem für jede Komponente des Distanzvektors $\underline{d}_{lin}$ folgende Regel angewandt wird:

$$\begin{aligned} d_{lin\ i} = \min\{d_{lin\ i}, d_{HG}\} \text{ mit } d_{HG} &= 400 \text{ mm bei Handgestenerkennung } (DS1, DS2) \\ d_{HG} &= 250 \text{ mm bei Kopfgestenerkennung } (DS2) \end{aligned} \tag{7.4}$$

Programmintern wird jeder Distanzwert $d_{lin\ i} = d_{HG}$ als Hintergrund interpretiert; d.h. im Erfassungsbereich des entsprechenden Sensors befindet sich kein Objekt.

Die so erhaltenen Distanzvektoren $\underline{d}_{lin}[t_j]$ werden später (siehe Kap. 7.4.3) in die Merkmalvektoren $\underline{M}[t_j]$ aufgenommen, welche für die Mustererkennung herangezogen werden.

7.4.2 Trendvektor-Berechnung und zeitliche Segmentierung

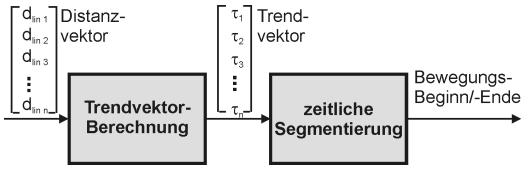


Abb. 7.12: Blockdiagramm Trendvektor-Berechnung und zeitliche Segmentierung**.**

Ermittlung der Trendvektoren

Hauptziel der Trendvektorberechnung ist die Gewinnung von Merkmalen, welche im Gegensatz zu den Distanzvektoren unabhängig sind von der absoluten Position des bewegten Körperteils im Sensorfeld. Jede der *n* Komponenten eines Trendvektors enthält Information über die Geschwindigkeit des erfassten Objekts relativ zum zugehörigen Distanzsensor. Die Berechnung der Trendvektoren könnte durch einfache komponentenweise Differenzbildung von aufeinanderfolgenden Distanzvektoren geschehen. Da die Messwerte jedoch mit einem gewissen Grundrauschen behaftet sind, wird zur Bestimmung der Relativgeschwindigkeit eine größere Anzahl von Distanzvektoren herangezogen, um eine höhere Robustheit zu erreichen. Der Begriff *Trend* ist hierbei zu verstehen als eine aus der Vergangenheit abgeleitete Grundrichtung der Entwicklung der gemessenen Distanzen.

Dazu wird der aktuelle Distanzvektor nach jedem Abtastzyklus in einem Puffer von bestimmter Länge w_{tr} - bewährt hat sich eine Zeitfensterlänge von $w_{tr} = 10$ Vektoren - gespeichert, wobei nach dem *FIFO*[19]-Prinzip mit jedem Neuzugang der älteste Vektor aus dem Puffer gelöscht wird (siehe Abb. 7.13).

[19] Abkürzung für englisch: First In First Out.

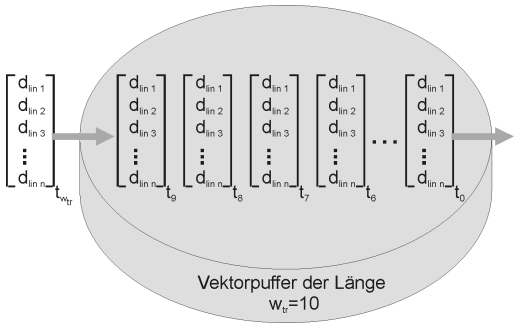


Abb. 7.13: FIFO-Puffer zur Speicherung von w_{tr} Distanzvektoren.

Nach jeder Aktualisierung des Vektorpuffers werden nun für die im Puffer befindlichen Distanzvektoren komponentenweise die Regressionsgeraden $d_{reg\ i}[j]$ berechnet, wobei j hier die (zeitliche) Position des jeweiligen Distanzvektors im Vektorpuffer darstellt:

$$d_{reg\ i}[j] = m_i \cdot j + b_i; \qquad j = \{1, 2, ..., w_{tr}\} \tag{7.5}$$

Abb. 7.14 zeigt exemplarisch die Regressionsgerade für eine Komponente $d_{lin\ i}$ für den Fall, dass sich ein Objekt vom Sensor i entfernt.

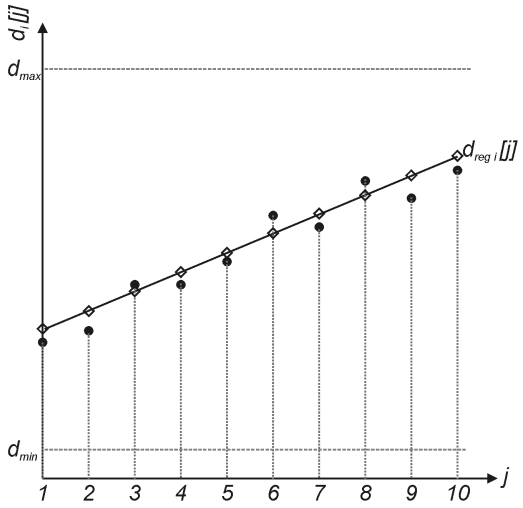


Abb. 7.14: Regressionsgerade $d_{reg\ i}[j]$ durch w_{tr}=10 Distanzmesswerte einer Distanzvektor-Komponente $d_{lin\ i}$.

Als Schätzverfahren wird hierfür die *Methode der kleinsten Quadrate* (MKQ) ausgewählt. Diese lineare Schätzfunktion ist eine Gerade, die so in die Datenpunkte $d_{lin\ i}[j]$ eingepasst ist, dass sie diese "gleichförmig" trennt, d.h. sie minimiert die Summe der vertikalen Abweichungsquadrate. Dabei wird davon ausgegangen, dass die zufälligen Messfehler voneinander unabhängig sind und dass die Genauigkeit der Distanzmessung distanzunabhängig ist (*Homoskedastizität*).

Allgemein berechnet man für k Wertepaare ($x[i]$, $y[i]$) die Steigung m und den Achsenabschnitt b der Regressionsgerade $y = m \cdot x + b$ wie folgt:

$$m = \frac{\sum_{j=1}^{k} x_j y_j - \frac{\sum_{j=1}^{k} x_j \cdot \sum_{j=1}^{k} y_j}{k}}{\sum_{j=1}^{k} x_j^2 - \frac{(\sum_{j=1}^{k} x_j)^2}{k}} \tag{7.6}$$

$$b = \frac{\sum_{j=1}^{k} y_j - m \sum_{j=1}^{k} x_j}{k} \tag{7.7}$$

Als repräsentative Größe für die Relativgeschwindigkeit zwischen Objekt und Sensor ist dabei im Folgenden nur die Steigung m von Interesse. Da die Abtastung der Sensoren mit der festgelegten Frequenz f_{abt} erfolgt (d.h. die Messungen erfolgen in äquidistanten Zeitabständen) gilt hier:

$$x_j = j \tag{7.8}$$

Dadurch lässt sich Gl. (7.6) mit

$$m = m_i, \quad y_j = d_{lin\,i}[j], \quad k = w_{tr} \tag{7.9}$$

wie folgt vereinfachen:

$$m_i = 6 \cdot \frac{2 \cdot \sum_{j=1}^{w_{tr}} j d_{lin\,i}[j] - (w_{tr}+1) \cdot \sum_{j=1}^{w_{tr}} d_{lin\,i}[j]}{w_{tr}^3 - w_{tr}}; \quad i = \{1, 2, ..., n\} \tag{7.10}$$

Nach jedem Abtastzyklus werden auf diese Weise die Steigungen m_i der Regressionsgeraden berechnet und bilden die Komponenten des Trendvektors $\underline{\tau}$. Sie repräsentieren ein Maß für die aktuell vorliegenden Relativgeschwindigkeiten unter Mitberücksichtigung einer durch w_{tr} festgelegten Anzahl von Distanzmessungen aus der Vergangenheit. Die angewandte Regressionsanalyse wirkt sich dabei wie ein Tiefpassfilter aus. Je größer die Länge w_{tr} des Vektorpuffers gewählt wird, desto stärker wird auch der Geschwindigkeitsverlauf geglättet. Es wird also einerseits das störende Signalrauschen unterdrückt, andererseits können aber auch kurze, schnelle Richtungswechsel des erfassten Objekts nicht mehr aufgelöst werden. Als sinnvoller Kompromiss wird die Größe des FIFO-Puffers sowohl bei der Kopf- als auch bei der Handgestenerkennung[20] auf $w_{tr} = 10$ gesetzt.

[20] Da w_{tr} nicht nur an die Art der Bewegungsabläufe, sondern auch an die jeweils verwendete Abtastfrequenz f_{abt} ($f_{abt,Hand} \neq f_{abt,Kopf}$!) angepasst werden muss, ist die Tatsache, dass sich die Einstellung $w_{tr} = 10$ bei *beiden* Erkennungssystemen als optimal erwies, als zufällig anzusehen.

Solange sich ein Objekt im Sensorfeld relativ zum Sensor *i* nicht bewegt, schwankt m_i um den Wert Null. Entfernt sich das Objekt vom Sensor, so ergibt sich eine Regressionsgerade mit positiver Steigung m_i (siehe auch Abb. 7.14), welche sich näherungsweise proportional zur Relativgeschwindigkeit $v_{Objekt/Sensor}$ verhält. Dementsprechend erhält man für m_i negative Werte, wenn sich das Objekt dem Sensor annähert.

Die Trendvektoren $\underline{\tau}[t_j]$ enthalten somit positionsunabhängige Merkmale und werden später (siehe Kap. 7.4.3) mit den Distanzvektoren $\underline{d}_{lin}[t_j]$ zu den Merkmalsvektoren $\underline{M}[t_j]$ vereint.

Abb. 7.15 zeigt exemplarische Trendvektorverläufe für Kopfgesten unter Verwendung der in Kap. 7.3.2 beschriebenen Sensoranordnung (siehe Abb. 7.6).

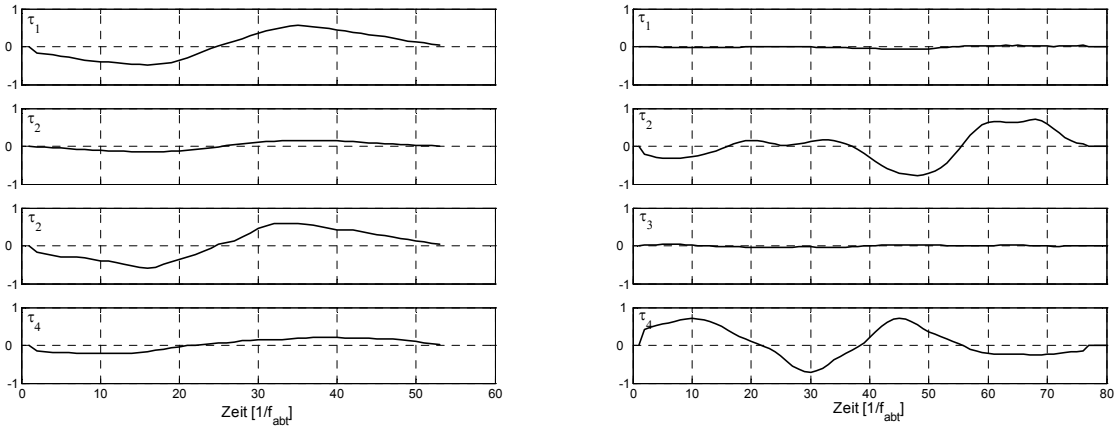


Abb. 7.15: Trendvektorverläufe $\underline{\tau}[t_j]$ für Kopfnicken (links) und Kopfschütteln (rechts); die Komponenten τ_i wurden gemäß Kap. 7.4.4 normiert.

Zeitliche Segmentierung

Um eine Geste zu klassifizieren, muss das System zunächst in der Lage sein, Beginn und Ende einer Bewegungssequenz zu erkennen, damit die zwischen diesen Zeitpunkten erfassten Distanz- und Trendvektoren aufgezeichnet werden können. Zur Realisierung dieser *zeitlichen Segmentierung* wird folgende Randbedingung für die Gestenausführung festgelegt:

Das bewegte Körperteil muss sich vor Beginn der Geste und nach deren Ende für eine jeweils festgelegte Zeit in Ruhe befinden (oder den Erkennungsbereich verlassen).

Diese Voraussetzung ermöglicht die Segmentierung durch *Bewegungsdetektion*, d.h. bei einer Geste handelt es sich um einen abgeschlossenen Bewegungsablauf. Als Maß für eine vorliegende Bewegung wird der Betrag *a* (*activity*) des Trendvektors $\underline{\tau}$ herangezogen:

$$a_j = \left|\underline{\tau}[t_j]\right| = \sqrt{\sum_{i=1}^{n} \tau_i^2[t_j]} \tag{7.11}$$

Abb. 7.16 zeigt den zeitlichen Verlauf des Bewegungsmaßes *a*, welcher sich aus dem (unnormierten) Geschwindigkeitsverlauf der Geste *Kopfschütteln* (siehe Abb. 7.15 rechts) ergibt. Obwohl sich der Kopf vor und nach der Gestenausführung in Ruhe befindet, ergeben sich für das Bewegungs-

maß *a* Werte, die leicht von Null abweichen. Dies ist in erster Linie auf Messrauschen zurückzuführen und ließe sich durch ein entsprechend langes Zeitfenster w_{tr} bei der Trendberechnung unterdrücken. Wie bereits erwähnt (siehe *Ermittlung der Trendvektoren* in diesem Kap.), würde sich dann jedoch eine zu geringe zeitliche Auflösung für die eigentliche Kopfbewegung ergeben.

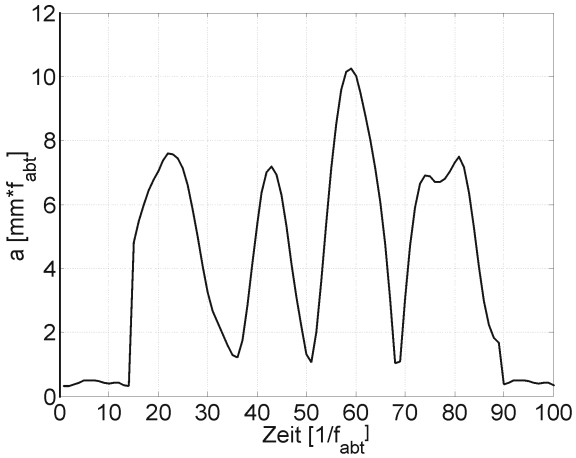


Abb. 7.16: Geste als abgeschlossene Bewegungssequenz; Bewegungsmaß *a* für die Geste *Kopfschütteln*[21] (vgl. Abb. 7.15 rechts).

Für die automatische zeitliche Segmentierung werden folgende Regeln angewandt:
Der Beginn einer Geste wird zum Zeitpunkt t_{start} detektiert, wenn das Bewegungsmaß a_j für mindestens th_{start} direkt aufeinanderfolgende Trendvektoren den Bewegungsschwellwert a_{min} überschreitet.

$$a_j >^{!} a_{\min} \quad \text{für} \quad j = t_{start} - th_{start} - 1, t_{start} - th_{start}, ..., t_{start} \tag{7.12}$$

Entsprechend endet die Geste zum Zeitpunkt t_{stop}, wenn a_{min} für mindestens th_{stop} aufeinanderfolgende Trendvektoren unterschritten wird:

$$a_j <^{!} a_{\min} \quad \text{für} \quad j = t_{stop} - th_{stop} - 1, t_{start} - th_{stop}, ..., t_{stop} \tag{7.13}$$

Für eine sinnvolle Wahl der Parameter a_{min}, th_{start} und th_{stop} werden abhängig vom Anwendungsbereich folgende Überlegungen angestellt:

- Parameterwahl für Kopfgestik

Der Bewegungsschwellwert a_{min} muss deutlich über dem Bewegungsmaß $a_j[t_j]$ liegen, welches sich bei ruhendem (bzw. sich langsam bewegendem) Kopf ergibt. Er bestimmt somit maßgeblich das Ansprechverhalten des Systems.
Zudem wird die Empfindlichkeit beeinflusst durch den Schwellwert th_{start}. Durch die Wahl eines entsprechend hohen Wertes für th_{start} lässt sich ein ungewolltes Ansprechen vermeiden, so dass z.B. kurze abrupte Kopfbewegungen, welche aufgrund von Beschleunigungskräften im fahrenden Auto

[21] Bei der verwendeten Abtastfrequenz $f_{abt,Kopf}$ = 10 Hz kann der hier auftretende Maximalwert von a ≈ 10mm/f_{abt}^{-1} interpretiert werden als ein Geschwindigkeitsbetrag von 0,1m/s.

auftreten, ignoriert werden.
Bei der Wahl des Pausenschwellwertes th_{stop} muss sichergestellt sein, dass die Kopfgeste wirklich als beendet betrachtet werden kann, sich a_j also nicht in einem lokalen Minimum der Bewegungssequenz befindet. Die gewählte Parametereinstellung zeigt Tab. 7.3.

a_{min}	th_{start}	th_{stop}
1 mm/f_{abt}^{-1}	5	8

Tab. 7.3: Parameter der Bewegungsdetektion bei Kopfgestenerkennung.

- Parameterwahl für Handgestik

Bei der Handgestenerkennung befindet sich das bewegte Körperteil - die Hand - im Allgemeinen nur zur Ausführung einer Geste im Erkennungsbereich des Sensorfeldes. Ansonsten ergibt sich (durch die Hintergrundausblendung, siehe Gl. 7.4) für das Bewegungsmaß $a_j[t_j]$ jederzeit der Wert Null. Es ist daher sinnvoll, auch den Bewegungsschwellwert a_{min} auf Null zu setzen, sowie die Startschwelle th_{start} auf den Wert Eins. Dies hat zur Folge, dass ein Gestenbeginn detektiert wird, sobald die Hand von einem der Sensoren erfasst wird - man kann also hier von einer *Objektdetektion* sprechen.
Für die Wahl des Pausenschwellwertes th_{stop} gilt einerseits dieselbe Überlegung wie bei der Kopfgestenerkennung (kein lokales Bewegungsminimum). Andererseits kann es bei Gesten mit großer horizontaler Bewegungsamplitude vorkommen, dass die Hand den Sensorbereich während der Ausführung mehrmals kurzzeitig komplett verlässt und wieder eintritt. Durch einen entsprechend hohen Pausenschwellwert th_{stop} wird gewährleistet, dass dies nicht zu falschen Ende-Detektionen führt. Die gewählte Parametereinstellung zeigt Tab. 7.4.

a_{min}	th_{start}	th_{stop}
0 mm/f_{abt}^{-1}	1	15

Tab. 7.4: Parameter der Bewegungsdetektion bei Handgestenerkennung.

- Plausibilität der Gestendauer

Neben den Kriterien für Gestenbeginn und -ende werden zusätzliche Randbedingungen eingeführt, die sich auf die Dauer der Geste beziehen:

Die Ausführungsdauer unterliegt gewissen Schwankungen, da sich einerseits verschiedene Gestentypen in der Komplexität ihrer Ausführung unterscheiden und andererseits identische Gesten mit unterschiedlichen Geschwindigkeiten (intra- und interindividuell) ausgeführt werden. Dennoch lässt sich die Gesamtdauer Δt_{ges} einer sinnvollen Geste durch Erfahrungswerte eingrenzen auf ein Intervall:

$$\Delta t_{min} \leq^{!} \Delta t_{ges} \leq^{!} \Delta t_{max} \quad \text{mit} \quad \Delta t_{ges} = t_{stop} - t_{start} \tag{7.14}$$

Falls die Ausführungszeit Δt_{ges} nach der Segmentierung unter der Grenze Δt_{min} liegt, kann entweder die aufgezeichnete Bewegungssequenz verworfen werden oder die Aufzeichnung wird fortgesetzt, obwohl das Endekriterium bereits erfüllt wurde. Entsprechend erfolgt während einer stattfindenden

Bewegung ein erzwungener Abbruch bei Erreichen der Obergrenze Δt_{max}, wobei auch hier zur Wahl steht, ob die erfasste Sequenz verworfen oder ausgewertet wird.

Erfahrungsgemäß liegt die Gesamtdauer einer Kopfgeste in folgenden Fällen außerhalb des plausiblen Zeitintervalls: Während Unterschreitungen häufig durch unbeabsichtigtes Ansprechen der Bewegungsdetektion bei kleinen Kopfbewegungen auftreten, deuten Überschreitungen darauf hin, dass sich der Kopf nicht im Zentrum des Sensorfeldes befindet, wodurch starke Messwertschwankungen hervorgerufen werden. In beiden Fällen wäre eine Auswertung der erfassten Bewegungssequenz nicht sinnvoll, so dass diese verworfen wird.

Bei der Handgestenerkennung treten einerseits dann zu geringe Gestendauern auf, wenn die Bewegung nicht über dem Sensorfeld ausgeführt wird, sondern dieses nur kurz „streift". Andererseits ist dies der Fall, wenn die Ausführungsgeschwindigkeit so hoch ist, dass aufgrund der verwendeten Abtastfrequenz $f_{abt,Hand}$ = 25 Hz nur Fragmente der Geste erfasst werden.
Zeitüberschreitungen sind häufig auf unbeabsichtigtes Hantieren im Erfassungsbereich zurück zu führen. Somit ist es auch bei der Handgestik sinnvoll, Bewegungssequenzen mit nicht plausibler Gestendauer zu verwerfen.

Tab. 7.5 zeigt die bei der Plausibilitätsprüfung der Gesamtgestendauer gewählten Parameter für Hand- und Kopfgestik sowie die sich ergebenden Zeitintervalle unter Berücksichtigung der jeweils verwendeten Abtastfrequenz.

	Δt_{min}	Δt_{max}	$\Delta t_{min}/f_{abt}$	$\Delta t_{max}/f_{abt}$
Kopfgestik	5	40	0,5 s	4 s
Handgestik	10	100	0,4 s	4 s

Tab. 7.5: Parameter der Plausibilitätsprüfung.

7.4.3 Komplettierung des Musterverlaufs und Bildung der Merkmalvektoren

Komplettierung

Mit dem Erkennen eines Gestenbeginns werden die nachfolgenden Distanz- und Trendvektoren bis zum Gestenende aufgezeichnet; sie bilden die Bewegungssequenz.
Da jedoch zunächst Bewegung vorhanden sein muss ($a_j >^{!} a_{min}$, siehe Gl. 7.12), damit die Bewegungsdetektion anspricht, liegt der Zeitpunkt t_{start} des detektierten Gestenbeginns für $th_{start}>0$ (Kopfgestenerkennung) zeitlich hinter dem des tatsächlichen Startzeitpunkts $t_{start,real}$. Ein gewisser Informationsanteil über den Gestenanfang geht also verloren. Aus diesem Grund wird dem aufgezeichneten Musterverlauf nachträglich ein Vektorblock vorangestellt, der diese Information enthält. Dazu werden permanent Distanz- und Trendvektoren in zwei FIFO-Puffern (siehe Abb. 7.8 bzw. Abb. 7.17) der Länge th_{start} gespeichert. Nach der Detektion des Gestenendes werden die aktuellen

Inhalte der beiden Puffer der aufgezeichneten Bewegungssequenz vorangestellt, wodurch diese komplettiert wird[22].

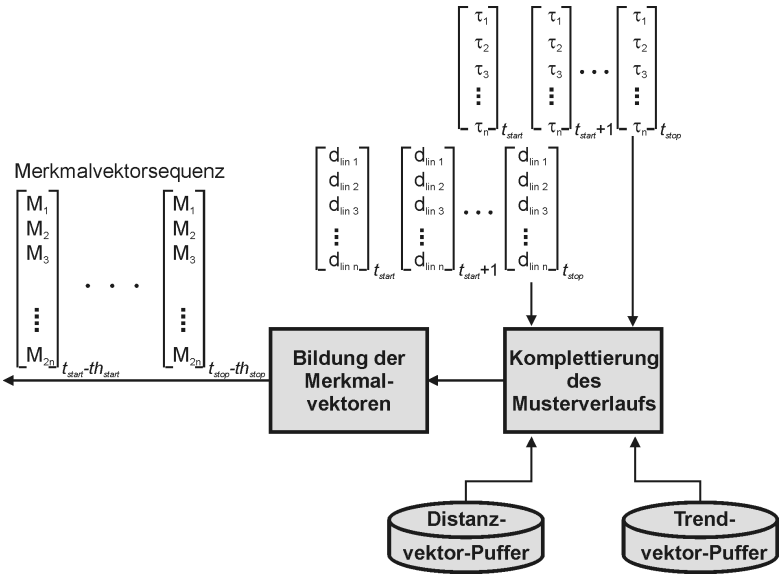


Abb. 7.17: Blockdiagramm Komplettierung des Musterverlaufs und Bildung der Merkmalvektoren.

Entsprechend wird ein gewisser Teil am Ende der komplettierten Bewegungssequenz nachträglich gelöscht. Dabei handelt es sich um diejenigen Distanz- und Trendvektoren, welche im Zeitraum von $t_{stop} - th_{stop}$ bis t_{stop} aufgezeichnet wurden und somit keine Information über die ausgeführte Geste enthalten.

Bildung der Merkmalvektoren

Der vollständige Musterverlauf besteht nun aus einer Distanzvektor- und einer Trendvektorsequenz über den Zeitraum:

$$\Delta t_{real} = t_{start,real} \ldots t_{stop,real} \quad \text{mit} \quad t_{start,real} = t_{start} - th_{start};\ t_{stop,real} = t_{stop} - th_{stop} \tag{7.15}$$

Diese werden nun zur Merkmalvektorsequenz $\underline{\underline{M}}=\underline{M}[t_{start,real}]\ldots \underline{M}[t_{stop,real}]$ zusammengefasst, indem die Komponenten eines Merkmalvektors $\underline{M}[t_j]$ folgendermaßen (für n=4 Sensoren) belegt werden:

$$\left[M_1, M_2, M_3, M_4, M_5, M_6, M_7, M_8\right]_j^{\mathrm{T}} = \left[d_{lin1}, d_{lin2}, d_{lin3}, d_{lin4}, 0, 0, 0, 0\right]_j^{\mathrm{T}} + \left[0, 0, 0, 0, \tau_1, \tau_2, \tau_3, \tau_4\right]_j^{\mathrm{T}} \tag{7.16}$$

Sie trägt die Information der ausgeführten Geste, welche später für die Klassifikation herangezogen wird.

[22] Da die Bewegungsdetektion bei der Handgestenerkennung (Objektdetektion mit th_{start}=0) sofort anspricht, ist hierbei keine Komplettierung nötig.

7.4.4 Filterung und Normierung

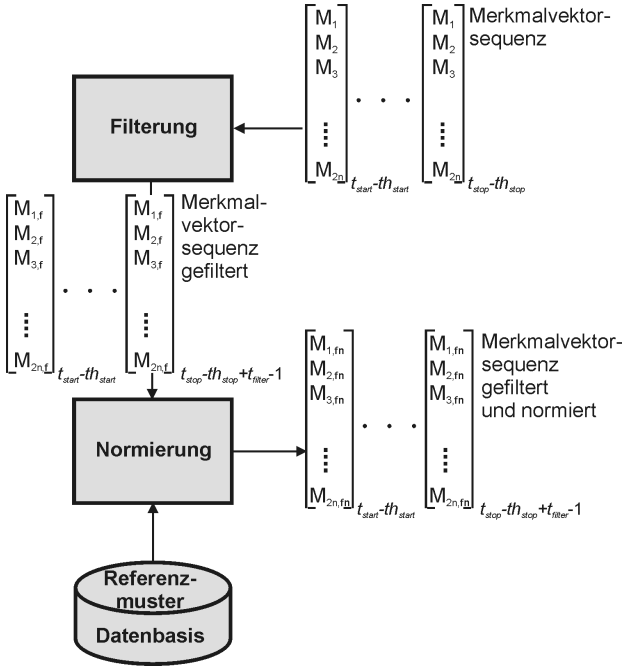


Abb. 7.18: Blockdiagramm Filterung und Normierung.

Filterung

Mit dem Ziel, die Erkennungsleistung des Gesamtsystems zu optimieren, wurde ein Filtermodul implementiert (siehe Abb. 7.18). Dadurch ist es möglich, die zeitlich segmentierte Merkmalvektorsequenz nachträglich komponentenweise zu glätten. Eine derartige Filterung erweist sich aus folgenden Gründen als vorteilhaft:

- Die Distanzmesswerte sind mit einem Grundrauschen behaftet (siehe auch Kap. 7.4.2).
- Durch Reflexionen an Objektkanten (z.B. der Hand) können einzelne Ausreißer in den Messwerten entstehen.

Bei der späteren Mustererkennung zeigte sich eine nachteilige Auswirkung dieser Effekte. Durch die Glättung soll eine Reduktion des Informationsgehaltes auf den für die Bewegung charakteristischen Signalverlauf erreicht werden. Dazu wird die Merkmalvektorsequenz komponentenweise einer zeitdiskreten Faltung mit dem Filter $h_i[t_j]$ unterzogen:

$$M_{i,f}[t_j] = M_i[t_j] * h_i[t_j] \tag{7.17}$$

$$\begin{aligned} h_i[t_j] &= h_{i,j} \quad \text{für} \quad 0 \le t_j \le t_{filter} \\ h_i[t_j] &= 0 \qquad \text{sonst} \end{aligned} \tag{7.18}$$

Die Merkmalvektoren sind für i=1...n mit Distanzmesswerten belegt. Tab. 7.6 zeigt die Parameter, welche sich für die Filterung dieses Vektorbereichs als sinnvoll erwiesen haben. Die Filteroperation entspricht somit einer ungewichteten gleitenden Mittelwertfilterung.

	t_{filter}	$h_{i,j}$
Kopfgestik	15	1/15
Handgestik	5	1/5

Tab. 7.6: Parameter der Filterimpulsantwort h für die Vektorkomponenten i=1...n.

Das relativ große Zeitfenster t_{filter}=15 bei der Kopfgestenerkennung resultiert aus der Tatsache, dass die Distanzmessungen besonders große Streuungen aufweisen. Diese werden in erster Linie dadurch hervorgerufen, dass das ausgesandte IR-Licht der Sensoren relativ diffus vom Kopf reflektiert wird. Aufgrund der inhomogenen Kopfoberfläche (Haare) kommt es vor, dass in den Empfänger eines Sensors Licht eintritt, welches nicht von diesem ausgesandt wurde. Der sich in diesem Fall ergebende Distanzmesswert weicht von der tatsächlichen Distanz ab und stellt einen Messwertausreißer dar.

Die gefilterte Merkmalvektorsequenz $\underline{\boldsymbol{M}}_{\boldsymbol{f}}$ verlängert sich durch die Faltungsoperation um (t_{filter}-1) Vektoren, so dass gilt:

$$\underline{\boldsymbol{M}}_{\boldsymbol{f}} = \{\underline{M}_f[t_{start,real}],...,\underline{M}_f[t_{stop,real} + t_{filter} - 1]\} \tag{7.19}$$

Die Merkmalvektoren sind für i= n +1...2n mit Geschwindigkeitswerten belegt. Es zeigte sich, dass die nachträgliche Filterung der Geschwindigkeitsverläufe nicht zur Verbesserung der Erkennungsleistung führt. Dies ist plausibel, da die Trendwerte bereits durch die angewandte Regressionsanalyse (siehe Kap. 7.4.2) einer optimalen Glättung unterzogen wurden. Die Parameter des Filters $h_i[t_j]$ werden daher so gewählt, dass sie lediglich einer Verschiebung entsprechen:

$$\begin{aligned} h_i[t_j] &= 1 \quad \text{für} \quad t_j = (t_{filter} - 1)/2 \\ h_i[t_j] &= 0 \qquad \text{sonst} \end{aligned} \tag{7.20}$$

Normierung

Das später hier angewandte Mustererkennungsverfahren *Dynamic Time Warping* (DTW; siehe Kap. 7.4.5) arbeitet abstandsbasiert, d.h. zur Klassifizierung eines unbekannten Musterverlaufs werden die Gesamtabstände zu allen vorhandenen Referenzmusterverläufen berechnet. Um dafür Sorge zu tragen, dass jede Komponente i einer Merkmalvektorsequenz $\underline{\boldsymbol{M}}_{\boldsymbol{f}}$ gleichberechtigt zum Gesamtabstand beiträgt, wird eine komponentenweise Normierung durchgeführt, woraus sich betragsmäßig einheitliche Dynamikbereiche innerhalb jeder Vektorkomponente ergeben. Dazu werden die einzelnen Komponenten N_i (i=1...2n) des Normierungsvektors $\underline{N}$ mit dem jeweils in der Vergangenheit beobachteten Betragsmaximum belegt. Als Datenbasis für frühere Beobachtungen wird der Trainingskorpus $\underline{\boldsymbol{T}}$ herangezogen. Dieser enthält für jede zu erkennende Gestenklasse k={1,...,K} eine bestimmte Anzahl l_k={1,...,L_k} an Referenzmusterverläufen $\underline{\boldsymbol{R}}_{k,l_k}$:

$$\underline{\boldsymbol{T}} = \{\underline{\boldsymbol{R}}_{1,l_1},...,\underline{\boldsymbol{R}}_{1,L_1},\underline{\boldsymbol{R}}_{2,l_2},...,\underline{\boldsymbol{R}}_{2,L_2},...,\underline{\boldsymbol{R}}_{K,1_K},...,\underline{\boldsymbol{R}}_{K,L_K}\} \tag{7.21}$$

Insgesamt enthält der Trainingskorpus $\underline{\boldsymbol{T}}$ also V Referenzmusterverläufe mit

$$V = \sum_{k=1}^{K} L_k \ . \tag{7.22}$$

Der Normierungsvektor $\underline{N}$ wird nun folgendermaßen belegt:

$$N_i = \max\left|R_i\right| \quad \forall\ 1 \le i \le 2n,\ \underline{R} \in \underline{\boldsymbol{T}} \tag{7.23}$$

Der Inhalt einer Komponente N_i stellt somit das zu erwartende Betragsmaximum der entsprechenden Komponente i eines Merkmalvektors $\underline{M_f}$ dar.

Da die Geschwindigkeitswerte innerhalb eines Merkmalvektors im Gegensatz zu den Distanzwerten vorzeichenbehaftet sind, wird der entsprechende Dynamikbereich halbiert. Insgesamt ergibt sich für die normierte Merkmalvektorsequenz $\underline{\boldsymbol{M}_{fn}}$ folgende Berechnungsvorschrift:

$$\begin{aligned} M_{i,\boldsymbol{fn}}[t_j] &= M_{i,\boldsymbol{f}}[t_j] / N_i \quad \text{für} \quad 1 \le i \le n \quad \text{(Distanzwerte)} \\ M_{i,\boldsymbol{fn}}[t_j] &= M_{i,\boldsymbol{f}}[t_j] / 2N_i \quad \text{für} \quad \text{n+1} \le i \le 2n \quad \text{(Geschwindigkeitswerte)} \end{aligned} \tag{7.24}$$

Da eine unbekannte Merkmalvektorsequenz größere Werte enthalten kann als die im Trainingskorpus vorhandenen Maxima, können die Grenzen der sich ergebenden Wertebereiche nicht exakt angegeben werden; diese Unsicherheit wird durch die Parameter $\delta_{1\ldots3}$ berücksichtigt, welche sich bei einem entsprechend großen Trainingskorpus dem Wert Null annähern:

$$\begin{aligned} 0 \le M_{i,\boldsymbol{fn}}[t_j] &\le 1 + \delta_1 \quad \text{für} \quad 1 \le i \le n \\ -0{,}5 - \delta_2 \le M_{i,\boldsymbol{fn}}[t_j] &\le 0{,}5 + \delta_3 \quad \text{für} \quad \text{n+1} \le i \le 2n \end{aligned} \tag{7.25}$$

Die Ermittlung des Normierungsvektors $\underline{N}$ erfolgt jeweils nach dem Einlesen der (noch unnormierten) Referenzmusterverläufe $\underline{\boldsymbol{R}}_{k,l_k}$ von der Festplatte. Diese werden daraufhin ebenfalls entsprechend Gl. 7.24 normiert.

7.4.5 Klassifikation

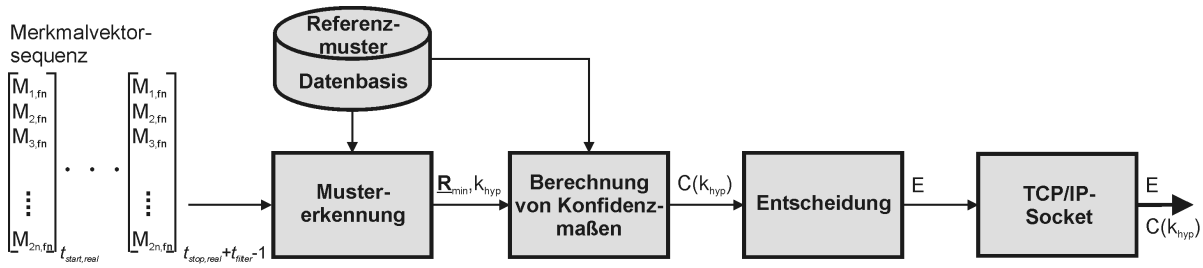


Abb. 7.19: Blockdiagramm Klassifikation.

Bei der Klassifikation wird der unbekannte Musterverlauf $\underline{\boldsymbol{M}_?}$, welcher sich aus einer Geste ergibt, mit allen vorhandenen Referenzmusterverläufen $\underline{\boldsymbol{R}}_{k,l_k}$ des Trainingskorpus $\underline{\boldsymbol{T}}$ verglichen. Dabei muss berücksichtigt werden, dass die Musterverläufe - selbst innerhalb einer Klasse k - aufgrund unterschiedlicher Ausführungsdynamik im Allgemeinen unterschiedliche Längen aufweisen. Eine Län-

genanpassung könnte durch lineare Inter- bzw. Extrapolation auf eine einheitliche Gesamtlänge (bzw. -dauer) erfolgen. Dies ist jedoch nicht sinnvoll, da beachtet werden muss, dass Bewegungsabläufe beim wiederholten Ausführen ein und derselben Geste nichtlineare zeitliche Verzerrungen aufweisen können. Es handelt sich hierbei um eine Problematik, die auch bei der Spracherkennung eine wichtige Rolle spielt (siehe z.B. [RUS97]). Das implementierte Klassifikationsverfahren muss also in der Lage sein, derartige Zeitverzerrungen zu berücksichtigen.

Mustervergleich

Zur Berechnung eines Gesamtabstands zwischen dem unbekannten Musterverlauf $\underline{\boldsymbol{M}}_?$ und einem Referenzmusterverlauf (im Weiteren *Referenz*) $\underline{\boldsymbol{R}}_{k,l_k}$ wird der sogenannte *Dynamic Time Warping* Algorithmus (DTW) verwendet. Dieser ermittelt die optimale zeitliche Verzerrungs- bzw. Zuordnungsfunktion, die den Gesamtabstand $D_{DTW}(\underline{\boldsymbol{M}}_?,\underline{\boldsymbol{R}}_{k,l_k})$ zwischen zwei Vektorsequenzen minimiert. Es handelt sich somit um einen *Abstandsklassifikator*.

Zur Berechnung des Gesamtabstands zwischen den beiden Musterverläufen $\underline{\boldsymbol{M}}_?$ und $\underline{\boldsymbol{R}}$ der Längen O bzw. P

$$\begin{aligned} \underline{M}_o &\in \underline{\boldsymbol{M}}_? \quad \text{mit} \quad 1 \le o \le O \\ \underline{R}_p &\in \underline{\boldsymbol{R}} \quad \text{mit} \quad 1 \le p \le P \end{aligned} \tag{7.26}$$

wird zunächst eine Distanzmatrix $\underline{\boldsymbol{D}}_{dist}$ aufgestellt, deren Komponenten $D_{dist}(o,p)$ jeweils den euklidischen Abstand der Merkmalvektoren $\underline{M}_o$ und $\underline{R}_p$ enthalten:

$$D_{dist}(o,p) = \left| \underline{M}_o - \underline{R}_p \right| \quad \text{für} \quad 1 \le o \le O;\ 1 \le p \le P \tag{7.27}$$

Nach der anschließenden Berechnung der akkumulierten Distanzmatrix $\underline{\boldsymbol{D}}_{akk}$ nach folgender Berechnungsvorschrift (siehe auch [RUS97])

$$\begin{aligned} \text{Anfangsbedingung:} \quad D_{akk}(o,p) &= D_{dist}(1,1) \quad \text{für} \quad o=1;\ p=1 \\ D_{akk}(j,k) = \min \begin{cases} D_{akk}(o-1,p) & + \quad D_{dist}(o,p) \quad \text{für} \quad 2 \le o \le O;\ 1 \le p \le P \\ D_{akk}(o-1,p-1) & + \quad 2D_{dist}(o,p) \quad \text{für} \quad 2 \le o \le O;\ 2 \le p \le P \\ D_{akk}(o,p-1) & + \quad D_{dist}(o,p) \quad \text{für} \quad 1 \le o \le O;\ 2 \le p \le P \end{cases} \end{aligned} \tag{7.28}$$

enthält das Matrixelement $D_{akk}(O,P)$ den minimalen Gesamtabstand

$$D_{DTW}(\underline{\boldsymbol{M}}_?,\underline{\boldsymbol{R}}) = D_{akk}(J,K)\ . \tag{7.29}$$

Auf diese Weise werden alle V (siehe Gl. 7.22) Gesamtabstände $D_{DTW,v}$ ($v=\{1,...,V\}$) des unbekannten Musterverlaufs $\underline{\boldsymbol{M}}_?$ zu allen im Trainingskorpus enthaltenen Referenzen $\underline{\boldsymbol{R}}_{k,l_k}$ berechnet. Die einfachste Methode, eine Klassifikationsentscheidung für $\underline{\boldsymbol{M}}_?$ zu treffen, besteht in der Zuordnung zu jener Klasse k, welche diejenige Referenz $\underline{\boldsymbol{R}}_{hyp}$ enthält, für den sich der kleinste Gesamtabstand $D_{DTW,min}$ ergibt. Es lassen sich zunächst jedoch keine Angaben über die Zuverlässigkeit der Entscheidung treffen. Hierfür bedarf es der Berechnung spezieller Konfidenzmaße unter Berücksichtigung zusätzlicher Informationen, die im folgenden Kapitel erläutert wird.

Berechnung von Konfidenzmaßen und Entscheidung

Konfidenzmaße dienen zur Bewertung der Zuverlässigkeit einer Klassifikationsentscheidung. Dies beinhaltet auch die Forderung, dass sich die Güte bzw. die Qualität der Ausführung einer Geste im Konfidenzmaß widerspiegeln muss. So soll sich bei „sauberer" Ausführung einer im Trainingskorpus enthaltenen Geste ein hoher *Score* (d.h. Konfidenzmaßwert) ergeben, wohingegen aus einer schlecht ausgeführten (oder gar unbekannten) Geste ein entsprechend niedriger Score resultieren soll. Ein derartiges Konfidenzmaß ermöglicht einerseits berechtigte Rückweisungen von Klassifikationshypothesen bei der Unterschreitung eines bestimmten Schwellwerts C_{min} (siehe *Entscheidung* in diesem Kap.). Andererseits soll das Konfidenzmaß später als Eingangsparameter (Merkmal) für ein automatisches Hilfe-System dienen, welches daraus Rückschlüsse auf die Unsicherheit bzw. Hilfebedürftigkeit des Benutzers zieht (siehe [NIE01] und [NIE02B]).

Der letztlich verwendete Algorithmus zur Konfidenzbewertung ergab sich aus der empirischen Auswertung plausibler Berechnungsvorschriften. Es wurden jeweils verschiedene Größen betrachtet, aus denen die Korrelation zwischen tatsächlich ausgeführter Geste und der Klassifikationshypothese k_{hyp} hervorgehen sollte. Dabei liegt allen angewandten Methoden die Überlegung zu Grunde, dass die jeweilige Größe in unterschiedlicher Weise widerspiegelt, wie sehr sich die Zuordnung des unbekannten Musters zur Hypothesenklasse von der zu anderen Klassen abhebt. Schließlich kristallisierten sich vier Verfahren heraus, die den oben genannten Anforderungen an ein sinnvolles Konfidenzmaß genügen. Sie liefern die Teilkomponenten c_1, c_2, c_3 und c_4, deren gewichtete Summe das endgültige Gesamt-Konfidenzmaß C_{ges} ergibt:

$$C_{ges} = \alpha \cdot c_1 + \beta \cdot c_2 + \chi \cdot c_3 + \delta \cdot c_4; \quad \alpha + \beta + \chi + \delta = 1 \tag{7.30}$$

Da die jeweilige Aussagekraft der Teilkomponenten - wie im Folgenden beschrieben - stark von den vorliegenden Randbedingungen abhängt, muss die Optimierung der Gewichtungsfaktoren α, β, χ und δ fallspezifisch erfolgen.

• Hypothese

Zunächst wird die *Hypothese* aufgestellt, dass der unbekannte Musterverlauf $\underline{\boldsymbol{M}}_?$ der Klasse k_{hyp} zuzuordnen ist. Dabei handelt es sich um diejenige Klasse, welche die Referenz $\underline{\boldsymbol{R}}_{hyp}$ mit dem minimalen Abstand D_{hyp} zum unbekannten Musterverlauf $\underline{\boldsymbol{M}}_?$ enthält:

$$\begin{aligned} k_{hyp} &= \arg\min_{k,l_k} \{D_{DTW}(\underline{\boldsymbol{M}}_?, \underline{\boldsymbol{R}}_{k,l_k})\} \quad \text{für} \quad k \in \{1,...,K\};\ l_k \in \{1,...,L_k\} \\ D_{hyp} &= D_{DTW}(\underline{\boldsymbol{M}}_?, \underline{\boldsymbol{R}}_{hyp}) \end{aligned} \tag{7.31}$$

Diese Hypothese soll von den im Folgenden beschriebenen Konfidenzmaßen hinsichtlich ihrer Zuverlässigkeit bewertet werden.

• Konfidenzmaß c_1

Als sehr einfaches und dennoch aussagekräftiges Maß erwies sich das Verhältnis zwischen dem sich ergebenden Minimalabstand D_{hyp} und dem Abstand $D_{next}=D_{DTW}(\underline{\boldsymbol{M}}_?,\underline{\boldsymbol{R}}_{next})$ zur „nächstbesten" Referenz $\underline{\boldsymbol{R}}_{next}$. Dies ist diejenige Referenz $\underline{\boldsymbol{R}}_{next}$, die den kleinsten Gesamtabstand zu $\underline{\boldsymbol{M}}_?$ aufweist und gleichzeitig *nicht* der Hypothesenklasse k_{hyp} angehört (siehe auch Gl. 7.33).

Mit der zusätzlichen Anforderung, einen sinnvollen Wertebereich ($0 \le c_1 \le 1$) zu erhalten, wird das Konfidenzmaß c_1 nun folgendermaßen definiert:

$$c_1 = 1 - \frac{D_{hyp}}{D_{next}} \tag{7.32}$$

Somit ergeben sich für c_1 kleine Werte ($c_1 \to 0$), wenn der Gesamtabstand von $\underline{\boldsymbol{M}}_?$ zur Referenz $\underline{\boldsymbol{R}}_{next}$ nur unwesentlich größer ist als der zu $\underline{\boldsymbol{R}}_{hyp}$ bzw. entsprechend hohe Werte ($c_1 \to 1$) für den umgekehrten Fall, dass das unbekannte Muster der besten Referenz deutlich ähnlicher ist als der nächstbesten.

Die Zuverlässigkeit dieses Konfidenzmaßes hängt stark von der Güte des Trainingskorpus ab. Aussagekräftige Werte ergeben sich nur dann, wenn beim Training sichergestellt wurde, dass jede Klasse ausschließlich repräsentative Referenzen, d.h. keine „Ausreißer", enthält. Eine hohe Gewichtung der Komponente c_1 setzt also ein entsprechend handverlesenes Trainingsmaterial voraus und bietet sich insbesondere bei solchen Trainingskorpora an, die nur wenige Referenzen pro Klasse[23] enthalten.

• Konfidenzmaß c_2

Die Komponente c_2 ist ein Maß dafür, wie sicher der unbekannte Musterverlauf $\underline{\boldsymbol{M}}_?$ *insgesamt* der Hypothesenklasse k_{hyp} zugeordnet werden kann, d.h. unter Betrachtung *aller* in ihr enthaltenen Referenzen. Als konkurrierende Klasse wird die „nächstbeste" Klasse k_{next} festgelegt. Dies ist diejenige Klasse, welche die Referenz $\underline{\boldsymbol{R}}_{next}$ (siehe *Konfidenzmaß* c_1) enthält:

$$k_{next} = \arg\min_{k,l_k} \{D_{DTW}(\underline{\boldsymbol{M}}_?, \underline{\boldsymbol{R}}_{k,l_k})\} \quad \text{für} \quad k \in \{1,...,K\} \wedge k \neq k_{hyp};\ l_k \in \{1,...,L_k\} \tag{7.33}$$

Nun werden die *mittleren Gesamtabstände* $\overline{D}_{hyp}$ und $\overline{D}_{next}$ des unbekannten Musterverlaufs $\underline{\boldsymbol{M}}_?$ zu den Klassen k_{hyp} und k_{next} berechnet:

[23] Im Extremfall wird jede Klasse durch nur *eine* Referenz repräsentiert.

$$\bar{D}_{hyp} = \frac{1}{L_k} \sum_{l_k=1}^{L_k} D_{DTW}(\underline{\boldsymbol{M}}_?, \underline{\boldsymbol{R}}_{k,l_k}); \quad k = k_{hyp} \tag{7.34}$$

$$\bar{D}_{next} = \frac{1}{L_k} \sum_{l_k=1}^{L_k} D_{DTW}(\underline{\boldsymbol{M}}_?, \underline{\boldsymbol{R}}_{k,l_k}); \quad k = k_{next} \tag{7.35}$$

Der *mittlere Gesamtabstand* $\bar{D}_k$ zu einer Klasse k lässt darauf schließen, wie gut ein unbekanntes Muster insgesamt in diese Klasse „passt". Dabei wird davon ausgegangen, dass sich die Referenzen innerhalb einer Klasse einander ähnlicher sind als denen anderer Klassen, da sie jeweils aus den Bewegungsabläufen ein und derselben Geste hervorgingen.

Das Konfidenzmaß c_2 wird nun folgendermaßen definiert:

$$c_2 = \min\{1 - \frac{\bar{D}_{hyp}}{\bar{D}_{next}}, 0\}^{24} \tag{7.36}$$

Die Komponente c_2 liefert insbesondere dann gute Ergebnisse, wenn sich die Bewegungsabläufe der im Trainingskorpus enthaltenen Gesten von Klasse zu Klasse stark unterscheiden. Sie ist allerdings sehr anfällig für *Out-Of-Vocabulary-Fälle*, die bei der Ausführung von Bewegungsabläufen, die nicht im Trainingskorpus $\underline{\boldsymbol{T}}$ enthalten sind, entstehen. Die Zuordnung zur zufällig ähnlichsten Klasse erfolgt dann häufig mit unerwünscht hohen Scores.

• Konfidenzmaß c_3

Das Konfidenzmaß c_3 erlaubt - ähnlich wie c_2 - eine Aussage darüber, wie gut sich die „Passung" des unbekannten Musterverlaufs in die Hypothesenklasse k_{hyp} von der in andere Klassen abhebt. Dabei wird allerdings nicht die nächstbeste Klasse zum Vergleich herangezogen, sondern *alle* verbleibenden Klassen, die durch die Menge $\boldsymbol{k}_{rest}$ angegeben werden:

$$\boldsymbol{k}_{rest} = \{1, ..., K\} \setminus k_{hyp} \tag{7.37}$$

Hierfür wird der *mittlere Gesamtabstand* $\bar{D}_{rest}$ des Musterverlaufs $\underline{\boldsymbol{M}}_?$ zu allen Klassen der Menge $\boldsymbol{k}_{rest}$ berechnet:

$$\bar{D}_{rest} = \frac{1}{V - K'} \sum^{k \in \boldsymbol{k}_{rest}} \sum_{l_k=1}^{L_k} D_{DTW}(\underline{\boldsymbol{M}}_?, \underline{\boldsymbol{R}}_{k,l_k}); \quad K' = |\boldsymbol{k}_{rest}|^{25} \tag{7.38}$$

---

[24] Durch den min-Operator wird das Konfidenzmaß bei dem sehr selten auftretenden Fall $\bar{D}_{hyp} > \bar{D}_{next}$ auf Null gesetzt.

[25] K' ist die Anzahl der Elemente der Menge $\boldsymbol{k}_{rest}$; V ist die Gesamtanzahl der vorhandenen Referenzmusterverläufe (siehe Gl. 7.22).

Entsprechend der Vorgehensweise für c_2 wird nun das Konfidenzmaß c_3 definiert:

$$c_3 = \min\{1 - \frac{\bar{D}_{hyp}}{\bar{D}_{rest}}, 0\} \tag{7.39}$$

Das Konfidenzmaß c_3 liefert gute Ergebnisse, wenn die vorhandenen Gestenklassen möglichst homogen im Merkmalraum verteilt liegen. Es erweist sich dann bei *Out-Of-Vocabulary-Fällen* robuster als c_2. Hier ergeben sich für c_3 entsprechend niedrige Werte, da der mittlere Gesamtabstand zur Hypothesenklasse $\bar{D}_{hyp}$ nur geringfügig kleiner ist als $\bar{D}_{rest}$ zu allen anderen Klassen. Die Verlässlichkeit sinkt jedoch drastisch, wenn sich mindestens *eine* der vorhandenen Klassen sehr stark von den übrigen unterscheidet, da sich für den mittleren Gesamtabstand $\bar{D}_{rest}$ dann generell relativ hohe Werte ergeben, woraus selbst bei Fehlerkennungen entsprechend hohe Konfidenzmaße c_3 resultieren. In diesem Fall liegt es nahe, zur Berechnung von $\bar{D}_{rest}$ nicht *alle* Klassen der Menge $\boldsymbol{k}_{rest}$ heranzuziehen, sondern nur diejenigen, welche einen bestimmten mittleren Maximalabstand $\bar{D}_{\max}$ zur Hypothesenklasse nicht überschreiten.

- Konfidenzmaß c_4

Wie bereits erwähnt (siehe *Konfidenzmaß* c_2), kann davon ausgegangen werden, dass die Referenzen *innerhalb* einer Klasse einander ähnlich sind, also einen relativ geringen Gesamtabstand zueinander aufweisen. Es ist somit plausibel, dass sich bei der *korrekten* Erkennung einer Geste eine gewisse Anzahl N_{hyp} an Referenzen $\underline{\boldsymbol{R}}_{k_{hyp},l}$ der Hypothesenklasse k_{hyp} unter den *N Besten*[26] Referenzen befindet. Es erwies sich als günstig, diese Anzahl N_{hyp} zur Berechnung des Konfidenzmaßes c_4 heranzuziehen.

Die Anzahl N der jeweils betrachteten *N Besten* ist variabel und wird gleichgesetzt mit der Anzahl der in der Hypothesenklasse vorhandenen Referenzen:

$$N = L_k \quad \text{mit} \quad k = k_{hyp} \tag{7.40}$$

Das Konfidenzmaß c_4 wird nun folgendermaßen definiert:

$$c_4 = \frac{N_{hyp}}{N} \tag{7.41}$$

Es leuchtet ein, dass diese Vorgehensweise nur dann Sinn macht, wenn für jede Klasse eine bestimmte Mindestanzahl $l_{min}>1$ an Referenzen vorhanden ist. Die Verlässlichkeit des Konfidenzmaßes c_4 setzt also einen entsprechend großen Trainingskorpus $\underline{\boldsymbol{T}}$ voraus; für $l_{min} \geq 5$ ergaben sich sinnvolle Resultate. Das Konfidenzmaß c_4 erweist sich als besonders robust gegen einzelne Ausreißer im Trainingsmaterial.

---

[26] Die „*N Besten*“ sind diejenigen Referenzen, welche die N geringsten Gesamtabstände zum unbekannten Muster aufweisen.

• Normierung der Konfidenzmaßkomponenten

Bei der hier angewandten DTW-Klassifikation ist zu beobachten, dass der beim Vergleich zweier Musterverläufe *derselben* Klasse resultierende Gesamtabstand mit zunehmender Komplexität des zugrunde liegenden Bewegungsablaufs größer wird. Grund dafür ist die Tatsache, dass bei wiederholter Ausführung einer komplizierten Geste - selbst von ein und derselben Person - zwangsläufig größere Variationen auftreten, als dies bei kurzen bzw. einfachen Gesten der Fall ist[27]. Dies führt dazu, dass komplexe Bewegungsabläufe generell schlechter aneinander angeglichen werden können als einfache und daher letztlich auch niedrigere Scores liefern.

Um diese Benachteiligung komplexer Gesten bei der Konfidenzbewertung zu berücksichtigen, wurden die klassenspezifischen Normierungsfaktoren $c_{1norm,k},..., c_{4norm,k}$ eingeführt, mit denen das Gesamt-Konfidenzmaß C_{ges} nachgewichtet wird (siehe Gl. 7.42).

$$C_{ges,norm} = \alpha \cdot c_1 \cdot c_{1norm,k_{hyp}}^{-1} + \beta \cdot c_2 \cdot c_{2norm,k_{hyp}}^{-1} + \chi \cdot c_3 \cdot c_{3norm,k_{hyp}}^{-1} + \delta \cdot c_4 \cdot c_{4norm,k_{hyp}}^{-1} \tag{7.42}$$

Dabei werden die Normierungsfaktoren $c_{1norm,k},..., c_{4norm,k}$ mit dem jeweils höchsten klassenspezifischen Konfidenzmaß $C_{1max,k},..., C_{4max,k}$, das sich bei der Reklassifikation des Trainingskorpus (siehe auch Kap. 7.7.5) ergab, belegt. Diese Maxima können als Maß dafür betrachtet werden, wie gut - d.h. mit welchen typischen Einzelkonfidenzwerten - eine bestimmte Geste unter optimalen Bedingungen erkannt wird.

• Gewichtung der Konfidenzmaßkomponenten

Da die jeweilige Aussagekraft der beschriebenen Konfidenzmaße von den vorliegenden Randbedingungen (z.B. Umfang des Trainingskorpus, Trennbarkeit der Klassen etc.) abhängt, erschien es sinnvoll, sie zunächst als gewichtete Summe im Gesamt-Konfidenzmaß C_{ges} bzw. $C_{ges,norm}$ zu vereinen (siehe Gl. 7.30 und Gl. 7.42). Dieses lässt sich wiederum flexibel durch die Wahl der Gewichtungsfaktoren α, β, χ und δ für den konkreten Einsatz optimieren. Beispielsweise kann die Gewichtung jedes Faktors proportional zu seiner Korrelation mit der Gestenausführungsqualität erfolgen. Zur Ermittlung dieser Korrelationen müssen umfangreiche Testreihen mit Gesten unterschiedlicher Qualität durchgeführt werden, welche jeweils subjektiv zu beurteilen sind.

Als Variante bietet sich eine automatisierte Vorgehensweise an. Zur Bewertung der einzelnen Konfidenzmaßkomponenten wird lediglich betrachtet, wie sie sich jeweils bei korrekten bzw. falschen Klassifikationen verhalten. Dabei wird an ein „gutes“ Konfidenzmaß die Anforderung gestellt, dass es bei korrekten Klassifikationen im Mittel höhere Werte liefern muss als bei falschen.

Zur Bestimmung der Gewichtungsfaktoren wird ein typischer Testdatensatz am trainierten System evaluiert, wobei die einzelnen Konfidenzmaße sowie die Klassifikationsergebnisse (korrekte bzw. falsche Klassifikation) aufgezeichnet werden. Abb. 7.20 zeigt die Mediane sowie die ersten und

[27] Als Beispiel für eine komplexe Geste sei hier das *Abheben eines virtuellen Telefonhörers* genannt, wohingegen eine *Winkbewegung nach rechts oder links* eine relativ einfache Geste darstellt.

dritten Quartile der Konfidenzmaßkomponenten, die sich bei der Klassifikation von 600 Testgesten (12 Klassen; 25 Referenzen und 50 Testmuster pro Klasse) ergaben.

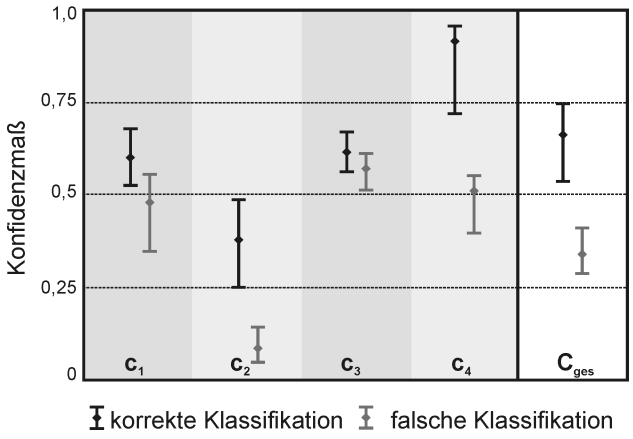


Abb. 7.20: Mediane sowie 1. und 3. Quartil der Konfidenzmaßkomponenten $c_1,...,c_4$ und des Gesamtkonfidenzmaßes C_{ges} (gewichtet nach Gl. 7.45; keine Normierung nach Gl. 7.42) aufgeteilt nach korrekter und falscher Klassifikation.

Erwartungsgemäß weisen die Mediane bei falscher Klassifikation grundsätzlich geringere Werte auf als bei korrekter. Es ist jedoch offensichtlich, dass sich die einzelnen Konfidenzmaße unterschiedlich gut als Indikator für die Verlässlichkeit einer Klassifikation eignen. Als besonders wenig aussagekräftig erweist sich in diesem Setup die Komponente c_3, da sich einerseits die Quartilbereiche stark überlappen und die Mediane zudem sehr nahe zusammen liegen - sie sollte daher entsprechend gering gewichtet werden.

Als einfache Methode zur Bestimmung der Gewichtungsfaktoren α, β, χ und δ werden die Differenzen der Mediane Δmed_k für korrekte und falsche Klassifikation herangezogen:

$$\Delta med_k = \text{med}(c_k; \text{korrekte Klassifikation}) - \text{med}(c_k; \text{falsche Klassifikation}) \quad \text{für } k \in \{1,...,4\} \tag{7.43}$$

Die Gewichtung erfolgt nun proportional zur Differenz Δmed_k:

$$\alpha = \frac{\Delta med_1}{N_{med}};\ \beta = \frac{\Delta med_2}{N_{med}};\ \chi = \frac{\Delta med_3}{N_{med}};\ \delta = \frac{\Delta med_4}{N_{med}} \quad \text{mit} \quad N_{med} = \sum_{k=1}^{4} \Delta med_k \tag{7.44}$$

Für das oben genannte Beispiel (siehe Abb. 7.20) ergeben sich nach Gl. 7.44 folgende Gewichtungsfaktoren:

$$\alpha = 0{,}139;\ \beta = 0{,}348;\ \chi = 0{,}052;\ \delta = 0{,}461 \tag{7.45}$$

Wie gefordert, liefert das nach Gl. 7.45 gewichtete Gesamtkonfidenzmaß C_{ges} bei falschen Klassifikationen deutlich geringere Scores als bei korrekten (siehe Abb. 7.20 rechts).

• Entscheidung

Nach der erfolgten Klassifikation muss das Erkennungssystem letztlich eine möglichst optimale Entscheidung treffen. Dabei wird hier die Philosophie verfolgt, dass die Auswirkungen einer Fehl-

entscheidung nachteiliger sind als die einer Rückweisung, bei der die vorliegende Geste als „nicht erkannt“ eingestuft wird. Als Entscheidungskriterium wird das nach Gl. 7.30 (bzw. Gl. 7.42) berechnete Konfidenzmaß $C_{ges}(k_{hyp})$ herangezogen, welches mit dem Entscheidungsschwellwert $0 \leq C_{min} \leq 1$ verglichen wird. Für die Entscheidung E wird folgende Regel angewandt:

$$E = \begin{cases} \text{"Geste der Klasse } k_{hyp} \text{ erkannt"} & \text{für } C_{ges}(k_{hyp}) \geq C_{\min} \\ \text{"keine Geste erkannt"} & \text{sonst} \end{cases} \tag{7.46}$$

Die Entscheidung E wird daraufhin zusammen mit dem zugehörigen Konfidenzmaß C_{ges} (k_{hyp}) an die angeschlossene Gesten-Applikation (via TCP/IP; siehe Abb. 7.8) übertragen. Bei der Wahl des Entscheidungsschwellwerts sollte eine Abstimmung durch empirische Tests mit dem jeweils vorliegenden Randbedingungen erfolgen. In oben genanntem Beispiel (12 Klassen, 25 Referenzen pro Klasse; siehe Abb. 7.20) ergibt sich für C_{min}=0,5 ein günstiges Verhältnis zwischen den Häufigkeiten korrekter Entscheidungen und korrekter Rückweisungen.

Bei entsprechend niedriger Wahl des Schwellwerts C_{min} (z.B. C_{min}=0,1) kann die finale Entscheidung zur Applikation verlagert werden, falls diese über eine eigene Entscheidungsinstanz verfügt. Dadurch ist es möglich, die applikationsinterne Entscheidungsschwelle abhängig von Systemzustand und Gestentyp zu variieren. Dies erlaubt z.B. eine kontextabhängige Anpassung von C_{min} an die jeweilige Schwere der Folgen einer Fehlerkennung.

7.4.6 Moduswechsel und Ermittlung der Regeldistanz

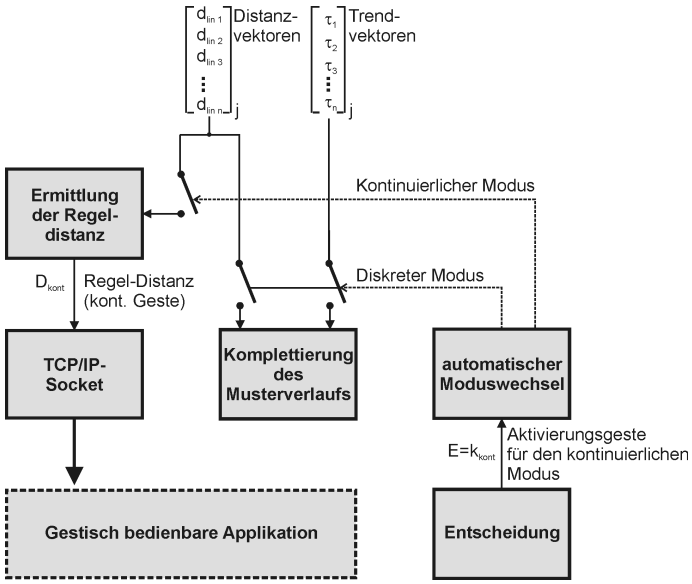


Abb. 7.21: Blockdiagramm Moduswechsel.

Die bislang erfolgte Systembeschreibung bezog sich ausschließlich auf die Erkennung *diskreter* Hand- und Kopfgesten. In den Handgestenerkenner wurde darüber hinaus eine Instanz zur Verarbeitung *kontinuierlicher Handgesten* implementiert (siehe Kap. 4.1). Dies ermöglicht das stufenlose Einstellen einer Regelgröße - z.B. der Musiklautstärke einer angeschlossenen Applikation - durch vertikale Handbewegungen. Die Wertzuweisung der Regelgröße erfolgt proportional zur *z*-Position der Hand im Sensorfeld und wird mit der Abtastfrequenz $f_{abt,Hand}$=25 Hz aktualisiert.

Moduswechsel zwischen diskreter und kontinuierlicher Gestenerkennung

Das Erkennungssystem wird also je nach Art der Gestik entweder im *kontinuierlichen* oder im *diskreten* Modus betrieben, wobei letzterer der Standardmodus ist. Der Moduswechsel vom diskreten in den kontinuierlichen Modus wird vom Benutzer durch eine spezielle (diskrete) Aktivierungsgeste k_{kont} initiiert. Die Aktivierungsgeste wird ausgeführt, indem die flache Hand in das Sensorfeld bewegt wird und dort für eine kurze Zeit stillsteht (*zeitkodierter Moduswechsel*; siehe Kap. 4.3.2). Nach einer erfolgten Klassifikationsentscheidung $E=k_{kont}$ für die Aktivierungsgeste wechselt das System so lange in den kontinuierlichen Modus, bis die Hand den Erfassungsbereich der Sensoren wieder verlassen hat (gemäß der Gestenende-Detektion; siehe Gl. 7.13). Danach wird automatisch in den diskreten Modus zurückgewechselt.

Ermittlung der Regeldistanz

Im kontinuierlichen Modus erfolgt keine Mustererkennung; die aktuellen Distanzvektoren werden an ein Modul umgeleitet, welches aus jeweils einem Vektor $\underline{d}_{lin}[t_j]$ den repräsentativen Distanzwert $D_{kont}[t_j]$ ermittelt (siehe Abb. 7.21). Dazu werden die Vektorkomponenten $d_{lin,4}$ und $d_{lin,5}$ herangezogen, welche von den Sensoren *S4* bzw. *S5* stammen (siehe Kap. 7.3.2.; *Sensor-Array für die Handgestenerkennung*). Während Sensor *S4* vom Typ *DS1* (Messbereich 100...800 mm) ist, handelt es sich bei *S5* um einen Nahbereichsensor des Typs *DS2* (Messbereich 40...300 mm). Durch die kombinierte Auswertung der beiden Sensorsignale zur Bestimmung des Distanzwerts D_{kont} wird der Ausführungsbereich der kontinuierlichen Gestik theoretisch auf 40...800 mm erweitert. Wie bei der diskreten Gestik wird jedoch auch hier eine Hintergrundausblendung (siehe Gl. 7.4) durchgeführt, so dass sich ein Gesamtbereich von 40 mm...400 mm längs der z-Achse ergibt.

Zunächst werden die Distanzwerte $d_{lin,4}[t_j]$ und $d_{lin,5}[t_j]$ einer gleitenden Mittelwertfilterung (Vorgehensweise gemäß Gl. 7.17) unterzogen, um das vorhandene Signalrauschen zu reduzieren.

Abhängig von der tatsächlichen Distanz zwischen Hand und Sensorfeld muss nun eine Entscheidung darüber getroffen werden, welcher der beiden Distanzmesswerte $d_{lin,4}$ und $d_{lin,5}$ als Regeldistanz D_{kont} herangezogen werden soll. Diese Entscheidung muss sicherstellen, dass der verwendete Distanzwert aus einer Messung im zulässigen Arbeitsbereich des zugehörigen Sensors stammt. Dazu wird zunächst die Distanzschwelle D_{sw} festgelegt, welche in folgendem Wertebereich liegen muss:

$$d_{min,DS1} < D_{sw} < d_{max,DS2}; \quad d_{min,DS1} = 100 \text{ mm}; \quad d_{max,DS2} = 300 \text{ mm} \tag{7.47}$$

Die Entscheidung könnte nun folgendermaßen definiert werden:

$$D_{kont} = \begin{cases} d_{lin,5} & \text{für} \quad d_{lin,5} < D_{sw} \quad \text{(Distanzwert des Nahbereichsensors)} \\ d_{lin,4} & \text{für} \quad d_{lin,5} > D_{sw} \wedge d_{lin,4} > D_{sw} \quad \text{(Distanzwert des Fernbereichsensors)} \end{cases} \tag{7.48}$$

Bei dieser Vorgehensweise ergaben sich jedoch an der Umschaltposition D_{sw} unerwünschte Sprünge des Distanzwertes D_{kont}, da die beiden Komponenten $d_{lin,4}$ und $d_{lin,5}$ nie exakt identische Distanzwerte aufweisen. Dies ist einerseits auf Ungenauigkeiten bei der Messung zurückzuführen und andererseits auf die Tatsache, dass sich die Hand meist nicht genau waagerecht im Sensorfeld be-

findet, wodurch sich tatsächlich unterschiedliche Distanzen zu den Sensoren *S4* bzw. *S5* ergeben. Zur Vermeidung der Distanzwertsprünge am Umschaltpunkt D_{sw} werden daher die zwei Sigmoidfunktionen σ_1 und σ_2 zur Gewichtung der Distanzmesswerte $d_{lin,4}$ und $d_{lin,5}$ eingeführt:

$$\begin{aligned}\sigma_1(d_{lin}) &= 0,5\cdot[1-\tanh(f\cdot(d_{lin}-D_{sw}))]\\ \sigma_2(d_{lin}) &= 1-0,5\cdot[1-\tanh(f\cdot(d_{lin}-D_{sw}))]\end{aligned} \tag{7.49}$$

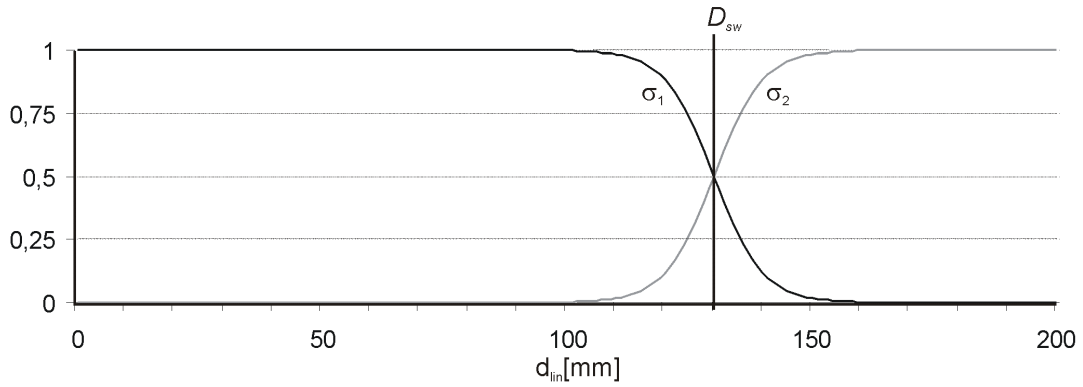


Abb. 7.22: Sigmoidfunktionen σ_1 und σ_2 zur Gewichtung der Distanzmesswerte $d_{lin,4}$ und $d_{lin,5}$ mit D_{sw}=130 mm, f=1.

Die Regeldistanz D_{kont} wird nun definiert als die Summe der nach Gl. 7.50 gewichteten Distanzmesswerte $d_{lin,4}$ und $d_{lin,5}$:

$$D_{kont} = \sigma_1(d_{lin,5})\cdot d_{lin,5} + \sigma_2(d_{lin,4})\cdot d_{lin,4} \tag{7.50}$$

Für die in Abb. 7.22 dargestellte Parameterwahl (f=1; D_{sw}=130[mm]) erfolgt der gewünschte „weiche“ Übergang bei der Speisung der Regeldistanz D_{kont} mit den beiden Sensorsignalen $d_{lin,4}$ und $d_{lin,5}$.

Bei der Konstruktion der Messvorrichtung sollte zudem dafür gesorgt werden, dass die Distanz zwischen Hand und Sensor den minimal messbaren Abstand $d_{min,DS2}$=40 mm des Nahbereichsensors nicht unterschreiten kann. Dies kann beispielsweise durch das Anbringen einer Abdeckungsscheibe in entsprechendem Abstand zu den Sensoren geschehen.

7.5 Implementierung

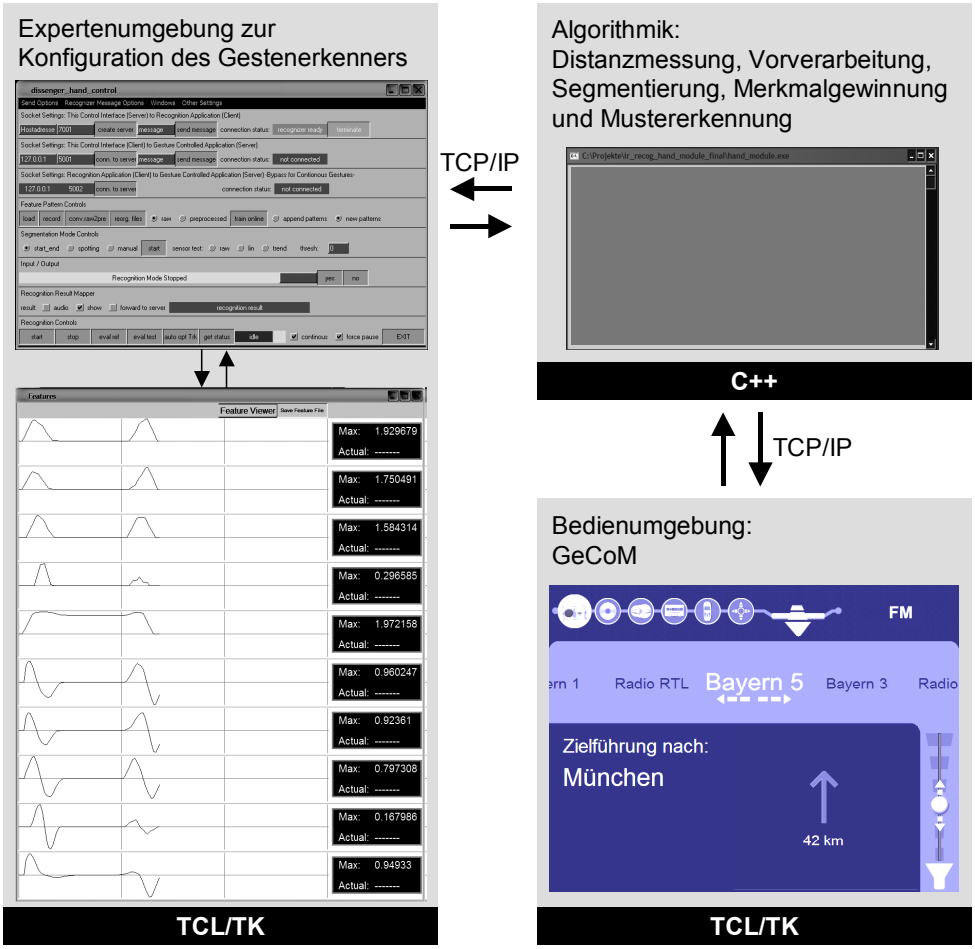


Abb. 7.23: Implementierung der Einzelmodule; Expertenumgebung bestehend aus Konfigurationsoberfläche (links oben) und Merkmal-Visualisierungsmonitor (links unten), Programmkern der Gestenerkennung (rechts oben) und GECOM als gestische Bedienumgebung (rechts unten).

Abb. 7.23 zeigt eine schematische Darstellung der vorhandenen Einzelmodule sowie die grobe Kommunikationsstruktur über TCP/IP-Netzwerkverbindungen. Die Implementierung des Gestenerkennungssystems erfolgte PC-basiert unter Verwendung des Betriebssystems MICROSOFT WINDOWS 98. Dabei wurden alle innerhalb dieses Kapitels beschriebenen Algorithmen in der Programmiersprache C++ umgesetzt. Zudem existiert eine grafische Expertenumgebung, die zur Konfiguration des Erkenners dient. Sie wurde in der plattformunabhängigen Skriptsprache Tcl/Tk implementiert und beinhaltet neben einem Visualisierungsmonitor zur grafischen Anzeige der Musterverläufe eine Bedienoberfläche für folgende Einstellungen:

- Vorgabe von Erkennungsparametern (z.B. Empfindlichkeit der Bewegungsdetektion und Parameter für die Merkmal-Normierung sowie für die Konfidenzmaßberechnung)
- Verwaltung der Trainingsdatensätze (Aufnahme, Speichern, Laden)
- Parametervorgabe für die automatische Optimierung des Trainingskorpus und für die Klassifikation von Testmaterial (siehe Kap. 7.7.5)

- Management der TCP/IP-Verbindungen zwischen dem Erkennungssystem und der gestengesteuerten Applikation (IP-Adressen, Ports, Verbindungsauf- und -abbau etc.)

7.6 Systemressourcen

In der nachfolgenden Überlegung soll der bei diesem Verfahren vorliegende Datenstrom diskutiert werden. Die Kopfgestenerkennung verwendet ein Sensorarray aus $n = 4$ Sensoren. Bei einer Abtastfrequenz $f_{abt} = 10$ Hz ergibt sich für eine Geste, die z.B. zwei Sekunden andauert, eine Merkmalsequenz der Länge 2 s·10 Hz = 20. Da sich jeder Merkmalvektor aus $2 \cdot n = 8$ Komponenten zusammensetzt, besteht die Sequenz insgesamt aus 8·20 = 160 Zahlenwerten. Wird zur Speicherung dieser Werte ein 8 bit Datentyp verwendet, so ergibt sich für die betrachtete Geste ein Speicherplatzbedarf von 8 bit·160 = 1280 bit bzw. 0,15625 KByte. Somit erzeugt eine Kopfgeste einen Datenstrom von etwa 0,078 KByte/s (entsprechend 0,244 KByte/s bei der Handgestenerkennung mit $n = 5$ und $f_{abt} = 25$ Hz). Im Vergleich zur videobasierten Gestenerkennung handelt sich hierbei also um ausgesprochen geringe Systemanforderungen: [MOR00] etwa führt in seinen Arbeiten Datenraten (*nach* der Merkmalextraktion) von bis zu 15 KByte/s an.

Zur Überprüfung der Echtzeitfähigkeit wurde der Gestenerkenner auf einer INTEL PENTIUM 1 (133 MHz) Hardwareplattform getestet, deren Performance mit künftig geplanten Fahrzeugbordsystemen vergleichbar ist. Der PC wurde dabei von einer gleichzeitig laufenden Multimedia-Anwendung stark ausgelastet. Dennoch gelang die echtzeitnahe Gestenerkennung mit Antwortzeiten von deutlich weniger als 100 ms ohne jegliche Engpässe.

7.7 Erkennungsergebnisse

7.7.1 Gesten- und Aktionsinventar

Das für die Evaluierung der Erkennungsleistung verwendete Trainings- und Testmaterial besteht aus *Gesten* und *Aktionen*, die für den realen Einsatz im Fahrzeug benötigt werden. Hierbei handelt es sich einerseits um das gesamte Gesteninventar, welches für die Interaktion mit dem prototypischen Infotainmentsystem GECOM definiert wurde (siehe Kap. 4). Darüber hinaus wurden zusätzliche Gestenklassen zur *Aktionserkennung* festgelegt. Bei diesen Aktionen handelt es sich um typische Körperbewegungen, die während einer Autofahrt auftreten, aber nicht fälschlicherweise als beabsichtigte Benutzereingaben interpretiert werden dürfen. So ist z.B. der Blick des Fahrers nicht permanent starr nach vorne auf das Verkehrsgeschehen gerichtet; die zu beobachtenden Blickabwendungen gehen im Allgemeinen einher mit Kopfbewegungen (z.B. „Schulterblick“). Es wurden daher vier Blickbewegungen in das Kopfgesteninventar aufgenommen (siehe Anh. A.5.2), um sicher zu stellen, dass diese Aktionen vom System als solche erkannt werden. Zudem werden die Aktionen *Kopf betritt den Erkennungsbereich* (z.B. Einsteigen ins Fahrzeug) und *Kopf verlässt den Erkennungsbereich* (z.B. Aussteigen) durch eigene Klassen repräsentiert. Entsprechend wurde bei der Handgestenerkennung die häufig auftretende Aktion *Gangschaltung betätigen* als eigene Klasse berücksichtigt.

Die Aktionserkennung unterscheidet sich aus Sicht des Erkennungssystems in keiner Weise von der Nutzgestenerkennung - Aktionen werden wie normale Gestenklassen behandelt, d.h. trainiert und

klassifiziert. In der vorliegenden Arbeit lösen erkannte Aktionen keine Systemreaktionen aus; sie dienen vielmehr als *Garbage-Modell*. Die explizite Nutzung bestimmter Informationen, etwa die der aktuellen Blickrichtung, erscheint jedoch dann sinnvoll, wenn eine entsprechende Instanz zu deren Auswertung bzw. Weiterverarbeitung vorhanden ist. Aus dem visuellen Fokus des Fahrers könnten dann beispielsweise Rückschlüsse auf seine momentanen Absichten oder seine kognitive Belastung gezogen werden. Im Idealfall könnte das Systemverhalten aufgrund dieser Informationen an die jeweilige Situation angepasst werden.

Insgesamt ergeben sich für die Handgestenerkennung zwölf Klassen - elf Gesten und eine Aktion - bzw. für die Kopfgestenerkennung acht Klassen - zwei Gesten und sechs Aktionen (siehe Anh. A.5).

7.7.2 Trainings- und Testmaterial

Um über die reine Erkennungsrate hinaus auch Aussagen über die Benutzerabhängigkeit bzw. Mehrbenutzertauglichkeit des Systems machen zu können, wurden sowohl für die Hand- als auch für die Kopfgestenerkennung jeweils zwei Testszenarien evaluiert: ein Einzelpersonentest und ein Mehrpersonentest. Dabei stammen Test- und jeweils zugehöriges Trainingsmaterial von der/den selben Person/en und sind disjunkt voneinander.

Die Trainingsreihenfolge der einzelnen Gesten war fest vorgegeben, wobei sichergestellt wurde, dass die Ausführungswiederholungen ein und der selben Geste nicht direkt hintereinander erfolgten. Dadurch sollte vermieden werden, dass die Gestenausführung durch häufiges Wiederholen in zunehmend „mechanischer“ bzw. unnatürlicher Weise stattfindet.

Um statistisch repräsentative Erkennungsraten zu erhalten, wurde das jeweilige Testmaterial in entsprechend großem Umfang (bis zu 600 Gesten) gesammelt. Zu jedem Testkorpus wurde ein Trainingskorpus von jeweils halber Datenmenge angelegt. Für die Darstellung des Zusammenhangs zwischen Trainingsumfang und Erkennungsleistung wurden diese Trainingsdaten schrittweise durch zufällige Entnahme von Referenzen bis zur minimalen Größe (eine Referenz pro Klasse) verkleinert. Folgendes Datenmaterial wurde schließlich für die Evaluierung verwendet:

Trainingskorpora

- $\underline{\boldsymbol{Tr}}_{\text{Hand,1P}}$: 1 Person, 12 Klassen à 25 Referenzen
- $\underline{\boldsymbol{Tr}}_{\text{Hand,4P}}$: 4 Personen, 12 Klassen à 20 Referenzen, jeweils 5 Referenzen pro Person
- $\underline{\boldsymbol{Tr}}_{\text{Kopf,1P}}$: 1 Person, 8 Klassen à 25 Referenzen
- $\underline{\boldsymbol{Tr}}_{\text{Kopf,3P}}$: 3 Personen, 8 Klassen à 9 Referenzen, jeweils 3 Referenzen pro Person

Testkorpora

- $\underline{\boldsymbol{Te}}_{\text{Hand,1P}}$: 1 Person, 12 Klassen à 50 Referenzen
- $\underline{\boldsymbol{Te}}_{\text{Hand,4P}}$: 4 Personen, 12 Klassen à 40 Referenzen, jeweils 10 Referenzen pro Person
- $\underline{\boldsymbol{Te}}_{\text{Kopf,1P}}$: 1 Person, 8 Klassen à 50 Referenzen
- $\underline{\boldsymbol{Te}}_{\text{Kopf,3P}}$: 3 Personen, 8 Klassen à 18 Referenzen, jeweils 6 Referenzen pro Person

7.7.3 Klassifikation mit Hidden-Markov-Modellen

Die Implementierung des DTW-Algorithmus als Mustererkennungsverfahren erfolgte in vorliegender Arbeit in erster Linie aufgrund der gegebenen Randbedingungen (*Fahrzeugtauglichkeit,* siehe Kap. 7.1), die ein echtzeitfähiges Verfahren unter Berücksichtigung geringer Hard- und Softwarekosten fordern.

In vielen anderen Einsatzbereichen, z.B. im Bürobereich, müssen derartige Randbedingungen nicht berücksichtigt werden. Die heute verfügbare Computerhardware wird bei gleichzeitig sinkenden Preisen immer leistungsfähiger, was den Einsatz entsprechend rechenintensiver Mustererkennungsverfahren erlaubt. So hat sich hier die Verwendung von *Hidden-Markov-Modellen* (HMMs) als besonders leistungsfähig erwiesen und für ein breites Spektrum an Klassifikationsaufgaben etabliert. Es handelt sich hierbei um ein statistisches Verfahren, welches auf der stochastischen Modellierung von Musterverläufen (z.B. Sprachsignalen) als Markov-Prozesse basiert. Somit sind HMMs in der Lage, Musterverläufe unter Angabe der zugehörigen Emissionswahrscheinlichkeit zu erzeugen. Bei *kontinuierlichen HMMs* werden die zustandsinternen Wahrscheinlichkeitsdichtefunktionen (WDFs) durch Gaußsche Normalverteilungen gebildet. Durch die Überlagerung mehrerer Normalverteilungen (*Mixturen*) können mehrgipflige WDFs modelliert werden (siehe [RUS97]); man spricht dann von sogenannten *Mischverteilungs-Modellen*. Jede vorhandene Klasse wird durch (mindestens) ein HMM repräsentiert. Die Klassifikation erfolgt üblicherweise, indem dasjenige HMM gesucht wird, welches einen unbekannten Musterverlauf mit der größten Wahrscheinlichkeit selbst erzeugen kann. Für detaillierte Beschreibungen zu Funktionsweise, Trainingsmethoden und Klassifikation sei an dieser Stelle auf [RAB89], [RUS94] und [RUS97] verwiesen.

Um aussagekräftige Angaben zur Leistungsfähigkeit des in dieser Arbeit eingesetzten DTW-Verfahrens machen zu können, wurde ein durchgängiger Vergleich zur heute üblichen HMM-Erkennung vollzogen. Dazu wurden kontinuierliche HMMs verwendet, welche mit den vorverarbeiteten Referenzmustern des DTW-Erkennungssystems gespeist wurden. Die Implementierung der HMMs erfolgte unter Verwendung des HTK TOOLKITS (siehe [YOU00]).

7.7.4 Erkennungsraten

Einzelpersonentest Handgestik

Abb. 7.24 zeigt die Erkennungsraten, die sich bei der Evaluierung des Testdatensatzes $\underline{\boldsymbol{Te}}_{\mathrm{Hand,1P}}$ unter Verwendung des Trainingsdatensatzes $\underline{\boldsymbol{Tr}}_{\mathrm{Hand,1P}}$ (bzw. zufällig ausgewählter Teilmengen von $\underline{\boldsymbol{Tr}}_{\mathrm{Hand,1P}}$) ergaben. Die DTW-Erkennungsleistung wird hierbei mit der von HMMs verglichen, welche mit neun[28] Zuständen und jeweils *einer* Gauß-Verteilung modelliert wurden.

[28] Bei dieser Anzahl ergaben sich in vorliegendem Fall maximale Erkennungsraten. Da HTK-intern grundsätzlich zwei zusätzliche Zustände (Start- und Endzustand) angelegt werden, enthalten die HMMs somit lediglich sieben emittierende Zustände; dies ist bei allen hier angegebenen Zustandsanzahlen zu berücksichtigen.

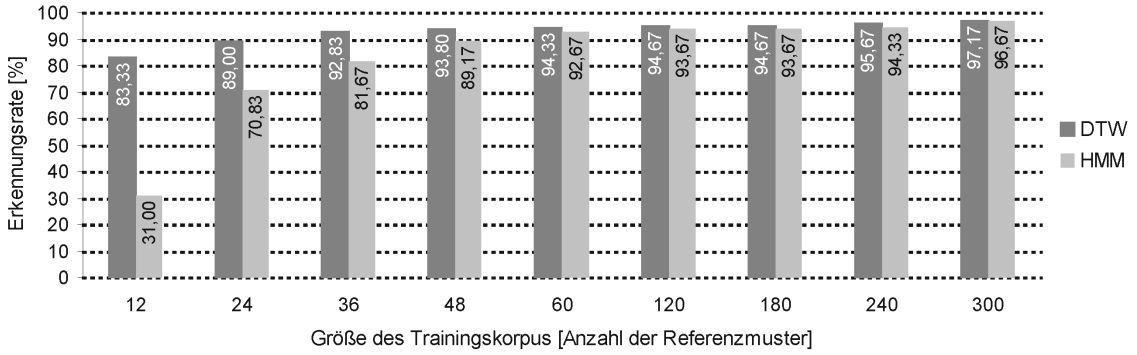


Abb. 7.24: Erkennungsraten des Einzelpersonentests für DTW und HMMs bei 12 Handgestenklassen.

Zunächst ist ersichtlich, dass das DTW-Verfahren bereits bei sehr geringer Trainingskorpusgröße akzeptable Erkennungsraten liefert (83,33 % bei nur *einer* Referenz pro Klasse). Darüber hinaus scheint die Erkennungsleistung mit zunehmendem Trainingsumfang deutlich früher eine Sättigung zu erreichen als bei der HMM-Erkennung. Erwartungsgemäß setzt das HMM-Verfahren einen relativ hohen Trainingsaufwand voraus, da es sich um ein statistisches Verfahren handelt[29].

Dahingegen war bei der DTW-Erkennung anzunehmen, dass die optimale Erkennungsleistung bei weitaus geringerer Anzahl an Trainingsreferenzen pro Klasse erreicht würde als die Evaluierung zeigt. Dieses Resultat legt die Vermutung nahe, dass klasseninterne Musterverläufe teilweise sehr hohe Varianzen aufweisen und somit alle vorkommenden Ausführungsvarianten nur durch einen entsprechend großen Trainingsumfang abgedeckt werden.

Des Weiteren zeigt Abb. 7.24, dass die DTW-Erkennung durchgängig höhere Erkennungsraten liefert als die HMMs, wo sich erst bei der maximalen Trainingskorpusgröße etwa gleichwertige Resultate ergeben. Dieser Umstand erscheint zunächst ungewöhnlich, wird jedoch in Anbetracht der Tatsache, dass es sich um HMMs mit jeweils lediglich *einer* Normalverteilung handelt, relativiert. Die HMM-Erkennungsleistung lässt sich durch die Verwendung mehrerer Gauß-Mixturen deutlich verbessern, wie aus Abb. 7.25 hervorgeht.

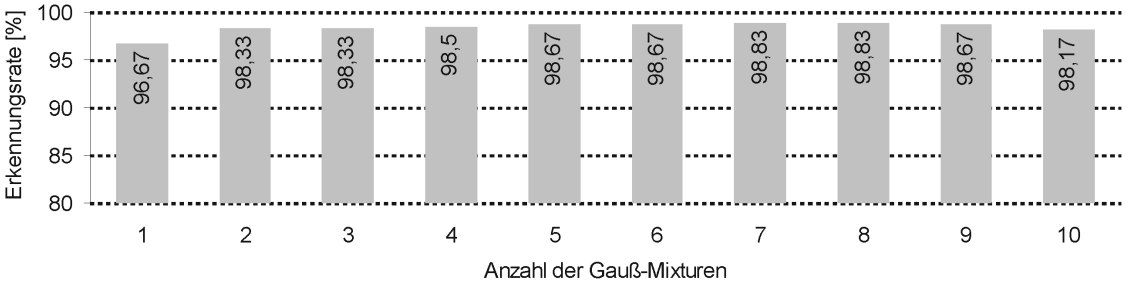


Abb. 7.25: Verbesserung der HMM-Erkennungsleistung durch die Verwendung mehrerer Gauß-Mixturen**.**

[29] Die Angabe der HMM-Erkennungsrate bei z.B. nur 12 Trainingsdatensätzen geschah dementsprechend lediglich aus Gründen der Vollständigkeit.

Auch hier wurde der Testkorpus $\underline{\boldsymbol{Te}}_{\text{Hand,1P}}$ unter Verwendung des vollständigen Trainingsdatensatzes $\underline{\boldsymbol{Tr}}_{\text{Hand,1P}}$ evaluiert (vgl. Abb. 7.24; Trainingskorpusgröße 300), wobei die Anzahl der Mixturen schrittweise erhöht wurde. Durch die Überlagerung von sieben Normalverteilungen steigt die Erkennungsrate auf 98,83 % und übertrifft somit die der DTW-Erkennung. Für die Erstellung der Multimixtur-Modelle muss jedoch ein entsprechend hoher Rechenaufwand betrieben werden. Bei allen folgenden Angaben zu HMM-Erkennungsraten wurde sowohl die Anzahl der Zustände als auch die der Gauß-Mixturen auf maximale Erkennungsleistung optimiert.

Einzelpersonentest Kopfgestik

Bei der Evaluierung des Testdatensatzes $\underline{\boldsymbol{Te}}_{\text{Kopf,1P}}$ unter Verwendung des Trainingsdatensatzes $\underline{\boldsymbol{Tr}}_{\text{Kopf,1P}}$ ergaben sich die in Abb. 7.26 dargestellten Erkennungsraten.

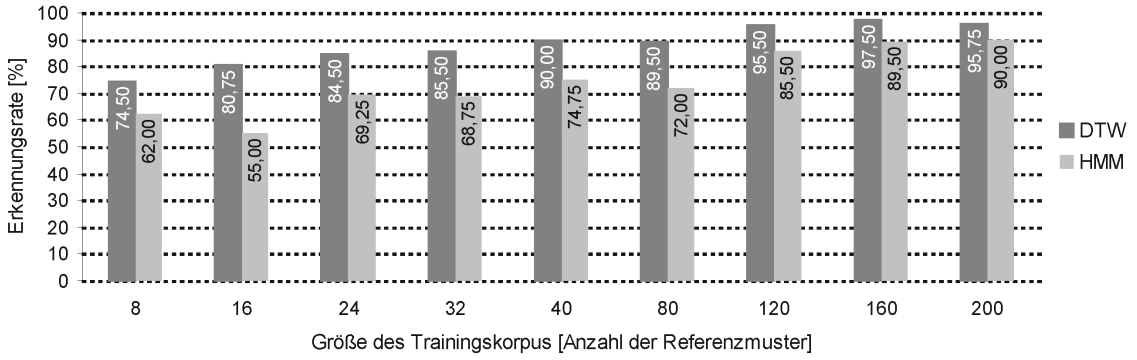


Abb. 7.26: Erkennungsraten des Einzelpersonentests für DTW und HMMs bei 8 Kopfgestenklassen.

Hohe Erkennungsraten setzten hier offensichtlich einen relativ großen Trainingsaufwand voraus. So wird bei der DTW-Erkennung erst mit 15 Trainingsreferenzen pro Klasse (also 120 Referenzen bei acht Klassen) eine akzeptable Erkennungsrate von über 95 % erreicht. Dies lässt sich mit den hohen klasseninternen Varianzen begründen, die folgendermaßen entstehen: Bereits bei geringen Sitzpositionsänderungen des Benutzers können sich für ein und dieselbe Kopfgeste stark unterschiedliche Musterverläufe ergeben. Bei der Aufnahme von Test- und Trainingsmaterial wurde die Sitzposition durch Verlassen und erneutes Betreten des Fahrzeugs nach jedem Aufnahmezyklus absichtlich variiert. Dabei umfasst ein Zyklus die Aufnahme von *einer* Referenz für alle vorhandenen Klassen.

Die HMM-Erkennung lieferte bei einer Parameterwahl von 6 Zuständen mit 10 Mixturen die besten Ergebnisse, welche dennoch nicht an die des DTW-Verfahrens heranreichen. Offenbar erfordern die stark variierenden Musterverläufe innerhalb der Klassen einen noch höheren als den hier betriebenen Trainingsaufwand. Folgt man dem Trend der HMM-Erkennungsrate, ist zu vermuten, dass sich die Erkennungsleistung durch weitere Vergrößerung des Trainingskorpus steigern ließe.

Mehrpersonentest Handgestik

Zur Untersuchung der Mehrbenutzertauglichkeit wurde gemischtes Test- ($\underline{\boldsymbol{Te}}_{\text{Hand,4P}}$) und Trainingsmaterial ($\underline{\boldsymbol{Tr}}_{\text{Hand,4P}}$) zusammengestellt, welches jeweils zu gleichen Anteilen Referenzmuster von vier Personen enthält. Abb. 7.30 zeigt die Erkennungsraten, die sich hierbei ergaben.

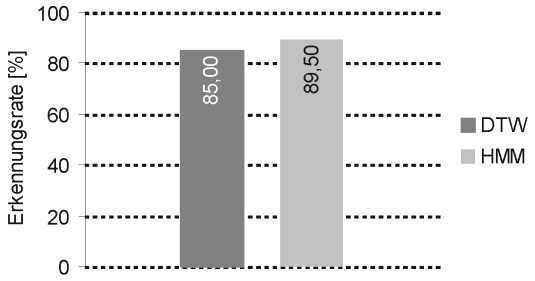


Abb. 7.27: Erkennungsraten des Mehrpersonentests (vier Personen) für DTW und HMMs (sieben Zustände, sechs Mixturen).

Für die DTW-Erkennungsrate ergibt sich mit 85 % ein relativ unbefriedigender Wert im Vergleich zum Einzelpersonentest. Dieses Resultat lässt zunächst darauf schließen, dass interpersonelle Unterschiede in der Gestenausführung zu Überlappungen von Musterverläufen unterschiedlicher Klassenzugehörigkeit im Merkmalraum führen. Im Gegensatz zur statistischen HMM-Erkennung sind derartige Überschneidungen unter Verwendung des DTW-Verfahrens schwer bzw. nicht handhabbar. Zur Überprüfung obiger These werden die DTW-Einzelergebnisse der vier Versuchspersonen (Vpn) näher betrachtet (siehe Abb. 7.28).

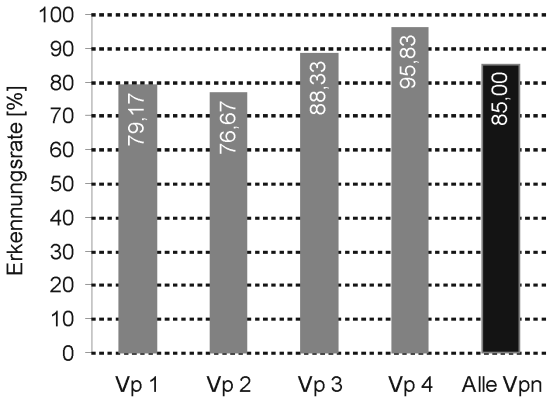


Abb. 7.28: DTW-Erkennungsraten des Mehrpersonentests aufgeschlüsselt nach Versuchspersonen.

Offensichtlich ist die schlechte Gesamterkennungsleistung in erster Linie auf die niedrigen Erkennungsraten der Versuchspersonen Vp 1 und Vp 2 zurückzuführen. Diese Datensätze wurden daraufhin isoliert und sodann getrennt voneinander klassifiziert (gemäß der Vorgehensweise des Einzelpersonentests). Hierbei ergaben sich jedoch ähnlich geringe Erkennungsleistungen, woraus gefolgert werden kann, dass die niedrige Gesamterkennungsrate nicht aus Überlappungen im Merkmalraum resultiert. Vielmehr zeigte sich, dass die Gestenausführungen von Vp 1 und Vp 2 bei der Datenaufnahme sehr stark variierten. Für eine verbesserte Erkennungsleistung müssten die Trainingskorpora von Vp 1 und Vp 2 entsprechend vergrößert werden, so dass alle Ausführungsvarianten abgedeckt werden.

Die HMM-Erkennung spielt im Mehrpersonentest erwartungsgemäß die Stärke eines statistischen Verfahrens aus und übertrifft die DTW-Erkennung um 4,5 %. Auch hier ist davon auszugehen, dass sich die Erkennungsleistung durch umfangreicheres Training weiter verbessern ließe.

Mehrpersonentest Kopfgestik

Das hierbei verwendete Kopfgestenmaterial (***Tr***$_{\text{Kopf,3P}}$ und ***Te***$_{\text{Kopf,3P}}$) stammt von drei Versuchspersonen. Wie bei der Handgestik liegt die DTW-Gesamterkennungsleistung im Mehrpersonentest mit 81,25 % deutlich unter der des Einzelpersonentests, wie aus Abb. 7.29 hervorgeht.

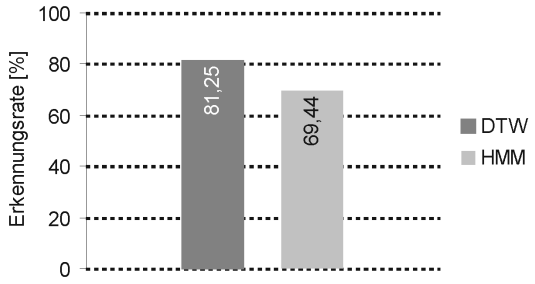


Abb. 7.29: Erkennungsraten des Mehrpersonentests (drei Personen) für DTW und HMMs (fünf Zustände, zwölf Mixturen).

Auch hier soll die Einzelbetrachtung der personenspezifischen Erkennungsraten (siehe Abb. 7.30) näheren Aufschluss über die Ursachen liefern.

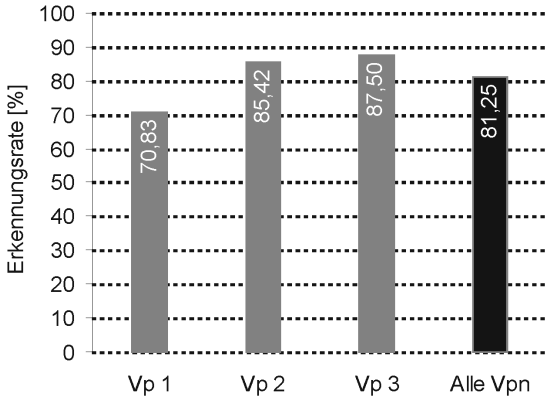


Abb. 7.30: DTW-Erkennungsraten des Mehrpersonentests aufgeschlüsselt nach Versuchspersonen.

Im hier verwendeten Trainingsmaterial ***Tr***$_{\text{Kopf,3P}}$ enthält jede Klasse neun Referenzen, d.h. nur *drei* von jeder Person. Die DTW-Erkennungsraten von Vp 2 und Vp 3 (siehe Abb. 7.30) entsprechen in etwa der des Einzelpersonentests bei *vier* Trainingsreferenzen pro Klasse (vgl. Abb. 7.26). Die Einzelerkennungsraten leiden also offensichtlich nicht unter der Mischung des Datenmaterials verschiedener Benutzer.

Die niedrige Erkennungsrate von Vp 1 lässt sich hier auf zwei Hauptursachen zurückführen:
Einerseits unterscheidet sich die Körpergröße von Vp 1 - und somit auch die Kopfposition im Sensorfeld - stark von den etwa gleichgroßen Teilnehmern Vp 2 und Vp 3. Da keine Anpassung der Kopfstützenhöhe erfolgte, weichen die Musterverläufe (der selben Klassenzugehörigkeit) von Vp 1 stark von denen der anderen beiden Versuchspersonen ab. Es ergaben sich dadurch zwar keine nachteiligen interpersonellen Überschneidungen im Merkmalraum, die Erkennungsrate von Vp 1 kann jedoch nicht aus dem Trainingsmaterial der andern Teilnehmer profitieren (und umgekehrt).

Andererseits wurden bei VP 1 auch in diesem Fall (vgl. Mehrpersonentest Handgestik) besonders starke Variationen in den einzelnen Gestenausführungen beobachtet. Aus diesem Grund erweist sich der Trainingsumfang von drei Referenzen pro Person zumindest für Vp 1 als zu gering, da nicht alle zu erwartenden Ausführungsvarianten repräsentiert werden.

Für das HMM-Verfahren ergaben sich bereits im Einzelpersonentest aufgrund des zu kleinen Trainingsumfangs relativ unbefriedigende Erkennungsraten (vgl. Abb. 7.26). Da die Musterverläufe im Mehrpersonentest wie vorangehend erwähnt nun noch größere klasseninterne Varianzen aufweisen, ergibt sich hier erwartungsgemäß eine noch schlechtere Erkennungsleistung von nur 69,4 %. Ein entsprechend umfangreicheres Training ist also bei der HMM-Erkennung unabdingbar.

Zudem sei hier am Rande noch erwähnt, dass sich die DTW-Gesamterkennungsrate auf 85 % erhöht, wenn die vier implementierten Blickrichtungen (siehe Anh. A.5.2) nur *einer* gemeinsamen Klasse zugeordnet werden. Es zeigte sich nämlich, dass sich Fehlklassifikationen bei Blickgesten häufig lediglich auf die erkannte Blickrichtung, nicht aber auf den Gestentyp an sich beziehen.

Diskussion der Erkennungsergebnisse

Die DTW-Erkennung erweist sich für das vorliegende Einsatzgebiet insgesamt als durchaus taugliches Klassifikationsverfahren. Insbesondere im Einzelpersonentest ergeben sich im Vergleich zum HMM-Verfahren hohe Erkennungsraten bei relativ geringem Trainingsaufwand.

Zudem ist der Trainingskorpus durch Online-Training jederzeit problemlos erweiterbar, wohingegen HMMs mit jeder Änderung der Datenbasis im Allgemeinen neu modelliert werden müssen. Der damit verbundene Rechenaufwand steigt mit der Komplexität der verwendeten Modelle. Wie die Evaluierung zeigt, ist die Verwendung von Mischverteilungs-Modellen in der vorliegenden Klassifikationsaufgabe Voraussetzung für akzeptable Erkennungsraten. Die Berechnung mehrgipfliger WDFs kann jedoch mit der im Fahrzeug verfügbaren Rechnerperformance keinesfalls echtzeitnah erfolgen.

Die Umfänge der verwendeten Trainingskorpora waren für die DTW-Erkennung meist ausreichend, jedoch zu gering für die statistischen HMMs. Nach [RUS97] erweist sich der Einsatz von HMMs nur dann als besonders günstig, wenn viel Trainingsmaterial vorliegt, da andernfalls keine ausreichende Schätzung der WDFs möglich ist. Diese Aussage wurde hier bestätigt. Dabei stellt sich jedoch unmittelbar die Frage, welcher Trainingsaufwand dem späteren Benutzer zugemutet werden kann.

7.7.5 Optimierung des Trainingskorpus

Da die Anzahl der vorhandenen Trainingsreferenzen bei der DTW-Erkennung in direktem Zusammenhang mit dem Rechenaufwand bei der Klassifikation steht, ist es wünschenswert, den Trainingskorpus möglichst klein zu halten. Ein weiterer Grund für dieses Bestreben ist der im Fahrzeug sehr eng bemessene Speicherplatz.

Aufgrund dieser Überlegungen wurde ein Algorithmus zur Optimierung des Trainingskorpus implementiert, dessen Grundidee sich an die Vorgehensweise von [HUN02] anlehnt. Ziel des Verfahrens

ist die sinnvolle Ausdünnung eines großen bzw. statistisch repräsentativen Trainingsdatensatzes, so dass sich hohe Erkennungsraten auf der Basis möglichst weniger Referenzen erzielen lassen.

Dazu werden einerseits redundante Musterverläufe identifiziert und entfernt. Redundanz liegt in diesem Zusammenhang vor, wenn zwei oder mehr Trainingsreferenzen der selben Klassenzugehörigkeit einander sehr ähnlich sind und daher zumindest eines dieser Muster nicht oder nur unwesentlich zur Erhöhung der Erkennungsleistung beiträgt.

Darüber hinaus soll die Klassentrennbarkeit verbessert werden, indem andererseits diejenigen Trainingsreferenzen, welche sich trotz *verschiedener* Klassenzugehörigkeit im Merkmalraum stark überschneiden, eliminiert werden. Hierbei handelt es sich meist um sogenannte „Ausreißer", die daher nicht als klassentypisch betrachtet werden.

Reklassifikation

Um Aussagen über die Qualität des gesammelten Datenmaterials treffen zu können, wird der Trainingskorpus $\underline{\boldsymbol{T}}$ reklassifiziert. Dazu wird jeweils *ein* Referenzmuster entnommen und als unbekanntes Testmuster $\underline{\boldsymbol{M}}_?$ betrachtet. Nun wird dieses Testmuster unter Verwendung des verbleibenden Datenmaterials klassifiziert. Dabei werden Erkennungsergebnis, Konfidenzmaße (siehe Kap. 7.4.5) und die DTW-Abstände zu allen übrigen Trainingsreferenzen gespeichert. Diese Prozedur erfolgt nacheinander für alle im Trainingskorpus enthaltenen Referenzmuster. Die Ergebnisse - insbesondere die Anzahl der beobachteten Fehlerkennungen - geben Aufschluss über die Güte des gesammelten Gestenmaterials.

Optimierungsalgorithmus

Das implementierte Optimierungsverfahren besteht in einem wiederholten Durchlauf der Reklassifikation, wobei der Trainingskorpus bis zur Erfüllung zuvor festgelegter Kriterien ausgedünnt wird. Hierbei handelt es sich insbesondere um die Forderung, dass in jeder Klasse eine Mindestzahl (≥2) an Referenzen verbleiben muss.

Die Eliminierung von Referenzen geschieht dabei wie folgt: Bei der Klassifikation eines dem Trainingskorpus entnommenen Testmusters $\underline{\boldsymbol{M}}_?$ ergäbe sich als ähnlichstes Referenzmuster $\underline{\boldsymbol{R}}_1$. Nun werden zwei Fälle unterschieden:

- Korrekte Klassifikation

Unter der Voraussetzung, dass es sich um eine korrekte Klassifikation handelt, $\underline{\boldsymbol{R}}_1$ also derselben Klasse angehört wie $\underline{\boldsymbol{M}}_?$, wird sodann $\underline{\boldsymbol{R}}_2$ ermittelt. Dies ist die zweitähnlichste Referenz mit derselben Klassenzugehörigkeit wie $\underline{\boldsymbol{R}}_1$.

Zur Identifizierung redundanter Trainingsreferenzen wird das Unterschiedsmaß u eingeführt:

$$u = \frac{D_{DTW}(\underline{\boldsymbol{M}}_?, \underline{\boldsymbol{R}}_2)}{D_{DTW}(\underline{\boldsymbol{M}}_?, \underline{\boldsymbol{R}}_1)} - 1 \quad \text{mit} \quad D_{DTW}(\underline{\boldsymbol{M}}_?, \underline{\boldsymbol{R}}_1) \le D_{DTW}(\underline{\boldsymbol{M}}_?, \underline{\boldsymbol{R}}_2) \tag{7.51}$$

Für u ergeben sich um so kleinere Werte, je geringer der Unterschied zwischen den beiden DTW-Abständen (siehe Gl. 7.51) ist. Im Extremfall nimmt u den Wert Null an, nämlich dann, wenn die

Abstände identisch sind. Aus sehr kleinen Werten für u kann indirekt gefolgert werden, dass die Musterverläufe $\underline{\boldsymbol{R}}_1$ und $\underline{\boldsymbol{R}}_2$ einander stark ähneln. Folglich besteht für die korrekte Klassifikation des Testmusters $\underline{\boldsymbol{M}}_?$ keine Notwendigkeit, *beide* Referenzen $\underline{\boldsymbol{R}}_1$ und $\underline{\boldsymbol{R}}_2$ im Trainingskorpus zu belassen. Eines der beiden Muster kann also entfernt werden, ohne die Gesamterkennungsleistung stark zu beeinträchtigen. Bei dem verwendeten Algorithmus wird eine Unterschiedsschwelle u_{thr} festgelegt, bei deren Unterschreitung ($u<u_{thr}$) das Muster $\underline{\boldsymbol{R}}_2$ zur Eliminierung in einer Löschliste vorgemerkt wird.

- Falsche Klassifikation

Im Falle einer Fehlklassifikation (das Referenzmuster $\underline{\boldsymbol{R}}_1$ gehört einer anderen Klasse an als das Testmuster $\underline{\boldsymbol{M}}_?$) liegt zunächst die Vermutung nahe, dass es sich bei $\underline{\boldsymbol{M}}_?$ um die einzige bzw. letzte vorhandene Referenz der entsprechenden Klasse handelt. Dies ist jedoch auszuschließen, da zuvor eine Mindestanzahl verbleibender Referenzen pro Klasse festgelegt wurde (siehe oben).

Es ist viel eher davon auszugehen, dass es sich entweder bei $\underline{\boldsymbol{R}}_1$ oder bei $\underline{\boldsymbol{M}}_?$ um einen qualitativ schlechten Repräsentanten der jeweiligen Klasse handelt. Die Voraussetzung eines genügend großen Trainingskorpus rechtfertigt in diesem Fall die Maßnahme, sowohl $\underline{\boldsymbol{R}}_1$ als auch $\underline{\boldsymbol{M}}_?$ für die Löschung vorzumerken.

Nach der Reklassifikation des Testmusters $\underline{\boldsymbol{M}}_?$ werden alle in der Löschliste befindlichen Referenzen unter Einhaltung folgender Randbedingungen eliminiert:

- Pro Iteration darf in jeder Klasse höchstens *eine* Referenz gelöscht werden, um unerwünschte Kaskadierungseffekte zu vermeiden.
- Eine Löschung erfolgt nur, wenn sie nicht zur Unterschreitung der vorgegebenen Mindestanzahl an Referenzen pro Klasse führt.

Zu Beginn der Ausdünnung wird die Unterschiedsschwelle u_{thr} auf den Wert Null gesetzt. Das beschriebene Optimierungsverfahren wird nun solange wiederholt, bis keine Referenzen mehr zur Löschung vorgemerkt werden. Daraufhin wird die Ähnlichkeitsschwelle u_{thr} geringfügig erhöht und die Prozedur beginnt erneut.

Diese Iteration erfolgt so lange, bis entweder ein zuvor festgelegter Maximalwert $u_{thr,max}$ für die Ähnlichkeitsschwelle erreicht wird, oder jede Klasse nur noch die vorgegebene Minimalanzahl an Referenzen enthält.

Der resultierende Trainingskorpus enthält nun einerseits gut trennbares Datenmaterial, welches fehlerfrei reklassifiziert wird. Andererseits werden die vorhandenen Klassen durch weitgehend redundanzfreie Referenzen repräsentiert, d.h. die jeweils zugehörigen Musterverläufe sind ausgewogen im Merkmalraum verteilt.

Abb. 7.31 zeigt exemplarisch die jeweilige Entwicklung von Trainingskorpus-Größe und daraus resultierender Erkennungsleistung für Hand- und Kopfgestik im Laufe des Ausdünnungsprozesses.

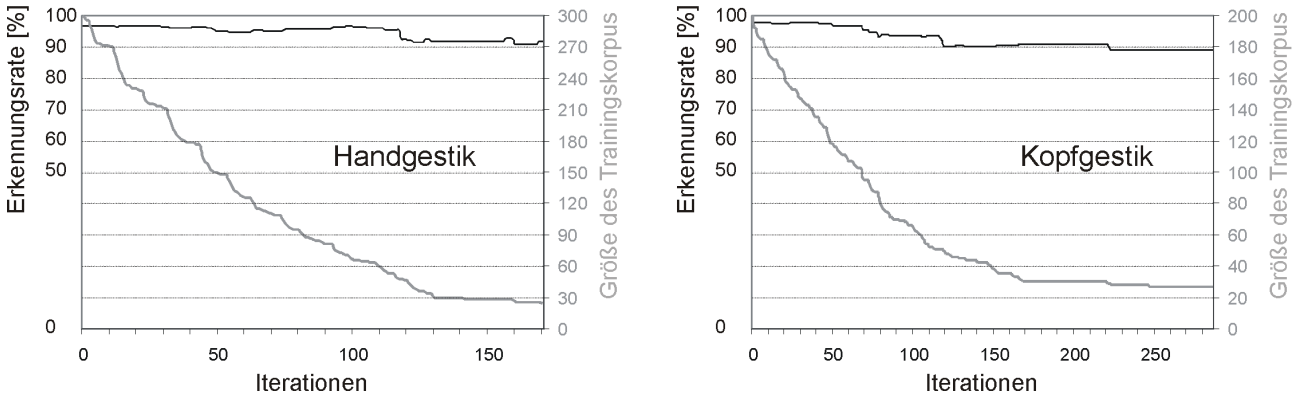


Abb. 7.31: Optimierung des Trainingskorpus; Erkennungsrate und Trainingskorpus-Größe für Handgestik (links) und Kopfgestik (rechts).

Als Ausgangsbasis dienten jeweils die Trainingskorpora der Einzelpersonentests ($\underline{\boldsymbol{Tr}}_{\text{Hand,1P}}$ und $\underline{\boldsymbol{Tr}}_{\text{Kopf,1P}}$; siehe Kap. 7.7.2). Nach jeder Iteration des Optimierungsalgorithmus wurden die aktuellen Erkennungsleistungen unter Verwendung der Testdaten $\underline{\boldsymbol{Te}}_{\text{Hand,1P}}$ bzw. $\underline{\boldsymbol{Te}}_{\text{Kopf,1P}}$ ermittelt.

Wie aus Abb. 7.31 hervorgeht, kann die Trainingskorpus-Größe in beiden Fällen bei nahezu gleichbleibenden Erkennungsraten stark reduziert werden. Bei der Handgestenerkennung kommt es erst dann zu einem deutlichen Einbruch in der Erkennungsleistung, wenn das Trainingsmaterial weniger als 50 Referenzen enthält (etwa bei 115 Iterationen). Offensichtlich wurde hier aufgrund einer zu hohen Unterschiedsschwelle u_{thr} bereits mit der Löschung nichtredundanter Referenzen begonnen. Es erscheint also ratsam, die Ausdünnung des Trainingskorpus kurz vor dem Erreichen dieses Einbruchs zu beenden. Immerhin wird das 300 Referenzen umfassende Ausgangsmaterial dadurch auf 16 % (48 Datensätze) reduziert, wobei die anfängliche Erkennungsrate von 96,8 % nur auf 95,5 % abfällt.

Auch bei der Kopfgestenerkennung lässt sich aufgrund entsprechender Überlegungen ein sinnvoller Kompromiss aus Trainingsumfang und dabei erzielbarer Erkennungsrate finden. Hierbei zeigt sich jedoch, dass die Ausdünnung nur in kleinerem Umfang erfolgen kann als bei der Handgestik. Wie bereits erläutert (siehe *Mehrpersonentest Kopfgestik*), weisen die Musterverläufe bei der Kopfgestik relativ hohe klasseninterne Varianzen auf, woraus die Notwendigkeit einer entsprechend großen Anzahl an Repräsentanten pro Klasse resultiert. So ergibt sich bei einer Ausdünnung des Trainingskorpus um 50 % (auf 99 Referenzen) eine Erkennungsrate von 96,75 % (vgl. Abb. 7.31 bei etwa 70 Iterationen). Dies ist im Vergleich zur Erkennungsrate von 97,8 % bei 200 Datensätzen ein hinnehmbarer Verlust. Bei weiterer Ausdünnung ergibt sich bei der Trainingskorpus-Größe von 52 Referenzen eine Erkennungsrate von immerhin noch 93,5 % (vgl. Abb. 7.31 bei 110 Iterationen).

Insgesamt erweist sich die angewandte Methodik zur Optimierung des Trainingskorpus somit als sinnvolle Maßnahme zur Reduzierung von Rechenaufwand und Speicherplatzbedarf.

8

Résumé

Vorgestellt wurde ein Gesamtsystem zur berührungslosen Bedienung eines Infotainmentsystems im Kraftfahrzeug, wobei das Augenmerk in erster Linie auf die Bereitstellung der in dieser Domäne neuen Eingabemodalität *Gestik* gerichtet war. Die Motivation bestand darin, den Autofahrer bei der Bedienung der mittlerweile sehr großen Funktionsumfänge zu entlasten, indem die heute überwiegend haptische Steuerung über Drehknöpfe und Tasten um alternative Informationskanäle erweitert wird. Dies geschieht mit der Absicht, den Informationsaustausch zwischen Mensch und Maschine „an die sensorischen, motorischen und kognitiven Fähigkeiten und Eigenschaften des Menschen anzupassen“ [LAN02].

Um dieser Zielsetzung gerecht zu werden, wurde allen hier durchgeführten Realisierungen der *Usability-Aspekt* vorangestellt. Diese Vorgehensweise wurde insofern als äußerst wichtig erachtet, zumal ad hoc schwer abgeschätzt werden konnte, ob durch die Nutzung der Gestik im Fahrzeug prinzipiell eine Verbesserung der heutigen Situation herbeigeführt werden kann.

Nachdem grundlegende Voruntersuchungen [ZOB01] vielversprechende Ergebnisse hinsichtlich Intuitivität sowie Benutzerakzeptanz erbrachten, wurden die speziell im Fahrzeugeinsatz sehr wichtigen Aspekte der kognitiven Belastung detailliert untersucht. Hierbei zeigte sich die Gestik aufgrund signifikant geringerer Ablenkungseffekte der haptischen Bedienung überlegen.

Als Konsequenz wurde ein gestisch bedienbares Gesamtsystem umgesetzt, wobei auch hier die Anforderungen - sowohl an die Bedienoberfläche als auch an den Gestenerkenner - Berücksichtigung fanden, die sich im Rahmen schritthaltender Benutzerstudien ergaben. Es resultierte ein lauffähiger Prototyp, der die Möglichkeit einer rein gestischen Bedienung großer Funktionsumfänge im Fahrzeug unter Beweis stellt. Es sei an dieser Stelle abermals darauf hingewiesen, dass das Anliegen dieser Arbeit nicht darin bestand, den Einsatz von ausschließlich gestischer Interaktion im Fahrzeug zu propagieren. Vielmehr sollte sie das hohe Potenzial dieser Eingabemodalität aufzeigen, deren offensichtlichen Vorteile in zukünftigen *multimodalen* Systemen gewinnbringend eingesetzt werden können. Ein erster Schritt in diese Richtung wurde hier bereits mit der Anbindung einer sprachverstehenden Komponente getan. Im Sinne einer komfortablen, benutzerfreundlichen Kommunikation

zwischen Mensch und Maschine erscheint die Kombination von Sprachbedienung mit anderen Interaktionsmitteln [LAN01] - wie etwa der hier betrachteten Gestik - als durchaus anzustrebende Zielsetzung. Wie die vorliegende Arbeit zeigt, kommt es bei der Optimierung einer Bedienumgebung auf Gestik einerseits, sowie auf Spracheingaben andererseits, keineswegs zu Realisierungskonflikten, da die grundlegenden Gestaltungskriterien weitestgehend disjunkt sind und (eben aus diesem Grunde) gemeinsam realisiert werden können.

Neben dem Konzept für eine ergonomische Bedienumgebung wurde in dieser Arbeit ein pragmatischer Ansatz für die Realisierung eines Gestenerkennungssystems vorgestellt. Es galt hierbei die Prämisse, grundsätzliche fahrzeugspezifische Rahmenbedingungen mit einer robust funktionierenden Technologie in Einklang zu bringen. Mit der sensorbasierten Gestenerkennung gelang die Realisierung eines ressourcenschonenden sowie echtzeitfähigen Systems. Das implementierte Mustererkennungsverfahren *Dynamic-Time-Warping* erwies sich selbst für die Klassifikation personenübergreifender Datensätze als durchaus leistungsfähig. Diese Multiuser-Fähigkeit beschränkt sich jedoch auf *kleine* Benutzergruppen und setzt voraus, dass sich die interpersonellen Musterverläufe unterschiedlicher Gesten im Merkmalraum nicht zu sehr überlappen. Um eine hohe Klassentrennbarkeit auch für eine große Anwenderzahl zu gewährleisten wäre es jedoch erforderlich, die Trainingsdatensätze voneinander zu isolieren und jeweils benutzerspezifisch für die Erkennung einzusetzen. Diese Maßnahme ist unter geringem Aufwand realisierbar, da einerseits wenig Trainingsmaterial pro Person benötigt wird und dieses andererseits einen sehr geringen Speicherplatzbedarf aufweist. Die Auswahl der jeweiligen Trainingsdatensätze könnte hierbei automatisch erfolgen: Derzeit vorhandene Bestrebungen, die im Fahrzeug angebotenen Funktionalitäten zu personalisieren, beinhalten auch Maßnahmen zur automatischen Benutzererkennung - als Beispiel sei hier die kommerziell verfügbare Fahreridentifizierung per Fingerabdruck der AUDI AG genannt.

Die im Rahmen dieser Arbeit entwickelte Merkmalgewinnung unter Verwendung von IR-Sensoren, erbrachte überzeugende Resultate. Es ist jedoch auch anzumerken, dass die erreichbare „Bewegungsauflösung“ aufgrund der - bewusst so gewählten - geringen Sensoranzahl nicht an die von videobasierten Verfahren heranreicht. Dies hat die Konsequenz, dass geringe Unterschiede in Bewegungsabläufen nicht erfasst werden können, woraus wiederum folgt, dass sich die implementierten Gesten deutlich voneinander unterscheiden müssen, um klassifiziert werden zu können. Eine Verbesserung der räumlichen Auflösung setzt also die Verwendung einer entsprechend höheren Anzahl an Einzelsensoren voraus. Es wäre in diesem Zusammenhang wünschenswert, auf integrierte IR-Sensor-Arrays zurückgreifen zu können, mit dem gleichzeitigen Vorteil einer weiteren Miniaturisierung.

Abschließend betrachtet liefert das vorgestellte System einen grundlegenden Beitrag in Richtung einer benutzerfreundlicheren Mensch-Maschine-Interaktion im Fahrzeug, wirft jedoch auch neue Fragestellungen und Probleme auf, die zum jetzigen Zeitpunkt kaum erschöpfend bewertet werden können. Eine mögliche Antwort auf dieses Dilemma liefert folgendes Zitat:

„Auf viele der durch Forschung und Technik erzeugten Probleme gibt es keine andere Antwort als durch neue Forschung und bessere Technik.“
HUBERT MARKL (ehemaliger Präsident der deutschen Max-Planck-Gesellschaft).

A

Anhang

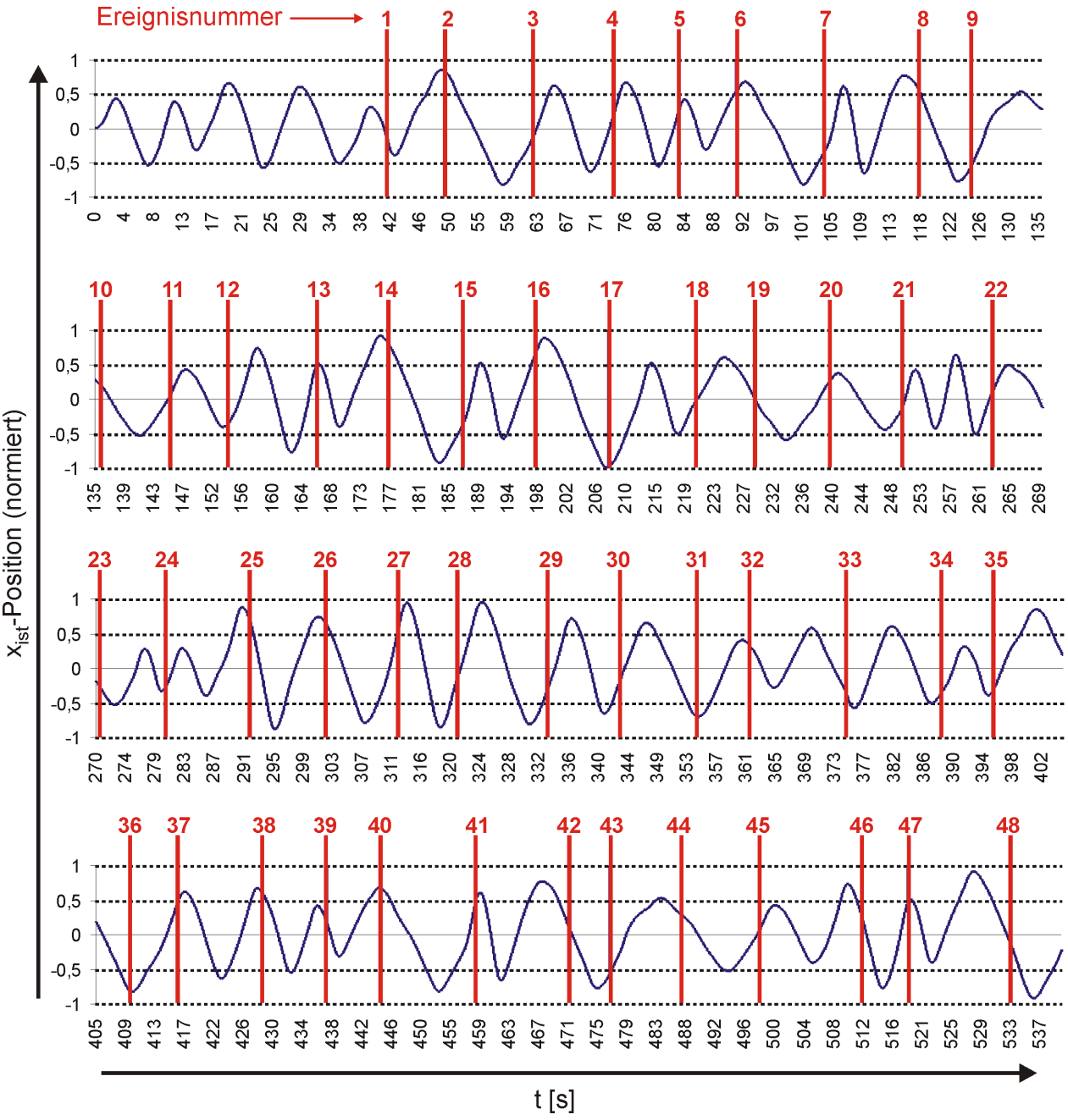
A.1 Untersuchung von Ablenkungseffekten

A.1.1 Ereignisse

Ereignis-Nr.	Bedienung	Modalität	Objekt
1	-	-	5
2	Telefon	Haptik	-
3	Erhöhen	Gestik	2
4	Telefon	Gestik	-
5	-	-	2
6	Auswahl	Haptik	5
7	Telefon	Haptik	2
8	-	-	5
9	Auswahl	Gestik	5
10	Erniedrigen	Haptik	2
11	Telefon	Gestik	2
12	-	-	5
13	Erhöhen	Haptik	5
14	-	-	5
15	Auswahl	Gestik	2
16	Erhöhen	Gestik	-
17	-	-	2
18	Auswahl	Haptik	-
19	Telefon	Haptik	2
20	-	-	2
21	Auswahl	Haptik	-
22	Erniedrigen	Gestik	5
23	Telefon	Gestik	5
24	-	-	5

Ereignis-Nr.	Bedienung	Modalität	Objekt
25	-	-	2
26	Erhöhen	Gestik	-
27	-	-	2
28	Erniedrigen	Haptik	-
29	Telefon	Haptik	-
30	Erniedrigen	Gestik	-
31	Telefon	Gestik	-
32	-	-	2
33	Erhöhen	Haptik	2
34	Erhöhen	Gestik	2
35	-	-	2
36	Erniedrigen	Haptik	5
37	-	-	5
38	Auswahl	Haptik	2
39	Erniedrigen	Gestik	-
40	Auswahl	Gestik	-
41	-	-	5
42	Erhöhen	Haptik	-
43	-	-	5
44	Erniedrigen	Gestik	5
45	Erniedrigen	Haptik	-
46	Auswahl	Gestik	-
47	-	-	5
48	Erhöhen	Haptik	-

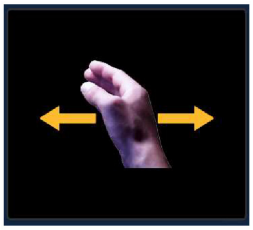
A.1.2 Verlauf der Soll-Marke mit Ereignissen



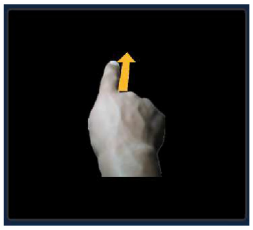
A.2 Audiovisuelles Hilfesystem in GeCoM

Bei einer Hilfeanforderung wird je nach Kontext eines der in A.2.1 dargestellten Bilder für acht Sekunden im Display (*Anzeigebereich 2*) angezeigt, während gleichzeitig der zugehörige Hilfetext gemäß A.2.2 via Sprachausgabe vorgelesen wird.

A.2.1 Visuelle Hilfe

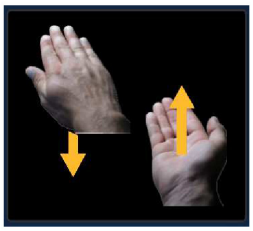


Hilfetyp 1



Hilfetyp 2

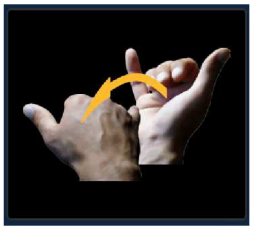
Hilfetyp 3



Hilfetyp 4

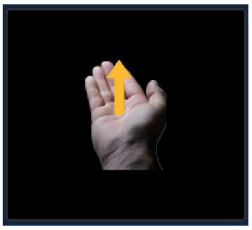
Hilfetyp 5

Hilfetyp 6

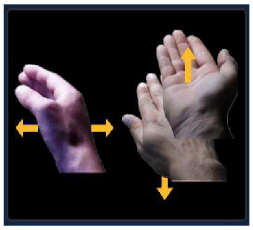


Hilfetyp 7

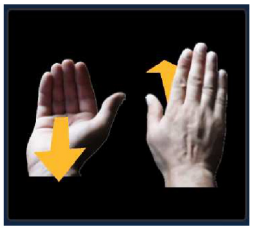
Hilfetyp 8



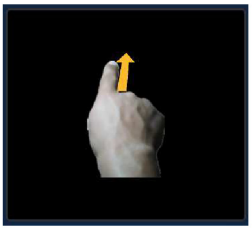
Hilfetyp 9



Hilfetyp 10



Hilfetyp 11



Hilfetyp 12

A.2.2 Auditive Hilfe

Kontext	Hilfetext	Hilfetyp
Geräteauswahl	„Um das Gerät zu wechseln, winken Sie nach rechts oder links.“	1
	„Um ein Gerät auszuwählen, zeigen Sie in Richtung des Displays.“	2
	„Sie können jederzeit zum Telefon wechseln, indem Sie das Abheben eines Telefonhörers imitieren.“	15
Radio	„Um zur Geräteauswahl zu wechseln, ziehen Sie an dem virtuellen Griff.“	3
	„Um den Sender zu wechseln, winken Sie nach rechts oder links.“	1
	„Um die Lautstärke zu ändern, winken Sie nach oben oder unten.“	4
	„Um den Ton stumm zu schalten, machen Sie eine waagerechte Wischbewegung.“	5
	„Um den Ton wieder anzuschalten, winken Sie nach oben.“	9
	„Um die Lautstärke stufenlos zu regeln, greifen Sie nach der virtuellen Kugel und bewegen diese nach oben oder unten.“	18
	„Sie können jederzeit zum Telefon wechseln, indem Sie das Abheben eines Telefonhörers imitieren.“	15
CD	„Um zur Geräteauswahl zu wechseln, ziehen Sie an dem virtuellen Griff.“	3
	„Um den Titel zu wechseln, winken Sie nach rechts oder links.“	1
	„Um die Lautstärke zu ändern, winken Sie nach oben oder unten.“	4
	„Um den Ton stumm zu schalten, machen Sie eine waagerechte Wischbewegung.“	5
	„Um den Ton wieder anzuschalten, winken Sie nach oben.“	9
	„Um die Lautstärke stufenlos zu regeln, greifen Sie nach der virtuellen Kugel und bewegen diese nach oben oder unten.“	18
	„Sie können jederzeit zum Telefon wechseln, indem Sie das Abheben eines Telefonhörers imitieren.“	15
CD-Wechsler	„Um zur Geräteauswahl zu wechseln, ziehen Sie an dem virtuellen Griff.“	3
	„Um die CD zu wechseln, winken Sie nach rechts oder links.“	1
	„Um eine CD auszuwählen, zeigen Sie in Richtung des Displays.“	2
	„Um die Lautstärke zu ändern, winken Sie nach oben oder unten.“	4
	„Um den Ton stumm zu schalten, machen Sie eine waagerechte Wischbewegung.“	5
	„Um den Ton wieder anzuschalten, winken Sie nach oben.“	9
	„Um die Lautstärke stufenlos zu regeln, greifen Sie nach der virtuellen Kugel und bewegen diese nach oben oder unten.“	18
	„Sie können jederzeit zum Telefon wechseln, indem Sie das Abheben eines Telefonhörers imitieren.“	15
Kassette	„Um zur Geräteauswahl zu wechseln, ziehen Sie an dem virtuellen Griff.“	3
	„Um den Titel zu wechseln, winken Sie nach rechts oder links.“	1
	„Um die Lautstärke zu ändern, winken Sie nach oben oder unten.“	4
	„Um den Ton stumm zu schalten, machen Sie eine waagerechte Wischbewegung.“	5
	„Um den Ton wieder anzuschalten, winken Sie nach oben.“	9
	„Um die Lautstärke stufenlos zu regeln, greifen Sie nach der virtuellen Kugel und bewegen diese nach oben oder unten.“	18
	„Sie können jederzeit zum Telefon wechseln, indem Sie das Abheben eines Telefonhörers imitieren.“	15
Telefon	„Um zur Geräteauswahl zu wechseln, ziehen Sie an dem virtuellen Griff.“	3
	„Um einen anderen Eintrag anzuzeigen, winken Sie nach rechts oder links.“	1
	„Um einen Eintrag auszuwählen, zeigen Sie in Richtung des Displays oder imitieren Sie das Abheben eines Telefonhörers.“	6
	„Um einen Anfangsbuchstaben zu wählen, greifen Sie nach der virtuellen Kugel und bewegen Sie diese nach rechts oder links.“	19

Kontext	Hilfetext	Hilfetyp
Aufbau eines Telefongesprächs	„Um den Verbindungsaufbau abzubrechen, imitieren Sie das Auflegen eines Telefonhörers oder machen Sie eine waagerechte Wischbewegung.“	13
Eingehender Anruf	„Um das Gespräch anzunehmen, imitieren Sie das Abheben eines Telefonhörers.“	6
	„Um das Gespräch abzulehnen, machen Sie eine waagerechte Wischbewegung.“	13
	„Um das Gespräch zu halten, heben Sie die Hand.“	8
	„Um die Funktion zu wechseln, winken Sie nach rechts oder links.“	1
	„Um eine Funktion auszuwählen, zeigen Sie in Richtung des Displays.“	2
	„Um das Gespräch anzunehmen oder abzulehnen, nicken Sie oder schütteln Sie den Kopf.“	17
Laufendes Telefongespräch	„Um das Gespräch zu beenden, imitieren Sie das Auflegen eines Telefonhörers.“	7
	„Um das Gespräch zu halten, heben Sie die Hand.“	8
	„Um die Funktion zu wechseln, winken Sie nach rechts oder links.“	1
	„Um eine Funktion auszuwählen, zeigen Sie in Richtung des Displays.“	2
Gehaltenes Telefongespräch	„Um das Gespräch fortzusetzen, imitieren Sie das Abheben eines Telefonhörers.“	6
	„Um das Gespräch zu beenden, imitieren Sie das Auflegen eines Telefonhörers.“	7
	„Um die Funktion zu wechseln, winken Sie nach rechts oder links.“	1
	„Um eine Funktion auszuwählen, zeigen Sie in Richtung des Displays.“	2
Navigation	„Um zur Geräteauswahl zu wechseln, ziehen Sie an dem virtuellen Griff.“	3
	„Um die Funktion zu wechseln, winken Sie nach rechts oder links.“	1
	„Um eine Funktion auszuwählen, zeigen Sie in Richtung des Displays.“	2
	„Sie können jederzeit zum Telefon wechseln, indem Sie das Abheben eines Telefonhörers imitieren.“	15
Zieleingabe	„Winken Sie in die Richtung, in die Sie den Sucher bewegen wollen.“	10
	„Um die Karte zu vergrößern, ziehen Sie sie zu sich her. Um die Karte zu verkleinern, schieben Sie sie von sich weg.“	11
	„Um das Navigationsziel auszuwählen, zeigen Sie in Richtung des Displays.“	12
	„Um die Zieleingabe abzubrechen, machen Sie eine waagerechte Wischbewegung.“	14
	„Greifen Sie den Sucher, um ihn stufenlos zu verschieben.“	20
	„Sie können jederzeit zum Telefon wechseln, indem Sie das Abheben eines Telefonhörers imitieren.“	15
Zielspeicher	„Um zur Geräteauswahl zu wechseln, ziehen Sie an dem virtuellen Griff.“	3
	„Um einen anderen Eintrag anzuzeigen, winken Sie nach rechts oder links.“	1
	„Um einen Eintrag auszuwählen, zeigen Sie in Richtung des Displays.“	2
	„Sie können jederzeit zum Telefon wechseln, indem Sie das Abheben eines Telefonhörers imitieren.“	15
Systemrückfrage	„Um zur Geräteauswahl zu wechseln, ziehen Sie an dem virtuellen Griff.“	3
	„Um zu einer anderen Antwort zu wechseln, winken Sie nach rechts oder links.“	1
	„Um die Antwort auszuwählen, zeigen Sie in Richtung des Displays.“	2
	„Um abzubrechen machen Sie eine waagerechte Wischbewegung.“	14
	„Um die Frage zu bejahen oder zu verneinen, nicken Sie oder schütteln Sie den Kopf.“	16
	„Sie können jederzeit zum Telefon wechseln, indem Sie das Abheben eines Telefonhörers imitieren.“	15

A.3 Usability-Untersuchung: Versuchsablauf

Einleitungstext

„Herzlich willkommen am Lehrstuhl für Mensch-Maschine-Kommunikation! Wir untersuchen in diesem Versuch ein neuartiges Bedienkonzept zur Steuerung von Geräten im Fahrzeug. Dabei handelt es sich um Geräte der Informations-, Kommunikations- und Unterhaltungselektronik, wie Radio, CD-Spieler, Telefon und Navigationssystem. Während üblicherweise zur Bedienung solcher Funktionen Tasten und Drehknöpfe verwendet werden, wird das vorliegende System mit Gesten gesteuert. Die Gesten können dazu mit der rechten Hand oder in einigen Fällen mit dem Kopf ausgeführt werden. In diesem Versuch soll geklärt werden, ob das Bedienkonzept intuitiv und benutzerfreundlich handhabbar ist. Es werden Ihnen im weiteren Verlauf einige Aufgaben gestellt, bei denen Sie alle Bedienabläufe mit Gesten bedienen sollen. Die Dauer der Versuche wird insgesamt ca. 60 bis 90 Minuten betragen. Im Voraus vielen Dank für Ihre Teilnahme als Versuchsperson!“

→ *Eingangsfragebogen (Versuchsperson)*

Teil 1: Selbstständige Exploration

„Bitte beschäftigen Sie sich in den nächsten Minuten selbständig mit dem Gerät um erste Eindrücke zu gewinnen. Versuchen Sie dabei die implementierten Funktionalitäten und deren Bedienung kennen zu lernen. Vergessen Sie nicht, dass sich das System ausschließlich mit Gesten steuern lässt. Versuchen Sie also, mit Gesten deutlich zu machen, was Sie erreichen wollen. Stellen Sie sich vor, dass Ihnen gegenüber eine Person sitzt mit der Sie nicht sprechen, sondern nur über Gesten kommunizieren können. Diese Person steuert dann für Sie die erwünschte Funktion. An der rechten Seite des Lenkrads befindet sich ein Bügel. Bitte halten Sie diesen permanent gedrückt und lassen Sie ihn nur zur Ausführung einer einzelnen Handgeste los. Dieser Bügel ist nur für die leichtere Versuchsauswertung nötig und wird natürlich im endgültigen System weggelassen. Bitte berücksichtigen Sie dies bei der Beurteilung des Systems. Das gleiche gilt natürlich für den Sensor am Handgelenk.

→ *Fragebogen Teil 1 (Versuchsperson)*

→ *Einschätzungsbogen Teil 1 (Versuchsleiter)*

Teil 2: Bedienung im stehenden Fahrzeug

„Im folgenden werden Ihnen einige Aufgaben gestellt, bei denen Sie bestimmte Bedienabläufe ausführen sollen. Alle auszuführenden Aufgaben werden von einer Frauenstimme vorgelesen. Bitte führen Sie ausschließlich diese Aufgaben aus - dies erleichtert uns die nachfolgende Auswertung. Falls Sie bei der Bedienung Schwierigkeiten haben oder einfach nur zusätzliche Informationen wünschen, so drücken Sie bitte die mit einem Fragezeichen gekennzeichnete Taste links vom Schaltknüppel. Bitte halten Sie den Bügel am Lenkrad weiterhin permanent gedrückt und lassen Sie ihn nur zur Ausführung einer einzelnen Handgeste oder zum Betätigen der Hilfetaste los. Bitte halten Sie den Bügel auch dann kurz mit der Hand fest, wenn Sie zwei oder mehrere, direkt aufeinanderfolgende Gesten ausführen wollen. Sind Sie bereit?

Teil 2, Block 1

1. Wählen Sie das Gerät Radio aus.

2. Wechseln Sie zum Sender Bayern 5.

3. Der Bericht interessiert Sie, daher machen Sie den Sender lauter.

4. Sie wollen einen anderen Sender hören. Wechseln Sie zu Radio Gong.

5. Die Musik ist ihnen zu laut, daher machen Sie wieder leiser.

6. Wechseln Sie zur Geräteauswahl.

7. Wählen Sie das Gerät CD-Spieler.

8. Wechseln Sie zu Titel 4.

9. Erhöhen Sie die Lautstärke um eine Stufe.

9. (Loop) Erniedrigen Sie die Lautstärke um eine Stufe.

10. Wechseln Sie zu Titel 2.

11. Ihnen gefällt die aktuelle CD nicht. Gehen Sie zum CD-Wechsler.

12. Wählen Sie CD 3, Titel 10.

13. Machen Sie sehr leise.

13. (Loop) Stellen Sie die vorherige Lautstärke ein.

→ *Fragebogen Teil 2, Block 1 (Versuchsperson)*

→ *Einschätzungsbogen Teil 2, Block 1 (Versuchsleiter)*

Teil 2, Block 2

14. Stellen Sie die Lautstärke wieder auf ein sinnvolles Maß ein.

15. Wechseln Sie zum Telefon.

16. Rufen Sie den Teilnehmer Anne an.

Versuchsleiter nimmt Gespräch an:

17. Sie möchten wieder CD hören. Wechseln Sie zu CD 4.

18. Wenn das Telefon gleich klingelt gehen Sie bitte ran.

Versuchsleiter (als Herr Hunsinger) ruft an:

19. Schalten Sie das Telefongespräch auf *halten*, damit ihr Gesprächspartner am Telefon nicht mithören kann.

20. Führen Sie nun das Telefongespräch fort.

21. Beenden Sie das Telefongespräch.

22. Sie möchten wieder Radio hören. Wählen Sie einen beliebigen Sender und stellen Sie sich eine angenehm leise Lautstärke ein.

→ *Fragebogen Teil 2, Block 2 (Versuchsperson)*

→ *Einschätzungsbogen Teil 2, Block 2 (Versuchsleiter)*

Teil 2, Block 3

23. Machen Sie wieder lauter.

24. Rufen Sie Herrn Mayer an.

25. Sie wollten vor ihrem Anruf noch eine Route berechnen lassen. Brechen Sie daher den begonnenen Rufaufbau ab.

26. Wechseln Sie ins Gerät Navigation.

27. Gehen Sie in die Zieleingabe.

28. Geben Sie direkt über die Karte das Ziel Frankfurt ein und aktivieren Sie anschließend die Zielführung.

29. Rufen Sie nun Herrn Mayer an.

Versuchsleiter nimm Gespräch an.

30. Beenden Sie das Gespräch.

31. Rufen Sie beim ADAC an.

Versuchsleiter nimmt Gespräch an.

32. Rufen Sie wieder Herrn Mayer an.

Versuchsleiter nimmt Gespräch an.

32. (Loop) Rufen Sie erneut beim ADAC an.

Versuchsleiter nimmt Gespräch an.

33. Beenden Sie das Gespräch.

34. Wechseln Sie wieder zum Radio.

35. Schalten Sie den Ton stumm.

35. (Loop) Stellen Sie die vorherige Lautstärke ein.

→ *Fragebogen Teil 2, Block 3 (Versuchsperson)*

→ *Einschätzungsbogen Teil 2, Block 3 (Versuchsleiter)*

Teil 2, Block 4

36. Schalten Sie den Ton wieder an.

37. Rufen Sie Herrn Hunsinger an.

Versuchsleiter nimmt Gespräch an.

38. Gehen Sie ins Navigationssystem.

39. Wechseln Sie zur Zieleingabe und setzen Sie den Sucher auf München.

40. Vergrößern Sie die Karte, bis im Stadtplan von München die Technische Universität zu erkennen ist.

41. Geben Sie nun das Ziel BMW ein.

Systemrückfrage: „Soll die Zielführung gestartet werden?“

Bevor die Versuchsperson die Frage beantworten kann, ruft der Versuchsleiter an.

42. Nehmen Sie das Telefongespräch an.

43. Brechen Sie die Frage ab.

44. Verkleinern Sie die Karte und wählen Sie Hannover als neues Navigationsziel.

45. Aktivieren Sie nicht die Zielführung, sondern lassen Sie Hannover in den Zielspeicher aufnehmen.

46. Sie möchten nicht gestört werden. Lehnen Sie daher jeden kommenden Telefonanruf ab.

Versuchsleiter ruft an.

47. Wählen Sie einen beliebigen Titel einer beliebigen CD.

48. Schalten Sie den Ton stumm.

→ *Fragebogen Teil 2, Block 4 (Versuchsperson)*

→ *Einschätzungsbogen Teil 2, Block 4 (Versuchsleiter)*

Teil 2, Block 5 (Fehlerblock)

49. Schalten Sie zunächst den Ton wieder an.

50. Machen Sie die CD sehr laut.

Funktioniert erst beim ca. dritten Versuch.

51. Wählen Sie das Radio aus.

Funktion nach rechts/links wird wahllos umgekehrt erkannt. Auswahl erst beim ca. dritten Versuch.

52. Gehen Sie ins Navigationssystem und wählen Sie den Zielspeicher.

53. Wählen Sie Berlin aus und aktivieren Sie die Zielführung.

54. Wechseln Sie zum CD-Spieler.

55. Schalten Sie den Ton stumm.

Dies wird zunächst einige Male wahllos als rechts oder links erkannt.

→ *Fragebogen Teil 2, Block 5 (Versuchsperson)*

→ *Einschätzungsbogen Teil 2, Block 5 (Versuchsleiter)*

Teil 3: Bedienung während simulierter Autofahrt

„Bitte fahren Sie zunächst einige Zeit frei, d.h. ohne zusätzliche Aufgabenstellung, durch die Straßen von Chicago, um sich an den Fahrsimulator zu gewöhnen. Bitte halten Sie sich - zumindest in

einem gewissen Rahmen - an die Verkehrsregeln. Es ist (wie in den USA üblich) erlaubt, an einer roten Ampel rechts abzubiegen.
Im folgenden werden Ihnen wieder einige Aufgaben gestellt. Falls Sie bei der Bedienung Schwierigkeiten haben oder einfach nur zusätzliche Information wünschen, drücken Sie bitte wie bisher die Hilfetaste links vom Schaltknüppel. Bitte halten Sie den Bügel am Lenkrad weiterhin permanent gedrückt und lassen Sie ihn nur zur Ausführung einer einzelnen Handgeste oder zum Betätigen der Hilfetaste los. Bitte denken Sie daran, den Bügel auch zwischen direkt aufeinanderfolgenden Gesten zu halten. Im Gegensatz zum vorigen Versuchsteil funktioniert das System wieder ohne simuliertes Fehlverhalten oder Fehlerkennungen. Sind Sie bereit?“

56. Schalten Sie zunächst den Ton wieder an.

57. Wechseln Sie zu Bayern 3.

58. Rufen Sie Peter an.

Versuchsleiter nimmt das Gespräch an.

59. Beenden Sie das Gespräch.

60. Fahren Sie rechts 'ran. Speichern Sie Karlsruhe im Zielspeicher ab.

61. Fahren Sie weiter. Sie wollen Kassette hören. Spulen Sie auf der eingelegten Kassette zu Titel 3.

62. Da Sie gerade durch den Verkehr abgelenkt sind lehnen Sie den gleich kommenden Anruf ab.

Versuchsleiter ruft an.

63. Spulen Sie zum nächsten Titel.

64. Machen Sie etwas leiser.

65. Lassen Sie sich vom Navigationssystem nach Kiel führen.

66. Wechseln Sie zum CD-Spieler und wählen Sie einen beliebigen Titel.

67. Sie erwarten einen dringenden Anruf. Nehmen Sie daher den nächsten Anruf sofort an.

Versuchsleiter ruft an.

68. Sie möchten sich kurz voll und ganz auf den Verkehr konzentrieren. Schalten Sie daher das Telefongespräch auf halten.

69. Nachdem der Verkehr etwas übersichtlicher geworden ist, führen Sie das Gespräch fort.

70. Halten Sie das Gespräch erneut.

71. Statt das Gespräch fortzusetzen entschließen Sie sich es direkt zu beenden.

72. Schalten Sie die Zielführung aus.

73. Wechseln Sie zu CD 2, Titel 4.

74. Machen Sie den Ton aus.

75. Aktivieren Sie die Zielführung nach Karlsruhe.

76. Wählen Sie Bayern 5.

77. Schalten Sie den Ton wieder an.

78. Rufen Sie Herrn Ruske an und sagen Sie ihm, dass Sie in Karlsruhe angekommen sind.

Versuchsleiter nimmt Gespräch an.

79. Wechseln Sie zum Radio.

80. Schalten Sie zum Abschluss den Ton stumm.

→ *Fragebogen Teil 3 (Versuchsperson)*

→ *Einschätzungsbogen Teil 3 (Versuchsleiter)*

A.4 Hardware-Komponenten

A.4.1 Infrarot-Distanz-Mess-Sensor

Original-Datenblatt des Sensors SHARP GP2D12.

SHARP GP2D12/GP2D15

GP2D12/GP2D15

General Purpose Type Distance Measuring Sensors

■ Features

1. Less influence on the color of reflective objects, reflectivity
2. Line-up of distance output/distance judgement type
   Distance output type (analog voltage) : **GP2D12**
   Detecting distance : 10 to 80cm
   Distance judgement type : **GP2D15**
   Judgement distance : 24cm
   (Adjustable within the range of 10 to 80cm)
3. External control circuit is unnecessary
4. Low cost

■ Applications

1. TVs
2. Personal computers
3. Cars
4. Copiers

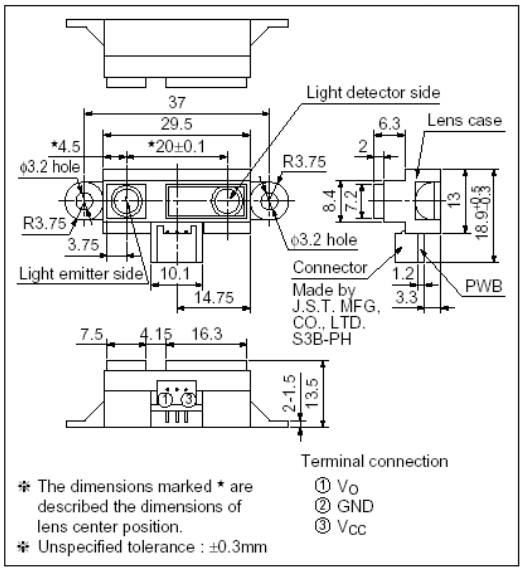
■ Absolute Maximum Ratings

(Ta=25°C, Vcc=5V)

Parameter	Symbol	Rating	Unit
Supply voltage	V_{CC}	−0.3 to +7	V
Output terminal voltage	V_O	−0.3 to Vcc +0.3	V
Operating temperature	T_{opr}	−10 to +60	°C
Storage temperature	T_{stg}	−40 to +70	°C

■ Outline Dimensions

(Unit : mm)



Notice	In the absence of confirmation by device specification sheets, SHARP takes no responsibility for any defects that may occur in equipment using any SHARP devices shown in catalogs, data books, etc. Contact SHARP in order to obtain the latest device specification sheets before using any SHARP device.
Internet	Internet address for Electronic Components Group http://www.sharp.co.jp/ecg/

■ Recommended Operating Conditions

Parameter	Symbol	Rating	Unit
Operating supply voltage	Vcc	4.5 to +5.5	V

■ Electro-optical Characteristics

(Ta=25°C, Vcc=5V)

Parameter		Symbol	Conditions	MIN.	TYP.	MAX.	Unit
Distance measuring range		ΔL	*1 *3	10	–	80	cm
Output terminal voltage	GP2D12	Vo	L=80cm *1	0.25	0.4	0.55	V
	GP2D15	VOH	Output voltage at High *1	Vcc –0.3	–	–	V
		VOL	Output voltage at Low *1	–	–	0.6	V
Difference of output voltage	GP2D12	ΔVo	Output change at L=80cm to 10cm *1	1.75	2.0	2.25	V
Distance characteristics of output	GP2D15	Vo	*1 *2 *4	21	24	27	cm
Average Dissipation current		Icc	L=80cm *1	–	33	50	mA

Note) L : Distance to reflective object.

*1 Using reflective object : White paper (Made by Kodak Co. Ltd. gray cards R-27 · white face, reflective ratio ; 90%).
*2 We ship the device after the following adjustment : Output switching distance L=24cm±3cm must be measured by the sensor.
*3 Distance measuring range of the optical sensor system.
*4 Output switching has a hysteresis width. The distance specified by Vo should be the one with which the output L switches to the output H.

Fig.1 Internal Block Diagram

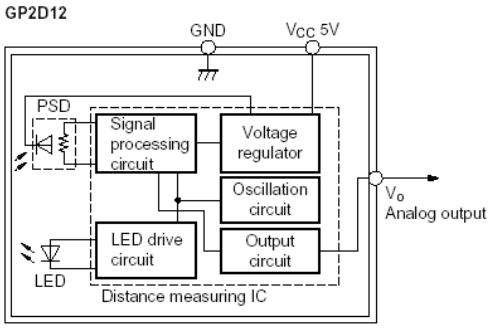


Fig.2 Internal Block Diagram

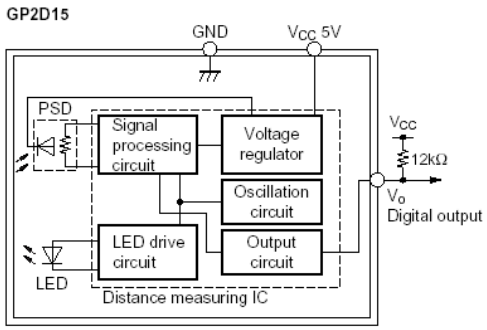


Fig.3 Timing Chart

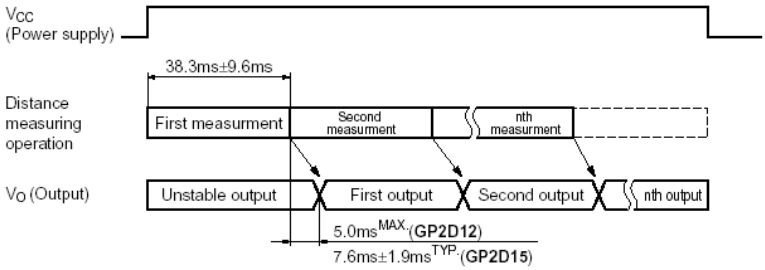


Fig.4 Distance Characteristics

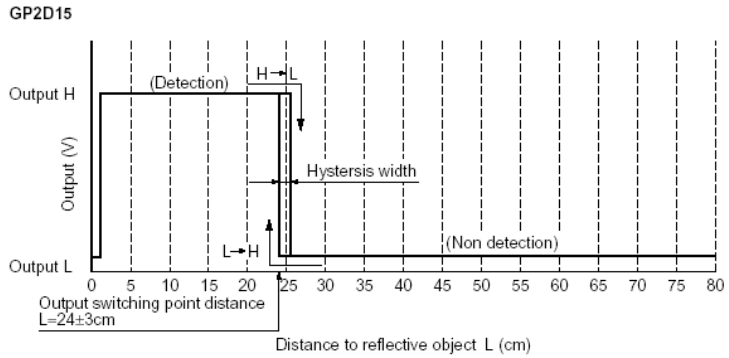


Fig.5 Analog Output Voltage vs. Surface Illuminance of Reflective Object

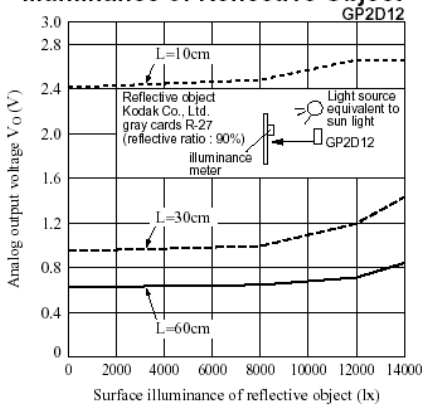


Fig.6 Analog Output Voltage vs.Distance to Reflective Object

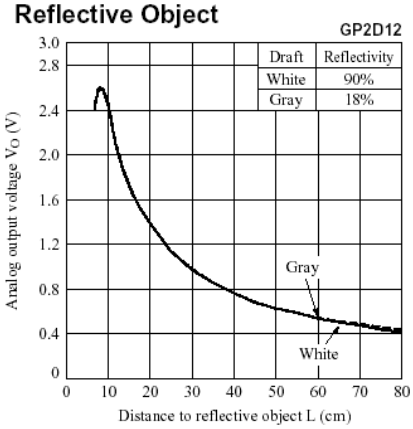


Fig.7 Analog Output Voltage vs.Ambient Temperature

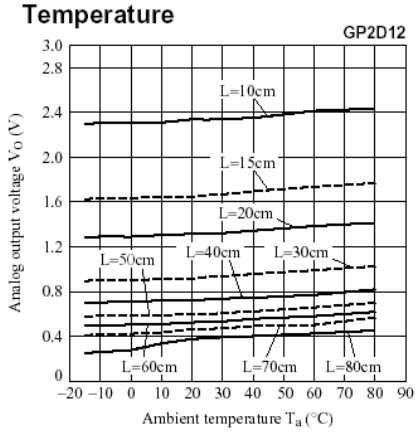
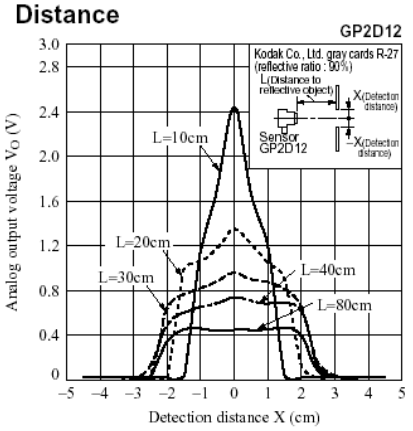
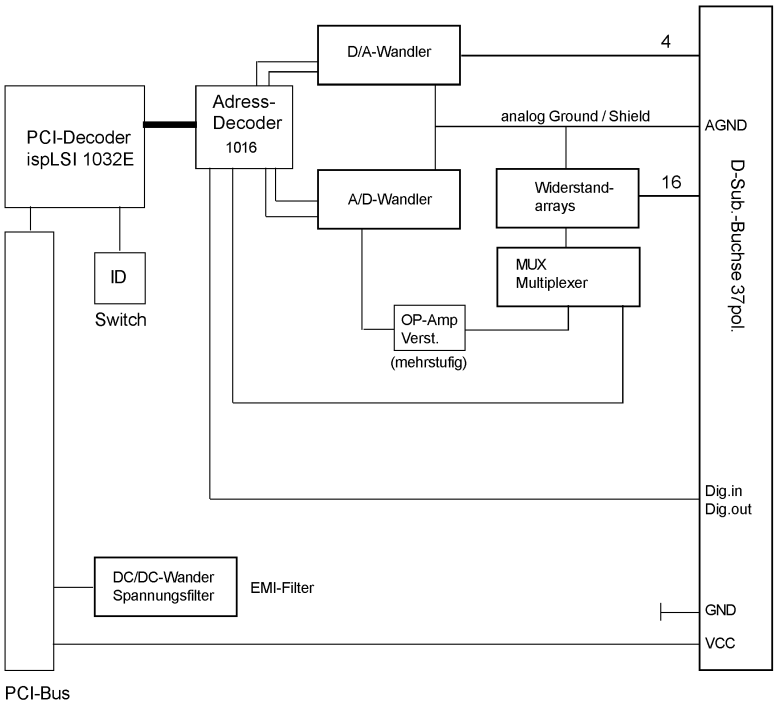


Fig.8 Analog Output Voltage vs.Detection Distance



A.4.2 A/D-Wandlerkarte KOLTER PCI AD12LC

Auszug aus dem Datenblatt der Firma KOLTER ELECTRONICS



Alle I/O Signale sind über eine 37-polige Sub-D-Buchse von außen am PC-Blech zugängig. Zur Versorgung externer Schaltungen sind zusätzlich, neben den beiden Digitalkanälen, aus dem PC die GND und +5 Volt-Leitung auf der Sub-D Buchse geführt. Bei der Verwendung der externen BNC-Box, die speziell für die PCI-AD-Karten entwickelt wurde und ebenfalls über KOLTER ELECTRONIC bezogen werden kann, ist der Transientenschutz (TS) überflüssig, da sich die entsprechenden Schutzbauteile bereits in der BNC-Box befinden.

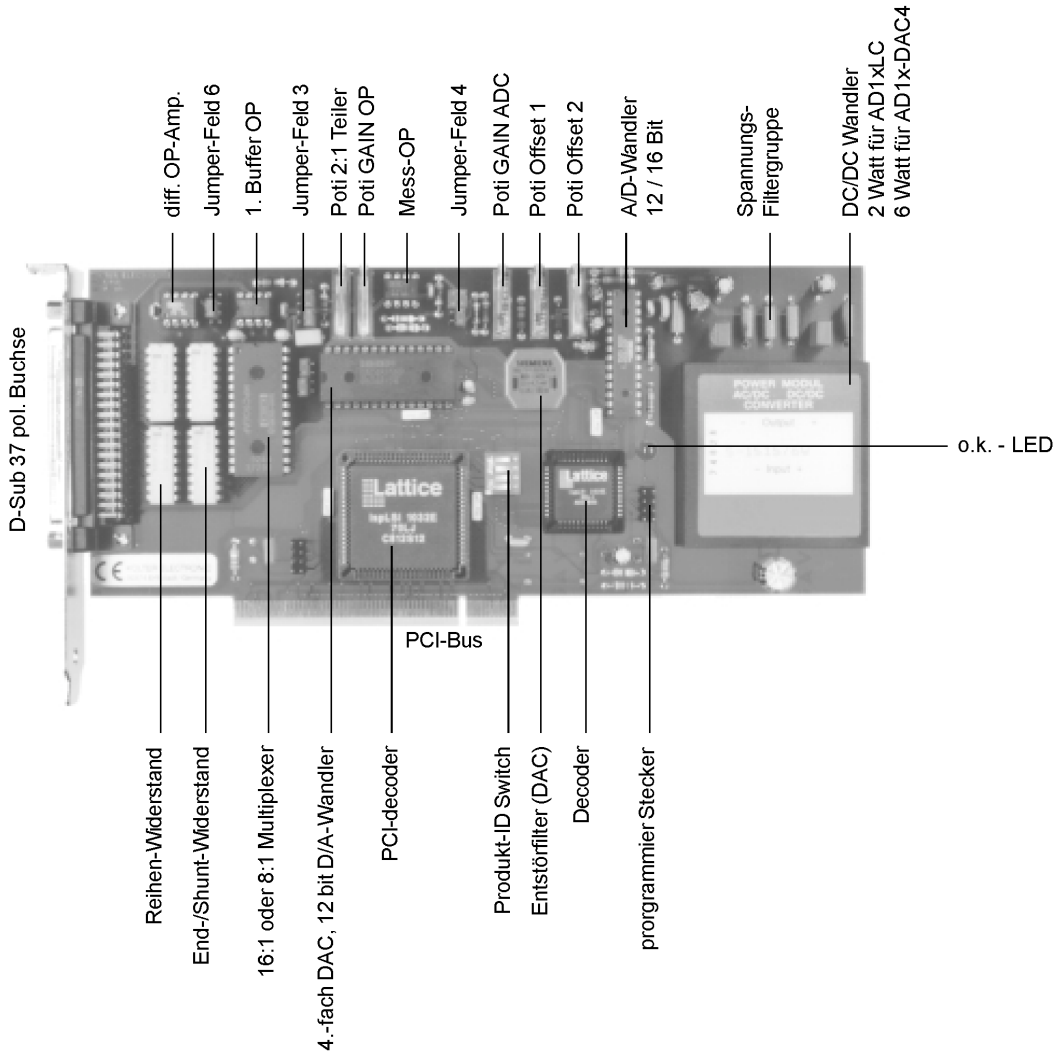
Die Einstellung der Kartenadressierung erfolgt automatisch über Plug&Play (PnP). Eine weitergehende Bauteilinitialisierung ist nicht erforderlich, A/D- und D/A-Wandler werden über die entsprechenden I/O-Port-Register direkt programmiert (siehe dazu die Programmbeispiele).

Andere Beispiele zur Kartenprogrammierung entnehmen Sie bitte den Texten auf der Diskette bzw. CD.

Zur Programmierung/Adressierung sind grundsätzlich folgende Offset-Adressen zu beachten:

in/out	Adresse	Data	Funktion	sonstiges
in	adr + 17	0/2	digital Input auf bit 1	restl. bits ohne Verwendung
out	adr + 16	0/1	digital Output auf bit 0	restl. bits ohne Verwendung
out	adr + 04	0..15	Multiplexer-Kanal setzen (ADC)	warten 0,25 bis 10 us
in	adr + 1	0/1	EOC-bit end-of-conversion	A/D-Wandlungsende abfragen
out	adr + 0	= 1	setze ADC auf lesen	R/C bit
out	adr + 0	= 0	convert ADC	A/D-Wandlung
out	adr + 0	= 1	high-byte von ADC aktiv	HB setzen
out	adr + 0	= 2	low-byte von ADC aktiv	LB setzen
in	adr + 0	0...255	ADC byte Register (Data-Port)	HB/LB lesen
in	adr + 29	xxx	D/A-Wandler-Register übergeben	DAC - latch
out	adr + 32	0...255	DAC A, lower byte setzen	Register DAC
out	adr + 36	0...255	DAC A, higher byte setzen	Register DAC
out	adr + 40	0...255	DAC B, lower byte setzen	Register DAC
out	adr + 44	0...255	DAC B, higher byte setzen	Register DAC
out	adr + 48	0...255	DAC C, lower byte setzen	Register DAC
out	adr + 52	0...255	DAC C, higher byte setzen	Register DAC
out	adr + 56	0...255	DAC D, lower byte setzen	Register DAC
out	adr + 60	0...255	DAC D, higher byte setzen	Register DAC

Kartenansicht und Bauteile



Allgemeines
Die Einstellung des DIP-Schalters darf nicht verändert werden. Der DIP-Schalter legt die Produkt-ID von 0x0012 Hex fest, damit das Rechner-BIOS eine entsprechende I/O-Adresse zuweisen kann und die Karte per Software identifiziert wird. Die LED zeigt an, ob die Karte richtig funktioniert beziehungsweise das PLSI richtig geladen wurde – also ein quasi Selbsttest.

Eingangsanpassung
Die PCI-AD Karte kann mit verschiedenen Widerstandarrays ausgerüstet werden, um so die zu messende Eingangsspannung optimal an den A/D-Wandler und Multiplexer anzupassen. Bei Herstellung beziehungsweise Auslieferung befinden sich Eingangsseitig zwei 10 Ω Widerstand-Arrays in Reihe zum Multiplexer und zwei Shunt-Arrays mit 100 kΩ gegen Analog-GND, um offenen A/D-Kanälen ein Null-Potential zuzuweisen.

A.5 Implementierte Gestenklassen und -aktionen

A.5.1 Handgestik

Gestenklasse	Kategorie	Reaktion GECoM
1) Winken nach rechts	kinemimisch	Zum nächsten Menüpunkt wechseln; Objekt (Navigationskarte) nach rechts bewegen
2) Winken nach links	kinemimisch	Zum vorigen Menüpunkt wechseln; Objekt (Navigationskarte) nach links bewegen
3) Winken nach oben	kinemimisch	Lautstärke erhöhen; Objekt (Navigationskarte) nach oben bewegen
4) Winken nach unten	kinemimisch	Lautstärke verringern; Objekt (Navigationskarte) nach unten bewegen
5) Winken nach vorne	kinemimisch	Objekt (Navigationskarte) verkleinern
6) Winken nach hinten	kinemimisch	Objekt (Navigationskarte) vergrößern
7) Zeigen nach vorne	symbolisch	Aktuellen Menüpunkt auswählen
8) Horizontale Wischbewegung	symbolisch	Aktuelle Funktion abbrechen; Lautstärke stumm schalten
9) Ziehen an virtuellem Griff	mimisch	Hauptmenü aufrufen (Geräteauswahl)
10) Virtuellen Telefonhörer abheben	mimisch	Zum Telefon wechseln; Teilnehmer anrufen; kommenden Anruf annehmen
11) Virtuellen Telefonhörer auflegen	mimisch	Telefon verlassen; Telefongespräch beenden; kommenden Anruf ablehnen
12) Gangschaltung betätigen	Aktion	Derzeit keine (Garbage-Modell)

A.5.2 Kopfgestik

Gestenklasse	Kategorie	Reaktion GECoM
1) Kopfnicken	symbolisch	Systemrückfrage bejahen; eingehenden Anruf annehmen
2) Kopfschütteln	symbolisch	Systemrückfrage verneinen; eingehenden Anruf ablehnen
3) Platznehmen auf dem Fahrersitz	Aktion	Derzeit keine (Garbage-Modell)
4) Aufstehen vom Fahrersitz	Aktion	Derzeit keine (Garbage-Modell)
5) Blickbewegung Frontscheibe → Mittelkonsole	Aktion	Derzeit keine (Garbage-Modell)
6) Blickbewegung Mittelkonsole → Frontscheibe	Aktion	Derzeit keine (Garbage-Modell)
7) Blickbewegung Frontscheibe → linke Seitenscheibe	Aktion	Derzeit keine (Garbage-Modell)
8) Blickbewegung linke Seitenscheibe → Frontscheibe	Aktion	Derzeit keine (Garbage-Modell)

A.6 Stichwortverzeichnis

A.7 Symbolverzeichnis

α_S	Signifikanzniveau
α, β , χ, δ	Gewichtungsfaktoren zur Berechnung des Gesamtkonfidenzmaßes C_{ges}
a_j	Bewegungsmaß für die Bewegungsdetektion zum Zeitpunkt j
a_{min}	Bewegungsschwellwert
c_1, c_2, c_3, c_4	Konfidenzmaße
C_{ges}	Gesamtkonfidenzmaß
C_{min}	Entscheidungsschwellwert bzw. Mindestkonfidenzmaß
d	tatsächliche Distanz zwischen Objekt und IR-Sensor
d_{HG}	Distanzschwellwert für die Hintergrundausblendung
d_{lin}	linearer Distanzwert
d_{min}	minimal messbare Distanz
d_{max}	maximal messbare Distanz
$d_{reg}[t_j]$	Distanzwert der Regressionsgerade zum Zeitpunkt j
$\underline{d}_{lin}[t_j]$	Distanzvektor zum Zeitpunkt j
D_{hyp}	kleinster sich ergebender DTW-Abstand beim Vergleich des unbekannten Musterverlaufs $\underline{\boldsymbol{M}}_?$ mit allen Referenzmustern des Trainingskorpus $\underline{\boldsymbol{T}}$
$\overline{D}_{hyp}$	mittlerer DTW-Abstand des unbekannten Musterverlaufs $\underline{\boldsymbol{M}}_?$ zur Klasse k_{hyp}
D_{next}	kleinster DTW-Abstand des unbekannten Musterverlaufs $\underline{\boldsymbol{M}}_?$ zu einem Referenzmuster $\underline{\boldsymbol{R}}_{next}$, das nicht der Hypothesenklasse k_{hyp} angehört
$\overline{D}_{next}$	mittlerer DTW-Abstand des unbekannten Musterverlaufs $\underline{\boldsymbol{M}}_?$ zur Klasse k_{next}
D_{kont}	Regeldistanz bei kontinuierlicher Gestik
D_{sw}	Distanzschwelle für das Umschalten zwischen Nahbereich- und Fernbereichsensor bei kontinuierlicher Gestik
$D_{DTW}(\underline{\boldsymbol{M}}_?,\underline{\boldsymbol{R}}_{k,l_k})$	DTW-Abstand zwischen dem unbekannten Musterverlauf $\underline{\boldsymbol{M}}_?$ und l_k-ter Referenzmustersequenz $\underline{\boldsymbol{R}}$ der Klasse k
$\underline{\underline{D}}_{dist}$	Distanzmatrix des DTW-Verfahrens
$\underline{\underline{D}}_{akk}$	akkumulierte Distanzmatrix des DTW-Verfahrens
$\overline{D}_{rest}$	mittlerer DTW-Abstand des unbekannten Musterverlaufs $\underline{\boldsymbol{M}}_?$ zu allen Klassen der Menge $\boldsymbol{k}_{rest}$
$h_i[t_j]$	zeitdiskrete Filterfunktion der i-ten Vektorkomponente

H_0	Nullhypothese
H_A	Alternativhypothese (bzw. Alternative)
e	Objekt-Fehlerkennungsrate
f_{abt}	Abtastfrequenz
ϑ	Winkel zwischen reflektiertem IR-Impuls und Horizontale
k	Gestenklasse
k_{hyp}	Hypothesenklasse
k_{next}	Gestenklasse, die das Referenzmuster $\underline{\boldsymbol{R}}_{next}$ enthält
$\boldsymbol{k}_{rest}$	Menge aller Gestenklassen ohne der Hypothesenklasse k_{hyp}
K	Anzahl der insgesamt vorhandenen Gestenklassen
L_k	Anzahl an Referenzmustern der Klasse k
μ_{Ξ}	Mittelwert der Verteilung der Differenzen aus haptischen und gestischen Beobachtungspaaren
m_i	Steigung der Regressionsgerade des i-ten IR-Sensors
M_i	i-te Komponente eines Merkmalvektors
$\underline{M}$	Merkmalvektor
$\underline{\boldsymbol{M}}$	Merkmalvektorsequenz
$\underline{\boldsymbol{M}}_f$	Merkmalvektorsequenz, gefiltert
$\underline{\boldsymbol{M}}_{fn}$	Merkmalvektorsequenz, gefiltert und normiert
$\underline{\boldsymbol{M}}_?$	unbekannter bzw. zu klassifizierender Musterverlauf
n	Gesamtanzahl der IR-Sensoren eines Arrays
$\underline{N}$	Normierungsvektor
o_{err}	Anzahl der falsch erkannten Objekte
o_{ges}	Gesamtanzahl der Objektdarbietungen
ρ	Regelfehler in Winkeldarstellung
r	Regelfehler als Effektivwert der Regelabweichung Δx
$\underline{\boldsymbol{R}}_{k,l_k}$	l_k-te Referenzmustersequenz der Klasse k
$\underline{\boldsymbol{R}}_{hyp}$	Referenzmustersequenz mit minimalem Abstand D_{hyp} zum unbekannten Musterverlauf $\underline{\boldsymbol{M}}_?$
$\underline{\boldsymbol{R}}_{next}$	Referenzmustersequenz mit Abstand D_{next} zum unbekannten Musterverlauf $\underline{\boldsymbol{M}}_?$
σ_1, σ_2	Sigmoidfunktionen zur Gewichtung der Distanzmesswerte bei kontinuierlicher Gestik

s_i	Signal des *i*-ten IR-Sensors
$S1, \ldots, S5$	IR-Sensor eins bis fünf
$\underline{\tau}_{lin}[t_j]$	Trendvektor zum Zeitpunkt *j*
t_{start}	Zeitpunkt des Gestenbeginns
t_{stop}	Zeitpunkt des Gestenendes
th_{start}	Mindestanzahl an Trendvektoren zur Detektion eines Bewegungsbeginns
th_{stop}	Mindestanzahl an Trendvektoren zur Detektion eines Bewegungsendes
Δt_{ges}	Gesamtdauer einer Geste
Δt_{min}	minimale Dauer einer Geste bei der Plausibilitätsprüfung
Δt_{max}	maximale Dauer einer Geste bei der Plausibilitätsprüfung
$\underline{\boldsymbol{T}}$	Trainingskorpus
$\underline{\boldsymbol{Tr}}_{\text{Hand},z\text{ P}}$	Trainingskorpus für die Handgestenerkennung mit Daten von *z* Personen
$\underline{\boldsymbol{Tr}}_{\text{Kopf},z\text{ P}}$	Trainingskorpus für die Kopfgestenerkennung mit Daten von *z* Personen
$\underline{\boldsymbol{Te}}_{\text{Hand},z\text{ P}}$	Testkorpus für die Handgestenerkennung mit Daten von *z* Personen
$\underline{\boldsymbol{Te}}_{\text{Kopf},z\text{ P}}$	Testkorpus für die Kopfgestenerkennung mit Daten von *z* Personen
u	Unterschiedsmaß zur Identifizierung redundanter Trainingsreferenzen
$U_{out}(d)$	distanzabhängige Ausgangsspannung eines IR-Sensors
$U_{app}(d)$	analytische Funktion zur Approximation der Ausgangsspannung
V	Gesamtanzahl an Referenzmustern im Trainingskorpus $\underline{\boldsymbol{T}}$
w_{tr}	Größe des Vektorpuffers bei der Regressionsgeraden-Berechnung
Δx	Regelabweichung
Δx_{hyp}	hypothetischer horizontaler Versatz
x_{ist}	horizontale Auslenkung der Ist-Marke
x_{soll}	horizontale Auslenkung der Soll-Marke
Ξ	Differenz eines Beobachtungspaares
X	Beobachtungsgröße
z	Ausführungszeit

Literatur

[AKY00] AKYOL, S.; CANZLER, U.; BENGLER, K.; HAHN, W.: *Gestengesteuerter Nachrichtenspeicher im Kraftfahrzeug.* 42. Fachausschusssitzung Anthropotechnik. 24.-25. Oktober, München, Deutsche Gesellschaft für Luft- und Raumfahrt, ISBN 3-932182-13-8, S. 319-328, 2000.

[ALP02] ALPERN, M.: *iWave: a car gesture interface to control navigation and entertainment.* Demo Session, IEEE Fourth International Conference on Multimodal Interfaces (ICMI'2002), Pittsburgh, PA, USA, 14.-16.10.2002.
Weitere Information: http://www.alpern.org/files/Alpern_Portfolio_bw.pdf

[AXT98] AXTELL, R.: *Gestures: the do's and taboos of body language around the world.* New York: Wiley, ISBN 0471183423, 1998.

[BOL80] BOLT, R. A.: *Put-that-there: Voice and Gesture at Graphics Interface.* Computer Graphics Journal of the Association of Computing and Machinery, Bd. 14, No. 3, S.262-270, 1980.

[CAI00A] CAIRNIE, N.; RICKETS, I. W.; MCKENNA S. J.; MCALLISTER, G.: *Using finger-pointing to operate secondary controls in automobiles.* Proceedings IEEE Intelligent Vehicles Symposium (IV 2000), Dearborn, MI, USA, 3.-5.10.2000. ISBN 0-7803-6363-9, S. 550-555.

[CAI00B] CAIRNIE, N.; RICKETS, I. W.; MCKENNA S. J.; MCALLISTER, G.: *A prototype adaptive finger-pointing interface for operating secondary controls in motor vehicles.* Proceedings IEEE International Conference on Systems, Man and Cybernetics (SMC 2000), Nashville, Tennessee, USA, 8.-11.10.2000. ISBN 0-7803-6586-0, S. 937-942.

[DEU98] DEUBEL, H., SCHNEIDER, W. X., PAPROTTA, I.: *Selective Dorsal and Ventral Processing: Evidence for a Common Attentional Mechanism in Reaching and Perception*. Visual Cognition, Vol. 5, Psychology Press (part of the Taylor & Francis Group), 1998. S. 81-107.

[EFR72] EFRON, D.: *Gesture, Race and Culture.* Mouton & Co., The Hague, The Netherlands, 1972 (Reprint of EFRON, D.: *Gesture and Environment*. King's Crown Press, New York, USA, 1941).

[FAS98] FASTENMEIER, W., GSTALTER, H.: *Ablenkungseffekte durch neuartige Systeme im Fahrzeug*. 2. Berliner Werkstatt Mensch-Maschine-Systeme, Pro Universitate Verlag, 1998.

[FIS03] FISCHER, S.: *Anwendung des Decision Tree bei multimodalen Interaktionen.* Bachelorarbeit, Fakultät für Elektrotechnik und Informationstechnik, Lehrstuhl für Mensch-Maschine-Kommunikation, Technische Universität München, 2003.

[FLE01] FLEISCHMANN, PH.: *Usability Engineering zur berührungslosen Mensch-Maschine-Interaktion mit dynamischen Gesten - Einsatz und Hilfebedarf.* Diplomarbeit, Fakultät für Elektrotechnik und Informationstechnik, Lehrstuhl für Mensch-Maschine-Kommunikation, Technische Universität München, 2001.

[FRA02] FRAEDRICH, W.: *Entwicklung und Evaluierung eines Gesamtkonzeptes zur gestischen Bedienung im Fahrzeug.* Diplomarbeit, Fakultät für Elektrotechnik und Informationstechnik, Lehrstuhl für Mensch-Maschine-Kommunikation, Technische Universität München, 2002.

[GEI98] GEIGER, M.: *Grundlagen für ein integriertes Bedienkonzept im Fahrzeug.* Diplomarbeit, Fakultät für Elektrotechnik und Informationstechnik, Lehrstuhl für Mensch-Maschine-Kommunikation, Technische Universität München, 1998.

[GEI01A] GEIGER, M.; ZOBL, M.; BENGLER, K.; LANG, M.: *Intermodal Differences in Distraction Effects while Controlling Automotive User Interfaces.* Proc. of the 9th Int. Conf. on Human-Computer Interaction (HCI International 2001), New Orleans, Louisiana, USA, 5.-10.8.2001. Ed.: Lawrence Erlbaum Ass., New Jersey, 2001. Vol. 1 "Usability Evaluation and Interface Design", S. 263-267.

[GEI01B] GEIGER, M.; NIESCHULZ, R.; ZOBL, M.; NEUSS, R.; LANG, M.: *Methods for Facilitation of Wizard-of-Oz Studies and Data Acquisition.* Proc. of the 9th Int. Conf. on Human-Computer Interaction (HCI International 2001), New Orleans, Louisiana, USA, 5.-10.8.2001. Ed.: Lawrence Erlbaum Ass., New Jersey, 2001. Poster Sessions: Abridged Proceedings, S. 191-193.

[GEI02A] GEIGER, M.; NIESCHULZ, R.; ZOBL, M.; LANG, M.: *Bedienkonzept zur Gestenbasierten Interaktion mit Geräten im Automobil - Gesture-Based Control Concept for In-Car Devices.* Tagungsband VDI/VDE - GMA Fachtagung USEWARE 2002, Darmstadt, 11.-12.06.2002. Düsseldorf: VDI-Verlag, 2002, Hrsg.: VDI. VDI-Berichte; 1678 "USEWARE 2002 Mensch-Maschine-Kommunikation/Design", S. 299-303.

[GEI02B] GEIGER, M.: *Kostenbewusste visuelle Interaktion über ein Distanzsensorenfeld.* Neue deutsche Patentanmeldung am 16.09.2002, Akz 102 42 890.5, Anmelder: Technische Universität München, Erfinder: Michael Geiger.

[HOF00] HOFMANN, M.; LANG, M.: *Belief Networks for a Syntactic and Semantic Analysis of Spoken Utterances for Speech Understanding.* Proc. ICSLP 2000, Peking, China, 16.-20.10.2000, China Military Friendship Publish, Vol. 2, S. 875-878.

[HOF01] HOFMANN, M.; LANG, M.: *Intention-based Probabilistic Phrase Spotting for Speech Understanding.* Proc. Of the Int. Symp. On Intelligent Multimedia, Video and Speech Processing, ISIMP 2001, Hong Kong, China, 2.-4.5.2001. Ed.: IEEE Hong Kong Chapter of Signal Processing, S.99-102.

[HOF03] HOFMANN, M.: *Intentionsbasierte maschinelle Interpretation von Benutzeraktionen.* Dissertation, Fakultät für Elektrotechnik und Informationstechnik, Technische Universität München, 2003 (eingereicht).

[HUN02] HUNSINGER, J.: *Multimodale Erfassung mathematischer Formeln durch einstufig-probabilistische semantische Decodierung.* Dissertation, Fakultät für Elektrotechnik und Informationstechnik, Technische Universität München, 2002.

[INT98] INTEL Corporation: *Intel Recognition Primitives Library Reference Manual*, Best.-Nr. 637785-007, 1998.

[JOH93] JOHANNSEN, G.: *Mensch-Maschine-Systeme*. Springer Verlag, ISBN 3-540-56152-8, Berlin-Heidelberg-New York,1993.

[KEN86] KENDON, A.: *Current Issues in the Study of Gesture*. In: NESPOULOUS, J.-L.; PERRON, P.; LECOURS, A. R. (eds.): *The Biological Foundation of Gestures: Motor and Semiotic Aspects,* Lawrence Erlbaum Ass., Hilldale, New Jersey, 1986, S. 23-47.

[LAN99A] LANG, M.: *Mensch-Maschine-Kommunikation 1.* Vorlesungsmanuskript, Lehrstuhl für Mensch-Maschine-Kommunikation, Technische Universität München, 1999.

[LAN99B] LANG, M.: *Mensch-Maschine-Kommunikation 2.* Vorlesungsmanuskript, Lehrstuhl für Mensch-Maschine-Kommunikation, Technische Universität München, 1999.

[LAN01] LANG, M.: *1. Kommunikationsdienste und Netze. 2. Informationsdarstellung.* Handbuch der Ergonomie, Band 2, Teil A-7.: Kommunikation und Information. Hrsg.: Bundesamt für Wehrtechnik und Beschaffung, Koblenz, 2001.

[LAN02] LANG, M.: *1. Sinnesorgane und Sinnesmodalitäten. 2. Interaktionsmodelle und Dialogformen.* Handbuch der Ergonomie, Band 4, Teil C-8.1.: Interaktion zwischen Mensch und Computer. Hrsg.: Bundesamt für Wehrtechnik und Beschaffung, Koblenz. Carl Hanser Verlag, München, 2002, 6. Ergänzungslieferung.

[MOR99] MORGUET, P.; LANG, M.: *Comparison of Approaches to Continous Hand Gesture Recognition for a Visual Dialog System.* Proceedings ICASSP 99 (Phoenix, Arizona, USA), IEEE, Vol. 6, 1999. S. 3549-3552.

[MOR00] MORGUET, P.: *Stochastische Modellierung von Bildsequenzen zur Segmentierung und Erkennung dynamischer Gesten.* Dissertation, Fakultät für Elektrotechnik und Informationstechnik, Technische Universität München, 2000.

[NEU00] NEUSS, R.: *Usability Engineering als Ansatz zum Multimodalen Mensch-Maschine-Dialog.* Dissertation, Fakultät für Elektrotechnik und Informationstechnik, Technische Universität München, 2000.

[NIE94] NIELSEN, J.: *Usability Engineering*. Morgan Kaufmann Publishers, San Francisco, CA, USA, Oktober 1994.

[NIE01] NIESCHULZ, R.; GEIGER, M.; BENGLER, K.; LANG, M.: *An Automatic, Adaptive Help System to Support Gestural Operation of an Automotive MMI.* Proc. of the 9th Int. Conf. on Human-Computer Interaction (HCI International 2001), New Orleans, Louisiana, USA, 5.-10.8.2001. Ed.: Lawrence Erlbaum Ass., New Jersey, 2001. Vol. 1 "Usability Evaluation and Interface Design", S. 272-276.

[NIE02A] NIESCHULZ, R.; SCHULLER, B.; GEIGER, M.; NEUSS, R.: *Aspekte effizienten Usability Engineerings.* Themenheft der Zeitschrift "it+ti", Schwerpunktthema "Usability Engineering", Oldenbourg Wissenschaftsverlag, München, 1/2002, S. 23-30.

[NIE02B] NIESCHULZ, R.; GEIGER, M.; ZOBL, M; LANG, M.: *Informartionsbedarf bei gestischer Interaktion im Fahrzeug - Need for Assistance in Automotive Gestural Interaction.* Tagungsband VDI/VDE - GMA Fachtagung USEWARE 2002, Darmstadt, 11.-12.06.2002. Düsseldorf: VDI-Verlag, 2002, Hrsg.: VDI. VDI-Berichte; 1678 "USEWARE 2002 Mensch-Maschine-Kommunikation/Design", S. 293-297.

[NIE03] NIESCHULZ, R.: Laufende Forschungsarbeiten am Lehrstuhl für Mensch-Maschine-Kommunikation, Technische Universität München,2003.

[NIG93] NIGAY, L.; COUTAZ, J.: *A Design Space for Multimodal Interfaces: Concurrent Processing and Data Fusion.* Tagungsband INTERCHI 1993, Amsterdam, Niederlande, S. 172-178, ACM Press, 1993.

[PAP99] PAPROTTA, I.: *Selektive Wahrnehmung und Handlungssteuerung, Die Kopplung von visueller Aufmerksamkeit und Bewegungszielselektion*. Shaker Verlag Aachen, ISBN 3-8265-6588-6,1999.

[PAY00] PAYER, M.: *Internationale Kommunikationskulturen.* 4. Nonverbale Kommunikation. 2. Gesten, Körperbewegungen, Körberhaltungen und Körperkontakt als Signale. Fassung vom 6. November 2000. Url: http://www.payer.de/kommkulturen/kultur042.htm.

[QUI93] QUINLAN, J. R.: *C4.5 Programs for Machine Learning*. Morgan Kaufmann, San Mateo, CA, USA, 1993.

[RAB89] RABINER, L. R.: *A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition.* Proc. of the IEEE, vol. 77, no. 2, S. 257-286, 1989.

[RUS94] RUSKE, G.: *Automatische Spracherkennung - Methoden der Klassifikation und Merkmalsextraktion.* Oldenbourg-Verlag, München, Wien, 1994.

[RUS97] RUSKE, G.: *Automatische Mustererkennung in der Sprachverarbeitung.* Kurzmanuskript zur Vorlesung, Lehrstuhl für Mensch-Maschine-Kommunikation, Technische Universität München, Stand SS 1997.

[SAC02] SACHS, L.: *Statistische Auswertungsmethoden.* Springer Verlag, ISBN 3-540-42448-2, 10. überarbeitete und aktualisierte Auflage 2002, Berlin-Heidelberg-New York, 2002.

[SIE98] SIEMENS AG: *Der mit Computern gestikuliert.* Archiv Forschung und Innovation 01.1998. http://w4.siemens.de/FuI/de/archiv/zeitschrift/heft1_98/artikel07.

[STU92] STURMAN, D. J.: *Whole Hand Input.* Massachusetts Institute of Technology, Boston, 1992.

[WEL99] WELCH, B. B.: Practical Programming in Tcl and Tk**,** ISBN 0-13-022028-0, Prentice-Hall, New Jersey, USA, 1999.

[YOU00] YOUNG, S.; KERSHAW, D.; ODELL, J.; OLLASON, D.; VALTCHEV, V.; WOODLAND, P.: *The HTK Book (for HTK Version 3.0)*, © 1995-1999 Microsoft Corporation.

[ZOB01] ZOBL, M.; GEIGER, M.; BENGLER, K.; LANG, M.: *A Usability Study on Hand Gesture Controlled Operation of In-Car Devices.* Proc. of the 9th Int. Conf. on Human-Computer Interaction (HCI International 2001), New Orleans, Louisiana, USA, 5.-10.8.2001. Ed.: Lawrence Erlbaum Ass., New Jersey, 2001. Poster Sessions: Abridged Proceedings, S. 166-168.

[ZOB02] ZOBL, M.; GEIGER, M.; MORGUET P.; NIESCHULZ, R.; LANG, M.: *Gestenbasierte Interaktion mit Geräten im Automobil - Gesture-Based Control of In-Car Devices.* Tagungsband VDI/VDE - GMA Fachtagung USEWARE 2002, Darmstadt, 11.-12.06.2002. Düsseldorf: VDI-Verlag, 2002, Hrsg.: VDI. VDI-Berichte; 1678 "USEWARE 2002 Mensch-Maschine-Kommunikation/Design", S. 305-309.

[ZOB03A] ZOBL, M.; GEIGER, M.; LANG, M.; RIGOLL G.: *A Realtime System for Hand Gesture Controlled Operation of In-Car Devices.* IEEE International Conference on Multimedia & Expo (ICME) 2003, Baltimore, MD, USA, 06. - 09.07.2003, zur Veröffentlichung angenommen.

[ZOB03B] ZOBL, M.; NIESCHULZ, R.; GEIGER, M.; LANG, M.; RIGOLL G.: *Gesture components for natural interaction with in-car devices.* The 5th International Workshop on Gesture and Sign Language based Human-Computer Interaction, Genova, Italy, 15. - 17.4.2003, zur Veröffentlichung angenommen.