

Institut für Informatik
der Technischen Universität München

**Kombination sensorischer Evidenzen
aus dem visuellen und dem
auditorischen Kanal**

Markus Hofbauer

Vollständiger Abdruck der von der Fakultät für Informatik der Technischen
Universität München zur Erlangung des akademischen Grades eines
Doktors der Naturwissenschaften (Dr. rer. nat.)
genehmigten Dissertation.

Vorsitzender: Univ.-Prof. Dr. J. Schmidhuber

Prüfer der Dissertation:

1. Univ.-Prof. Dr. Dr. h.c. mult. W. Brauer
2. Univ.-Prof. Dr. G. Hauske

Die Dissertation wurde am 26.01.2005 bei der Technischen Universität
München eingereicht und durch die Fakultät für Informatik am 08.06.2006
angenommen.

Zusammenfassung

Der Mensch empfängt mit seinen unterschiedlichen Sinnessystemen Informationen aus seiner Umwelt. Die verschiedenen Eigenschaften eines Objekts, wie z.B. Form, Geräusche, Geruch, Oberflächenbeschaffenheit, werden durch die verschiedenen Sinne getrennt registriert und im Gehirn wieder dem Objekt zugeordnet. Dabei findet auf unterschiedlichen Verarbeitungsebenen eine Interaktion der verschiedenen Signale statt. Diese Interaktion des multisensorischen Inputs kann dabei als Vorbild für die Datenfusion in technischen Systemen dienen.

In dieser Arbeit wurde die entstehende Kovariation der Signale aus dem auditiven und visuellen System untersucht. Zuerst wurde gezeigt, welche Methoden zur Datenfusion bei verschiedenen technischen Systemen angewendet werden. Anschließend wurden wichtige Erkenntnisse zur multimodalen Signalverarbeitung bei biologischen Systemen dargestellt. Ausgehend von einer Computersimulation wurde eine Hypothese über die Form der Informationsintegration beim Menschen aufgestellt. Diese Hypothese wurde anhand unterschiedlicher Wahrnehmungsexperimente untersucht. Die Modellierung der experimentellen Ergebnisse erfolgte auf unterschiedlichen Abstraktionsebenen. Dabei kamen sowohl Verfahren aus der Signalverarbeitung als auch ein kognitives Modell zur Evidenzkombination zur Anwendung.

In der Computersimulation wurde untersucht, welche statistischen Kovariationen im auditiven und visuellen Kanal bei typischen Relativbewegungen zwischen einem Beobachter und mehreren Objekten auftreten. Das sich dabei ergebende Muster war von der Relativbewegung zwischen Beobachter und Objekten in seiner Struktur unabhängig. Daraus ergab sich die Hypothese, dass die menschliche Signalverarbeitung an dieses spezielle Muster angepasst ist.

Die durchgeführten Wahrnehmungsexperimente lassen sich nach der Art der Bewegung und der Art der Stimuli unterscheiden. Zuerst wurde ein Detektionsexperiment mit einem sich in der Größe verändernden Quadrat und einem sich in der Amplitude ändernden 1 kHz Ton durchgeführt. Im zweiten Experiment wurden die aus dem ersten Experiment verwendeten Stimuli gegen einen Tigerkopf und ein Tigergebrüll ausgetauscht, um den Einfluss der Stimulusart auf die Wahrnehmungsleistung zu testen. Das nachfolgende Experiment bestand aus einer Absolutschwellenmessung bei einer seitlichen Objektbewegung mit abstrakten Stimuli. Beim letzten Experiment wurde die Präzision der Vorhersage über das Auftauchen eines bewegten „objekthaften“ Stimulus an einer bestimmten Position gemessen. Die gemessenen Interaktionseffekte zeigten deutliche Unterschiede. Wenn die Stimuli einem Objekt zugeordnet werden können, war die Integrationsleistung deutlich höher als bei abstrakten Stimuli. Bei der Modellierung der multisensorischen Interaktion konnte mit der Dempster-Shafer-Methode eine bessere Vorhersage der Daten erreicht werden als mit dem hybriden Modell aus Probability- und Linear-Summation.

Danksagung

Diese Arbeit konnte erst durch das Zusammenwirken vieler Menschen gelingen, bei denen ich mich an dieser Stelle für ihre Unterstützung bedanken möchte.

Frau Dr. Isolde von Bülow danke ich für ihren Einsatz als Koordinatorin des Graduiertenkolleg „Sensorische Interaktion in biologischen und technischen Systemen“ um eine nützliche, hilfreiche und freundliche interdisziplinäre Kommunikationsplattform.

Frau Prof. Dr. Kerstin Schill und Herrn Dr. Christoph Zetzsche danke ich für die Betreuung dieser Arbeit und die liebenswürdige Aufnahme in ihr Arbeitsumfeld.

Bei Herrn Prof. Dr. Dr. h.c. mult. Wilfried Brauer bedanke ich mich für die Annahme und für die freie Gestaltung dieser Arbeit.

Herrn Prof. Dr.-Ing. Gert Hauske danke ich für die konstruktiven Beiträge, welche diese Arbeit übersichtlicher und vollständiger machten.

Herrn Prof. Dr. Hans Distel danke ich für seine Unterstützung bei den formalen Aspekten dieser Arbeit.

Meiner Kolpingfamilie danke ich für die moralische Unterstützung, sowie für das Korrekturlesen der Arbeit.

Meiner Freundin Sigrid danke ich für ihre Liebe und ihre Geduld, die sie mir während der Entstehung dieser Arbeit entgegengebracht hat.

Inhaltsverzeichnis

1	Einführung	1
1.1	Grundlagen	1
1.2	Datenfusion in technischen Systemen	5
1.3	Datenfusion in biologischen Systemen	21
2	Methodik	33
2.1	Psychophysik	33
2.2	Statistik	35
2.3	Modelle	35
2.3.1	Maximum-Likelihood-Integrator	35
2.3.2	Linear Summation	37
2.3.3	Probability Summation	40
2.3.4	Dempster-Shafer Evidenztheorie	44
2.3.5	Minimum-Modell	48
3	Computersimulation eines multisensorischen Inputs	49
3.1	Methode	49
3.2	Ergebnis	51
3.3	Auswertung	51
4	Experimente zur Bewegung in Blickrichtung	59
4.1	Abstrakte Objektbewegung	59
4.1.1	Methode	61
4.1.2	Ergebnisse	61
4.1.3	Modellierung	63
4.1.4	Folgerungen	69
4.2	Natürliche Objektbewegung	70
4.2.1	Methode	70
4.2.2	Ergebnisse	71
4.2.3	Modellierung	71
4.2.4	Folgerungen	74
5	Experimente zur Bewegung in frontparalleler Ebene	77

5.1	Strukturlose Bewegung	77
5.1.1	Methode	78
5.1.2	Ergebnisse	79
5.1.3	Modellierung	80
5.1.4	Folgerungen	84
5.2	Objektbewegung	85
5.2.1	Methode	86
5.2.2	Ergebnisse	91
5.2.3	Modellierung	97
5.2.4	Folgerungen	101
6	Diskussion	107

Kapitel 1

Einführung

1.1 Grundlagen

Unser Gehirn verarbeitet Sinnesdaten aus unterschiedlichen Modalitäten. Jeder Sinn für sich liefert bereits eine Fülle an Information über unsere Umwelt. Dabei stellt sich zuerst die Frage nach der Notwendigkeit zur Informationserfassung im Allgemeinen und anschließend die Frage nach der Notwendigkeit zur Informationsintegration. Für biologische Systeme dient die Datenaufnahme in erster Linie dazu, den Organismus am Leben zu erhalten. Dazu zählt das Finden von Nahrung und die Erkennung von Feinden. Um mit anderen Lebewesen kommunizieren zu können, ist es auch notwendig, Informationen von anderen Lebewesen zu empfangen. Die menschliche Informationsverarbeitung ist in Abbildung 1.1 nach dem Modell von Streitz [Str87] schematisch dargestellt. Die Verarbeitungsstufen lassen sich in die nachstehenden vier aufeinander folgenden Schritte einteilen:

1. Informationsaufnahme,
2. Informationstransformation,
3. Informationsspeicherung,
4. Informationsabgabe.

Das Rezeptorsystem, bestehend aus sensorischem Register und perzeptuellem Prozessor dient der Aufnahme eines Reizes aus der Umwelt oder dem eigenen Organismus. Im perzeptuellen Prozessor wird die im sensorischen Register noch reiznah repräsentierte Information für die Weiterverarbeitung in der zentralen Verarbeitungseinheit umkodiert. Im perzeptuellen Prozessor findet eine Mustererkennung und eine syntaktische Analyse des Erkannten statt. Die erkannten Objekte erhalten eine Bezeichnung. Dazu greift der perzeptuelle Prozessor auf das im Langzeitgedächtnis vorhandene Wissen über Muster und Bezeichner zu.

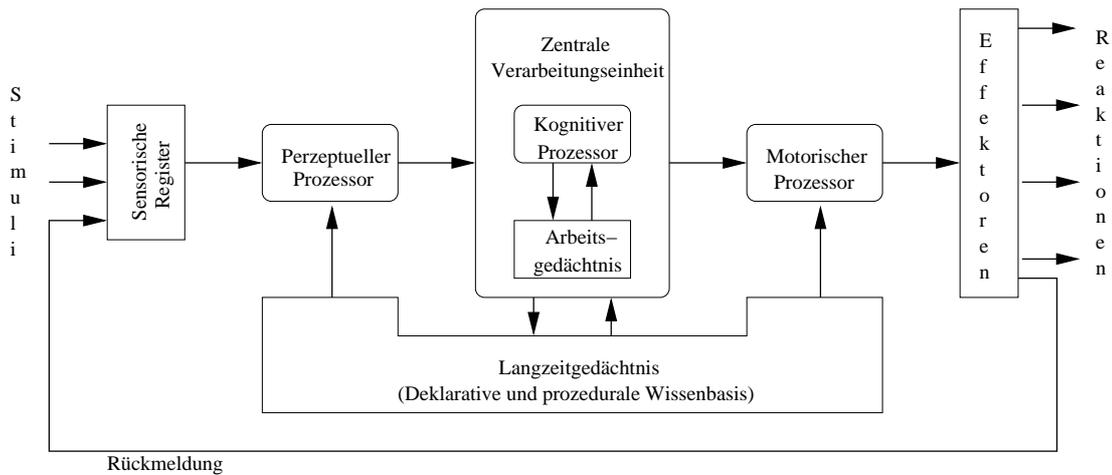


Abbildung 1.1: Das Modell der menschlichen Informationsverarbeitung nach Streitz.

Die zentrale Verarbeitungseinheit besteht aus dem aktiven kognitiven Prozessor und dem eher passiven Arbeitsgedächtnis. Der kognitive Prozessor wertet die vom perzeptuellen Prozessor aufbereitete Information aus. Es findet eine semantische Beschreibung und Bewertung der verarbeiteten Information im Hinblick auf die anstehende Aufgabe statt. Dazu wird zu Such- und Vergleichszwecken wieder auf das Langzeitgedächtnis zugegriffen. Daraus resultiert eine Handlungsanforderung, die eine Reaktionsauswahl bedingt.

Die Reaktionsauswahl dient als Eingabe für den motorischen Prozessor, der wiederum das im Langzeitgedächtnis gespeicherte entsprechende motorische Programm abrufen. Über die Effektoren wird das motorische Programm zur Ausführung gebracht. Diese Reaktion wird dann als externes Verhalten sichtbar.

Das Verhalten wird vom Informationsverarbeitungssystem wieder wahrgenommen. Durch diese Rückmeldungsschleife kann die Reaktion und dessen Auswirkungen im Hinblick auf die zu erledigende Aufgabe bewertet und gegebenenfalls korrigiert werden.

Durch die eben beschriebenen vier Schritte der Informationsverarbeitung wird unter anderem ein Abbild der Umwelt und der darin vorhandenen Strukturen erstellt. Besonders im ersten Schritt findet bei der Informationsaufnahme durch die Sinnesorgane eine gewaltige Datenreduktion statt und es geht viel Information über die Umwelt verloren. Dies liegt zum einen an einem begrenzten Arbeitsbereich unserer Sinnesorgane. Die physikalischen Reize müssen in einem bestimmten Bereich liegen, damit sie vom entsprechenden Sinnesorgan überhaupt verarbeitet werden. Ist der physikalische Reiz zu schwach, findet keine Reizung statt. Ist der Reiz zu stark, kann das Sinnesorgan beschädigt oder sogar zerstört werden. Beim Hörsinn z.B. sind die limitierten Faktoren die Amplitude und die Frequenz des

Schalls. Das Ohr hat einen Arbeitsbereich von ca. 0-140 dB für die Amplitude und ca. 20-20.000 Hz für die Frequenz des Signals [Bro99]. Der andere Faktor, der die Datenreduktion beeinflusst, ist die begrenzte Übertragungskapazität der Sinnessysteme. Die z.B. vom Auge erfasste optische Information beträgt ca. 10^6 bit/s [ZE99]. Die Informationsmenge, welche wir bewusst verarbeiten können, beträgt ca. 40-100 bit/s. Die Speicherfähigkeit des Gedächtnisses wird auf 1-10 Bit/s geschätzt. Bereits auf der Netzhaut des Auges findet durch die unterschiedliche Dichte von Stäbchen und Zapfen an bestimmten Stellen eine ganz unterschiedliche Informationsverarbeitung statt. Aufgabe der Sinnessysteme ist es, aus der Vielzahl an Informationen, die relevante Information herauszufiltern.

Bei der Datenverarbeitung durch technische Sensorsysteme soll nun anhand der Sensordaten auf den Zustand und die Struktur der Umwelt geschlossen werden. Dies ist aufgrund der komplexen Struktur der Umwelt nicht ohne weiteres möglich. Um trotzdem Rückschlüsse auf die Umwelt ziehen zu können, müssen bestimmte Annahmen über die Umwelt gemacht werden. Diese a priori Annahmen sind höchstens in einer künstlichen Welt sicher zu definieren. In unserer realen Umwelt kann nur auf bereits vorhandene Erfahrungen zurück gegriffen werden. Trotzdem können und müssen bestimmte Randbedingungen festgelegt werden. Dazu eignen sich die Gesetze aus der Physik und Mathematik. Dabei muss angenommen werden, dass diese Gesetze universelle Gültigkeit besitzen. Hinzu kommt, dass die Sensordaten verrauscht sind und somit die erfasste Information nicht exakt ist. Als Rauschen werden Störungen bezeichnet, die dem reinen Signal überlagert sind. Das Rauschen kann verschiedene Ursachen haben. Rauschen kann durch z.B. durch die verwendeten Komponenten entstehen (thermisches Rauschen), es kann aber auch von außen zugeführt werden, z.B. durch Nebengeräusche bei einer Sprachübertragung [Kü68]. Während unser Gehirn selbst mit extrem verrauschten Daten noch brauchbare Ergebnisse erzielt, stellt das Rauschen bei technischen Systemen ein ständig präsent Problem dar. Ein Mensch kann z.B. den Text einer "schlechten" Kopie relativ mühelos lesen und selbst völlig unleserliche Buchstaben aufgrund seiner kognitiven Fähigkeiten ergänzen. Ein technisches System hingegen, stößt bei einer solchen Aufgabe schnell an seine Leistungsgrenzen. Abbildung 1.2 und Abbildung 1.3 zeigen zur Illustration das Ergebnis eines technischen Filtervorgangs. Das Originalbild in Abbildung 1.2 (links) wurde mit einem Kantenfilter bearbeitet. Im Ergebnisbild (rechts) sind die Konturen gut erkennbar. Abbildung 1.3 (links) zeigt eine verrauschte Originalbild. Im bearbeiteten Bild (rechts) ist das Flugzeug nicht mehr zu erkennen. In verrauschten Bildern lassen sich Kanten, die vom visuellen System des Menschen noch gut erkannt werden, mit einem technischen Filter nur sehr schwer detektieren.

Zusätzlich hat jeder technische Sensor, wie unsere Sinne, nur ein begrenztes Auflösungsvermögen. Bei Systemen mit hohem Auflösungsvermögen müssen enorme Datenmengen verarbeitet werden. Dadurch steigt der Rechenaufwand und da-

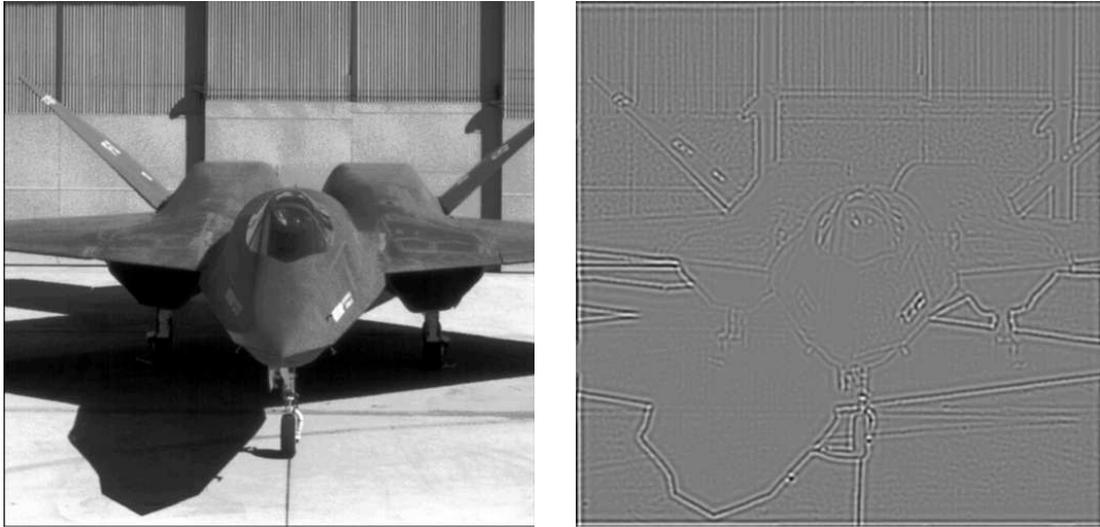


Abbildung 1.2: Beispiel eines Kantenfilters. Links das Originalbild, rechts das mit einem isotropen Bandpassfilter bearbeitete Bild. Die Konturen des Flugzeugs sind gut zu erkennen.

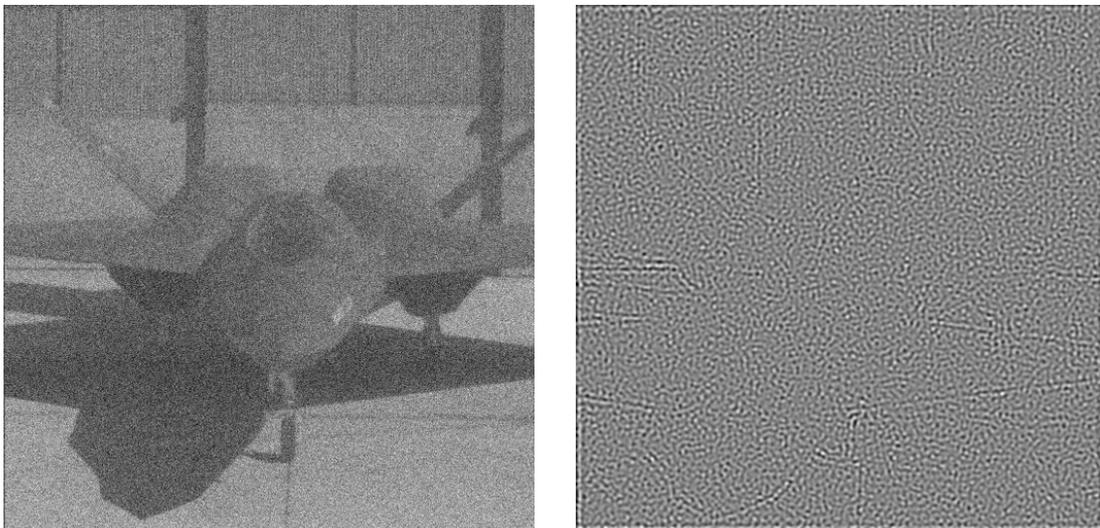


Abbildung 1.3: Beispiel eines Kantenfilters. Links das verrauschte Originalbild, rechts das mit einem isotropen Bandpassfilter bearbeitete Bild. Das Filter ist nicht mehr in der Lage die Kanten zu erkennen.

mit die Rechenzeit. Bei einfachen Systemen ist zwar eine schnellere Verarbeitung möglich, dafür ist die Qualität des resultierenden Signals in der Regel schlechter. Hieraus ergibt sich zwangsläufig die Idee der Informationsintegration aus verschiedenen Sensorkanälen, um möglichst schnell gesicherte Erkenntnisse über die Umwelt zu erlangen.

Die Eindrücke aus verschiedenen Modalitäten können im Gehirn einem Objekt zugeordnet werden. Um diese Leistung zu vollbringen, muss Information aus verschiedenen Kanälen auf irgendeine Art und Weise miteinander verknüpft werden.

Ein Ziel dieser Arbeit ist, herauszufinden, wo diese Verknüpfung für bestimmte Modalitäten stattfindet und wie sie beschrieben werden kann.

1.2 Datenfusion in technischen Systemen

Diese Arbeit ist eingebettet in einen neuen Forschungsbereich, der versucht unterschiedliche Sinnessysteme und deren Interaktion zu berücksichtigen. Die Multisensorische Forschung beschäftigt sich mit der Verknüpfung von auditorischen, visuellen, taktilen, haptischen, proprioceptiven, olfaktorischen und somatosensorischen Daten. Langfristiges Ziel ist es, möglichst viele Sinnessysteme miteinander in Bezug zu setzen. Meist werden aus Gründen der Komplexität nur zwei Modalitäten gleichzeitig untersucht. Nicht nur die Modalitäten sind vielfältig, sondern auch die möglichen Betrachtungsweisen der Interaktion. Im Folgenden werden beispielhaft einige verschiedene Anwendungen der Datenfusion in technischen Systemen erläutert. Damit soll zum einen die Notwendigkeit der Datenfusion deutlich werden und zum anderen die Komplexität dieser Systeme erkannt werden. Darüber hinaus soll auch gezeigt werden, welche Probleme in diesen Systemen auftreten und welche Unzulänglichkeiten damit verbunden sind.

Bildsequenzanalyse

Zur dynamischen Analyse von Bildsequenzen, welche mit verschiedenen Sensoren aufgenommen werden, können Neuronale Netzwerke eingesetzt werden. In [Hun92] werden die verschiedenen Stufen der Bildanalyse mit einem dreischichtigen Netzwerk abgebildet, welches in Abbildung 1.4 dargestellt ist. Als Eingabe dienen die Sensordaten zu einem bestimmten Zeitpunkt. In der ersten Schicht (unterer Cluster Analysis Level) wird mit einem selbstorganisierendem Kohonen-Netzwerk eine Gruppierung von Pixeln der jeweils zweidimensionalen Sensordaten vorgenommen.

In der zweiten Schicht (grau unterlegter Fusionsteil) werden die erkannten Bereiche integriert. Dabei werden sechs Fusionsfilter realisiert. Diese entsprechen

sechs Typen von Neuronen die im Optischen Tectum der Klapperschlange gefunden wurden [NH81]. Das AND-Neuron reagiert nur, wenn beide Eingangssignale vorhanden sind. Das OR-Neuron antwortet, wenn mindestens ein Signal am Eingang anliegt. Sowohl das S1E- als auch das S2E-Neuron liefert bei unimodaler Reizung eine schwache Antwort und bei bimodaler Reizung eine starke Antwort. Das S1I-Neuron antwortet stark, wenn nur ein Signal vom Sensor 2 vorliegt und wird inhibiert, falls von Sensor 1 ebenfalls ein Signal kommt. Das S2I-Neuron verhält sich entsprechend mit vertauschten Sensoreingängen.

In der dritten Schicht (oberer Cluster Analysis Level) werden aus den fusionierten Sensordaten zusammenhängende Bildbereiche extrahiert. Dies geschieht mittels eines Neuronalen Netzes, welches für jedes Pixel des Bildes einen eigenen Knoten besitzt. Diese Knoten sind zweidimensional angeordnet. Jeder Knoten hat eine Verbindung zu seinen vier nächsten Nachbarn. Das Netzwerk bildet somit die zweidimensionale Struktur des Eingangsbildes der Sensoren ab. Die Knoten erhalten ihre Eingabewerte aus den sechs Fusionsfiltern der zweiten Schicht und verändern dabei ihre Gewichte zu den Nachbarknoten. Anhand der endgültigen Gewichte zwischen den Knoten lassen sich zusammenhängende Regionen im Bild beschreiben. Aufgrund der komplexen Struktur dieser Verarbeitungsschicht war eine Implementierung auf einem Parallelrechner notwendig. Anschließend werden geeignete Merkmale generiert, welche die Region beschreiben. Dabei werden die Fläche, der Schwerpunkt, die Orientierung und die Exzentrizität der Region berechnet.

Diese drei Verarbeitungsschritte werden parallel mit dem zeitlich darauf folgenden Sensordatensatz durchgeführt. Danach wird aus den zwei Ergebnissen durch eine weitere Analyse eine mögliche Objektbewegung berechnet. Mittels einer Formveränderungsanalyse werden die Parameter einer möglichen Objektbewegung berechnet. Dazu wurde ein back-propagation Netzwerk verwendet, welches aus den Regionsmerkmalen die Translation oder Rotation der Region berechnet.

In den experimentellen Tests zur Segmentierung statischer Bilder wurde das Neuronale Netzwerk gegen einen Dempster-Shafer-Ansatz getestet. Für einige ausgewählte Ergebnisse liefern beide Verfahren gute Ergebnisse. Daraus lassen sich aber keine Schlüsse auf die tatsächliche Leistungsfähigkeit des Systems ziehen. Es wird auch ein Beispiel zur erfolgreichen Bewegungserkennung gezeigt. Das Hauptproblem dieses Verfahrens liegt aber wohl an den komplexen Berechnungen in der dritten Schicht.

Objekterkennung

Eine andere Anwendung der Datenfusion ist im Bereich der Objekterkennung zu finden. Hutchinson beschreibt in [HK92] eine Methode, mit der aus der Merkmalsextraktion mehrerer Sensoren und der Verknüpfung der Ergebnisse mittels der

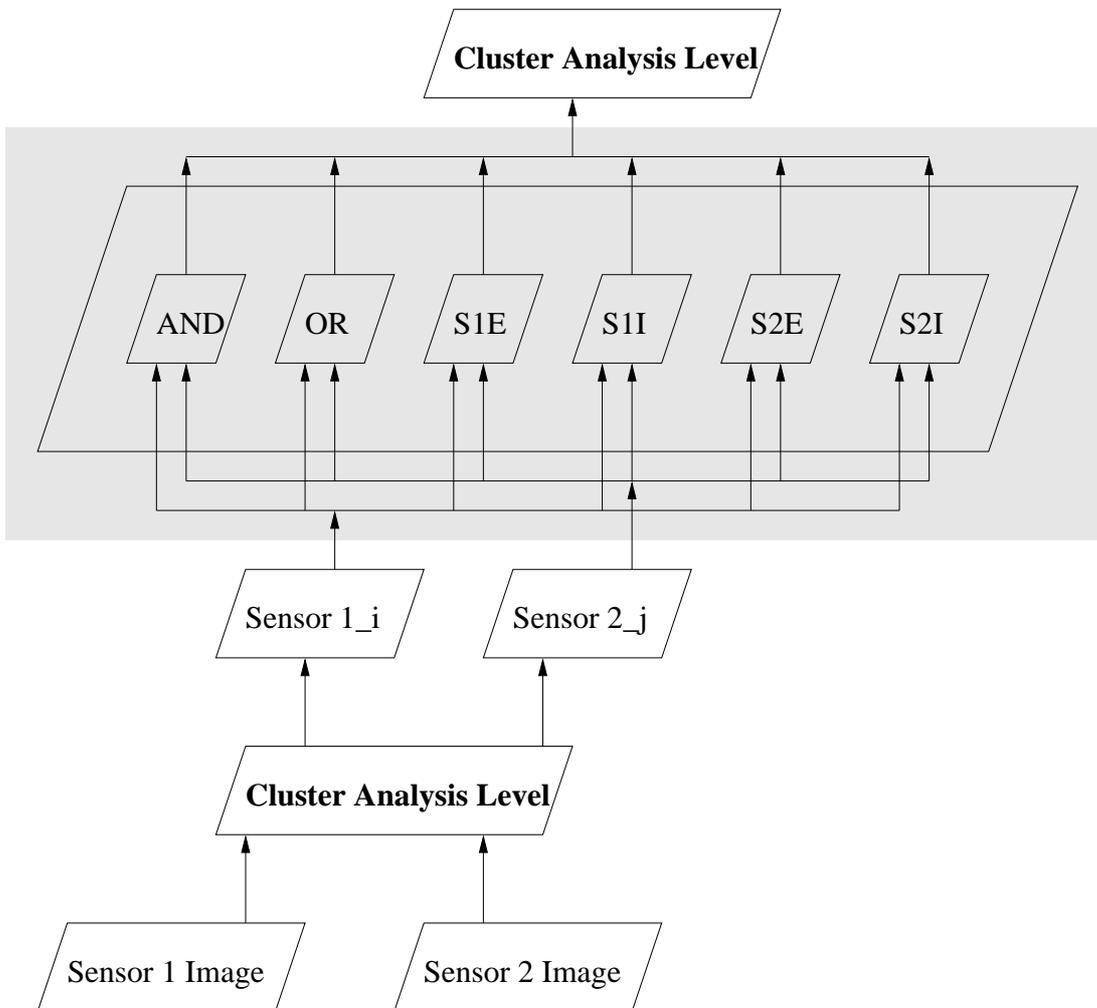


Abbildung 1.4: Die Bildsequenzanalyse gliedert sich in drei Teile: Bildvorverarbeitung mit Clusteranalyse (unterer Cluster Analysis Level), Fusionsteil (grau unterlegt) mit sechs Fusionsfiltern und Merkmalsextraktion mit Bewegungsanalyse (oberer Cluster Analysis Level).

Dempster-Shafer-Theorie ein dreidimensionales Objekt erkannt werden kann. Einer der Sensoren ist ein structured light scanner, womit dreidimensional bestimmte Merkmale des Objekts erkannt werden. Dabei wird eine Liste der Flächen des Objekts erstellt und deren Beziehung zueinander berechnet. Anschließend wird das Ergebnis mit den gespeicherten Modelldaten zu einem möglichen Objekt verglichen und Hypothesen über die Art des Objekts und dessen Lage im Raum aufgestellt. Der zweite Sensor ist eine Kamera zur Extraktion zweidimensionaler Merkmale. Die Merkmalsextraktion der Kamera ist weit weniger robust als die des Scanners, aber auch weniger rechenintensiv. Mit der Kamera als zusätzlichem Sensor soll auch die Anzahl der notwendigen Aufnahmen des Scanners verringert werden, da der Scanner für jede Aufnahme bewegt werden muss und dies zusätzlich Zeit kostet. Die aus dem Kamerabild extrahierten Merkmale beschränken sich auf Kanten und Löcher in der Oberfläche. Dabei steht nicht so sehr die Kantenerkennung, sondern die Überprüfung von Kanten im Vordergrund. Wenn bereits eine Hypothese über Art und Lage des Objekts erstellt wurde, kann berechnet werden, an welcher Stelle im Kamerabild eine Kante vorhanden sein muss. Diese Berechnung kann mit dem tatsächlichen Kamerabild verglichen werden. Somit kann die Hypothese bewertet werden. Als dritter Sensor existiert ein Kraft/Momenten-Sensor, mit dessen Fingern die Breite eines Objekts gemessen werden kann. Damit können wiederum bereits bestehende Hypothesen über ein Objekt überprüft werden. In Abbildung 1.5 ist schematisch der Programmablauf dargestellt.

Für die zu erkennenden Objekte müssen entsprechende Modelldaten vorhanden sein. Diese Modelle bestehen jeweils aus einem Aspektgraphen [KD79]. Ein Knoten eines Aspektgraphen beschreibt eine Menge von Blickrichtungen auf das Modell. Aus den Blickrichtungen eines Knotens kann immer nur ein Aspekt des Modells betrachtet werden. Ein Aspekt besteht dabei aus einer Menge von Merkmalen des Modells, wie z.B. bestimmte Kanten oder Flächen. Die Kanten des Aspektgraphen verbinden benachbarte Aspekte, bzw. Ansichten des Modells. Mit Hilfe dieses Aspektgraphen können somit alle möglichen und prinzipiell unterschiedlichen Ansichten eines Modells erfasst und strukturiert werden.

Die vom Sensor erfassten Merkmale S_j eines Objekts werden nun den Modellmerkmalen f_j^i zugewiesen. Für n erfasste Merkmale S_1, \dots, S_n ergibt sich dadurch eine Menge von Kombinationen S_j/f_j^i :

$$\theta_i = \{S_1/f_1^i, S_2/f_2^i, \dots, S_n/f_n^i\}$$

θ_i ergibt somit eine Objekt-Hypothese.

Die Menge aller möglichen Hypothesen wird in der Dempster-Shafer-Theorie als frame of discernment Θ bezeichnet. Θ ist somit die Menge aller möglichen Objekthypothesen.

$$\Theta = \{\theta_1, \theta_2, \dots, \theta_n\}$$

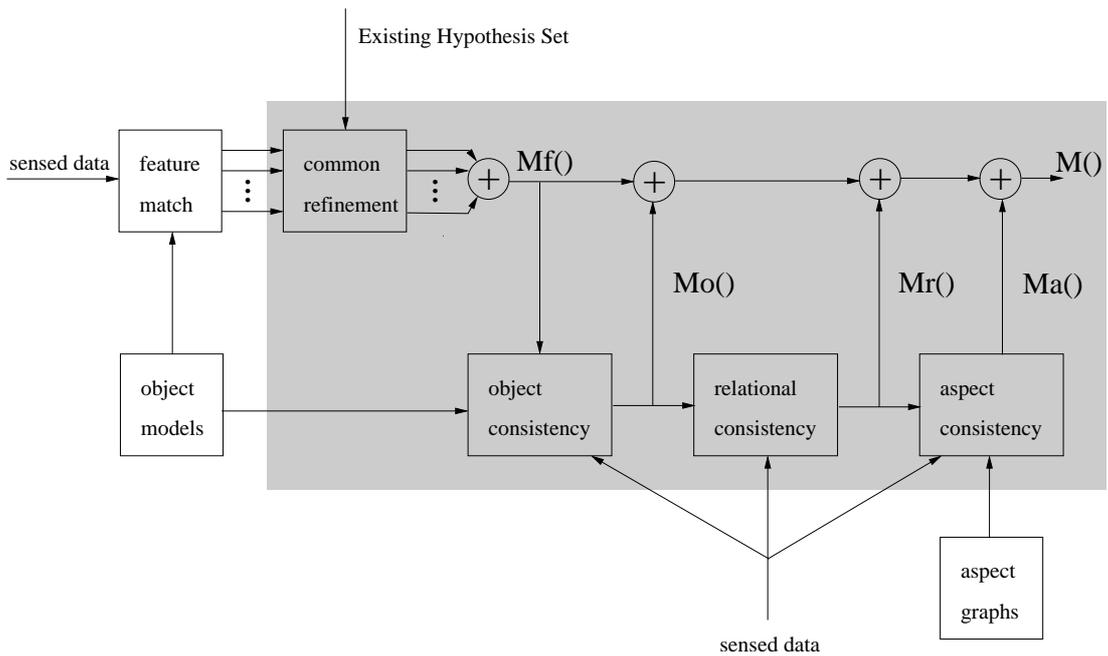


Abbildung 1.5: Blockdiagramm des Objekterkennungssystems. Die Objekthypothesenmengen $Mf()$, $Mo()$, $Mr()$ und $Ma()$ werden jeweils in verschiedenen Verarbeitungsstufen des Systems generiert und mittels der Dempster-Shafer-Methode kombiniert. Das Ergebnis dieser Verfeinerung $M()$ dient als Grundlage zur Berechnung der nächsten Sensoroperation. Der Fusionsteil ist grau unterlegt.

Das Maß für die Evidenz einer Hypothese wird in der Dempster-Shafer-Theorie als basic probability assignment (bpa) bezeichnet. Die Werte des bpa für die Hypothesen ergeben sich nun aus der Ähnlichkeit zwischen dem vom Sensor erfassten Merkmal und dem Modellmerkmal. Diese Ähnlichkeit kann z.B. bei einer Kante durch die Längendifferenz zwischen gemessener und Modellkante bestimmt werden. Das Ergebnis dieser Berechnungen ist in Abbildung 1.5 als Hypothesenmenge $Mf()$ bezeichnet. Dann folgen mehrere Konsistenzüberprüfungen. Damit wird die Anzahl der möglichen Hypothesen erheblich verringert. Bei jeder Überprüfung wird dabei der frame of discernment neu berechnet.

Bei der Prüfung auf **Objektkonsistenz** werden nur solche Objekthypothesen weiter berücksichtigt, welche nur aus konsistenten Entsprechungen bestehen. Eine inkonsistente Entsprechung besteht, falls zwei vom Sensor erfasste Merkmale eines Objekts, den Merkmalen von verschiedenen Modellen zugewiesen werden.

Beim Test auf **relationale Konsistenz** werden Ähnlichkeiten zwischen Beziehungen von sensorisch erfassten Merkmalen und Beziehungen zwischen Modellmerkmalen überprüft. Dabei kann z.B. getestet werden, ob der Winkel zwischen sensorisch erfassten Flächen den Modellannahmen entspricht.

Die letzte Überprüfung ist die **Aspektkonsistenz**. Für eine gegebene Objekthypothese kann mit Hilfe des Aspektgraphen eine mögliche Lage und Position des Objekts berechnet werden. Damit kann festgestellt werden, welche Objektmerkmale aus einer bestimmten Ansicht des Objekts erkannt werden sollen. Das zugehörige bpa berechnet sich dabei aus der Akkumulation positiver Evidenzen falls die erwarteten Merkmale in der Objekthypothese eine Entsprechung finden.

Anschließend soll die beste nächste Sensoroperation ausgewählt werden um die Position des Objekts genauer zu bestimmen. Aus der nach den Konsistenzprüfungen verbleibenden Hypothesenmenge $M()$ wird eine Objekthypothese θ_i ausgewählt. Mit dieser aus bestimmten Merkmalszuordnungen bestehenden Hypothese wird wiederum mit Hilfe des Aspektgraphen die entsprechende Objektansicht berechnet. Für jede Hypothese θ_i , ($i = 1, \dots, n$) aus $M()$ wird nun der Satz von Merkmalen, die sich aus der Objektansicht ergeben, vorhergesagt und die daraus resultierenden neuen Hypothesenmengen Θ_i berechnet.

Dann wird ein Maß A^{ij} für die Mehrdeutigkeit für jede der neuen Hypothesenmengen Θ_i berechnet. Der Index j bezieht sich dabei auf die momentane Objektansicht und damit auf die nächste Sensoroperation. Die Mehrdeutigkeit einer Hypothesenmenge ist am größten, wenn alle Hypothesen gleich wahrscheinlich sind. In diesem Fall steigt A^{ij} auch mit der Anzahl der Hypothesen. Die Mehrdeutigkeit sinkt, je mehr Evidenz auf eine Hypothese vereint werden kann. Die maximale Mehrdeutigkeit A_{max}^j aus A^{ij} wird der entsprechenden Sensoroperation zugewiesen. Mit allen Objekthypothesen θ_i wird entsprechend verfahren. Daraus ergibt sich eine Liste mit A_{max}^j . Die Sensoroperation, welche mit dem kleinsten Wert aus dieser Liste korrespondiert, wird als nächste Sensoroperation

ausgewählt.

Die mit dieser Methode vom Autor durchgeführten Experimente zeigen unter den gesetzten Rahmenbedingungen und der begrenzten Implementation brauchbare Ergebnisse. Leider ist nicht ersichtlich, wie sich das System unter komplexeren Umgebungsbedingungen und mit kompletter Implementierung verhält.

Kollisionsvermeidung

In [GZ96] wird versucht, mit Hilfe der Datenfusion eine effektive Kollisionsvermeidung bei Industrierobotern zu realisieren. Dabei geht es in erster Linie um Sicherheitsaspekte: Der Ausleger eines Roboters muss im Gefahrenfalle, d.h. falls sich Menschen oder andere Hindernisse im Aktionsradius des Roboters befinden, entweder seine Geschwindigkeit verlangsamen oder sofort anhalten. Der Autor betont dabei, dass das System unter Echtzeitbedingungen sicherheitsrelevante Entscheidungen treffen muss. Der Systemablauf ist schematisch in Abbildung 1.6 dargestellt. Zur Erkennung, ob sich ein Objekt im Bereich des Roboters aufhält, kommen verschiedenste Sensoren zum Einsatz. Dabei handelt es sich um Ultraschallsensoren, kapazitive Sensoren, Infrarotsensoren, Mikrowellensensoren, Drucksensoren, Lichtschranken oder Kameras. Diese verschiedenen Sensoren decken auch ganz verschiedene Bereiche in der Umgebung des Roboters ab. Jeder Sensor hat dabei seine eigene Landkarte. Diese Landkarten müssen zusammengefasst werden. Anschließend findet auf der gemeinsamen Landkarte die Detektion statt und danach muss eine Sicherheitsentscheidung getroffen werden. Das System gliedert sich somit in drei Teile: Datenintegration, Gefahrendetektion und Entscheidungsfindung.

In der Simulation dieser Methode wird die Umgebung des Roboters in etliche quadratische Felder aufgeteilt. Einzelne Sensoren überdecken dabei eine Anzahl Felder (siehe Abbildung 1.7).

Um nun die unterschiedlichen Sensorbereiche zu integrieren, würde sich laut Graham der Dempster-Shafer Algorithmus gut eignen. Er führt dabei mehrere Vorteile dieser Methode an:

“It has a mathematical rigor which allows it to be readily used with probabilistic estimates of sensory performance. It has a consistent rule for combination which is fully associative and commutative, so that the order of combination is irrelevant. It allows the inclusion of nonconclusive or ambiguous evidence (ignorance) without biasing the results. In short, this approach seems to match well with much of the physical intuition about sensory processing.”

Als Nachteil wird aber der potentiell hohe Rechenaufwand angeführt. Damit scheidet die Dempster-Shafer-Theorie als Fusionsalgorithmus für diese Echtzeit-

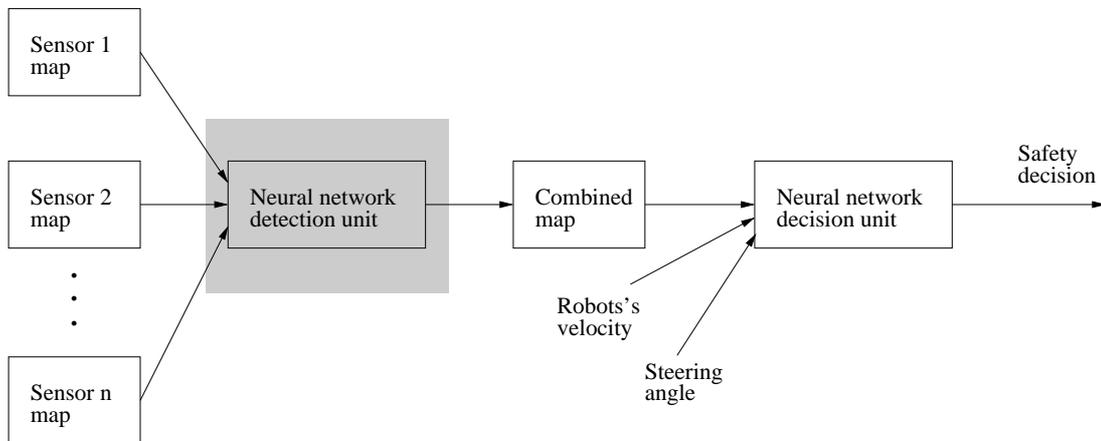


Abbildung 1.6: Schematische Darstellung des Kollisionsvermeidungssystems. In der grau unterlegten detection unit werden die Daten aus den verschiedenen Sensorkarten integriert. In der decision unit wird das Gefahrenpotential berechnet und eine Sicherheitsentscheidung getroffen.

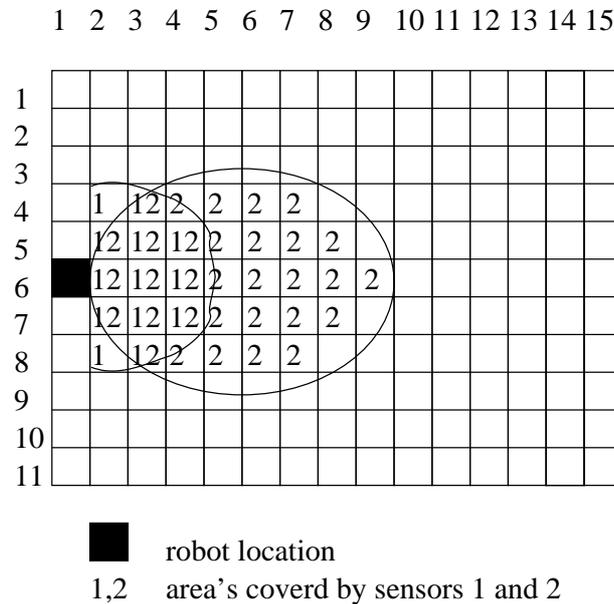


Abbildung 1.7: Der Arbeitsbereich eines Roboters wurde in 11x15 Felder unterteilt. Zwei Sensoren decken zwei sich überlappende Regionen des Arbeitsbereichs vor dem Roboter ab.

anwendung aus. Stattdessen werden nun Neuronale Netze verwendet, mit denen die Integrations- und Detektionseinheit sowie die Entscheidungseinheit realisiert werden. Bei den Neuronalen Netzen handelt es sich um feedforward-Netze mit einer back-propagation Trainingsmethode.

Die Integrations- und Detektionseinheit wird mit den Ergebnissen des Dempster-Shafer-Algorithmus trainiert. Jede Zelle des Umgebungsbereichs des Roboters bildet dabei eine Hypothese; somit bildet der ganze Umgebungsbereich den frame of discernment.

Die Entscheidungseinheit bekommt als Eingabe folgende Parameter:

1. Die Landkarte, auf denen sich die potentiellen Kollisionszonen befinden. Die Zonen sind dabei als belief-Werte in den Feldern der Landkarte kodiert.
2. Die Geschwindigkeit des Roboters.
3. Die Bewegungsrichtung des Roboters.

Die möglichen Ausgaben des Neuronalen Netzes lauten: weiterfahren, langsamer werden oder stopp. Grundlage der Entscheidung ist dabei das Skalarprodukt zweier Vektoren: Der Geschwindigkeitsvektor des Roboters und der aus dem Hindernis berechnete "virtual repulsive force vector". Dieser zeigt in Richtung des Roboters und wird umso stärker, je größer das Hindernis ist und je näher es sich am Roboter befindet. Trainiert wird die Entscheidungseinheit durch einen menschlichen Experten, der heuristisch die Schwellen für die Sicherheitsentscheidungen setzt.

In der Simulation wurden die beiden Neuronalen Netze zuerst getrennt getestet. Die Integrations- und Detektionseinheit liefert nach dem Training offensichtlich brauchbare Ergebnisse für die verwendeten Testfälle im Millisekundenbereich. Leider ist nicht ersichtlich wie lange der Dempster-Shafer-Algorithmus für die Berechnungen benötigt. Bei der Simulation für die Entscheidungseinheit wurde eine kleinere Landkarte gewählt (11*11 Felder). Das Training erfolgte mit zufälligen Werten für die Parameter: Roboterposition, Geschwindigkeitsvektor und Hindernisposition. Es ist nicht ersichtlich, wie der belief-Wert für die Position des Hindernisses berechnet wird. In 92% der Fälle erfolgte eine richtige Klassifikation.

Beim Test des gesamten Systems unter restriktiveren Rahmenbedingungen traf das System ebenfalls für 92% der Fälle richtige Entscheidungen. Durch unterschiedliche und teilweise stark eingeschränkte Randbedingungen in den verschiedenen Simulationen kann über die tatsächliche Leistungsfähigkeit des Systems keine klare Aussage getroffen werden.

Anlagensteuerung

Ein weiterer wichtiger Fall für die Anwendung der Datenfusion zeigt sich in der Kraftwerkstechnik. Hier sind viele verschiedene und redundante Sensoren zu finden, welche für den Betrieb des Kraftwerks von enormer Bedeutung sind. In [GA99] wird ein Fusionsalgorithmus beschrieben, um eine Gasturbine trotz veräuschter und divergierender Sensorwerte ständig mit bestem Wirkungsgrad zu betreiben. Der Wirkungsgrad ist am größten, wenn die Temperatur in der Brennkammer den maximal zulässigen Wert hat. Diese Temperatur kann nur indirekt gemessen werden und die Messungen sind etlichen Störfaktoren unterworfen, wie z.B. Änderungen der Umgebungstemperatur und des Umgebungsdrucks, des Kompressordrucks, Vibrationen, Korrosion, Materialfehlern, etc. Dadurch wird es notwendig die Daten aus gleichartigen und unterschiedlichen Sensoren, wie Temperatur- und Druckmesser, an verschiedenen Stellen der Anlage zu integrieren, um gesicherte Aussagen über den Systemzustand treffen zu können. Die Aufgaben des Algorithmus sind:

1. Redundante, aber unsichere, Sensorinformationen zu sammeln, um zu entscheiden, welche Sensorkanäle mit welcher Gewichtung in die Berechnung mit eingehen.
2. Rauschen und Fehlfunktionen der Sensoren auszufiltern, um das System trotz des Ausfalls eines oder mehrerer Sensoren stabil zu halten.

In Abbildung 1.8 ist der Fusionsalgorithmus schematisch dargestellt. Der Fusions- teil ist grau unterlegt. Die eingehenden Sensordaten werden zunächst auf Plausibilität geprüft und erhalten einen Konfidenzwert. Bei der Datenfusion werden die aktuellen Sensorwerte, der vorherige Systemzustand und die Ergebnisvorhersage aus der letzten Berechnung mitverwendet. z^{-1} deutet dabei einen Wert aus der vorhergehenden Berechnung an. Nach der Datenfusion wird der Ergebniswert an einen Controller weitergereicht und es ergibt sich ein neuer Systemzustand α , der wiederum Einfluss auf die nächste Berechnung hat. Der zugewiesene Konfidenzwert ist ein Maß für die Zuverlässigkeit des Messwerts und ist abhängig von der speziellen Sensorcharakteristik, dem vorhergesagten Wert und den physikalischen Grenzen des Sensorwerts. Damit wird sichergestellt, dass sinnlose Messwerte in der Berechnung nicht berücksichtigt werden. In Abbildung 1.9 wird die Vergabe des Konfidenzwerts deutlich. Liegt der Messwert z außerhalb dem Intervall $[v_l; v_r]$, welches die physikalischen Grenzen des Sensors darstellen, erhält der Messwert die Konfidenz 0 und geht nicht mit in die Fusion ein. Liegt der Messwert innerhalb des Intervalls, nimmt der Konfidenzwert Werte zwischen 0 und 1 an. Der Konfidenzwert wird umso größer, je näher der Messwert am vorhergesagten Wert \hat{x} liegt. Die Plausibilitätsfunktion $f(z)$ bildet ein adaptives Filter. Die Funktion ist asymmetrisch, dynamisch und von der Sensorzuverlässigkeit und den Umgebungsbedingungen abhängig. Damit lässt sich das unterschiedliche Reaktions-

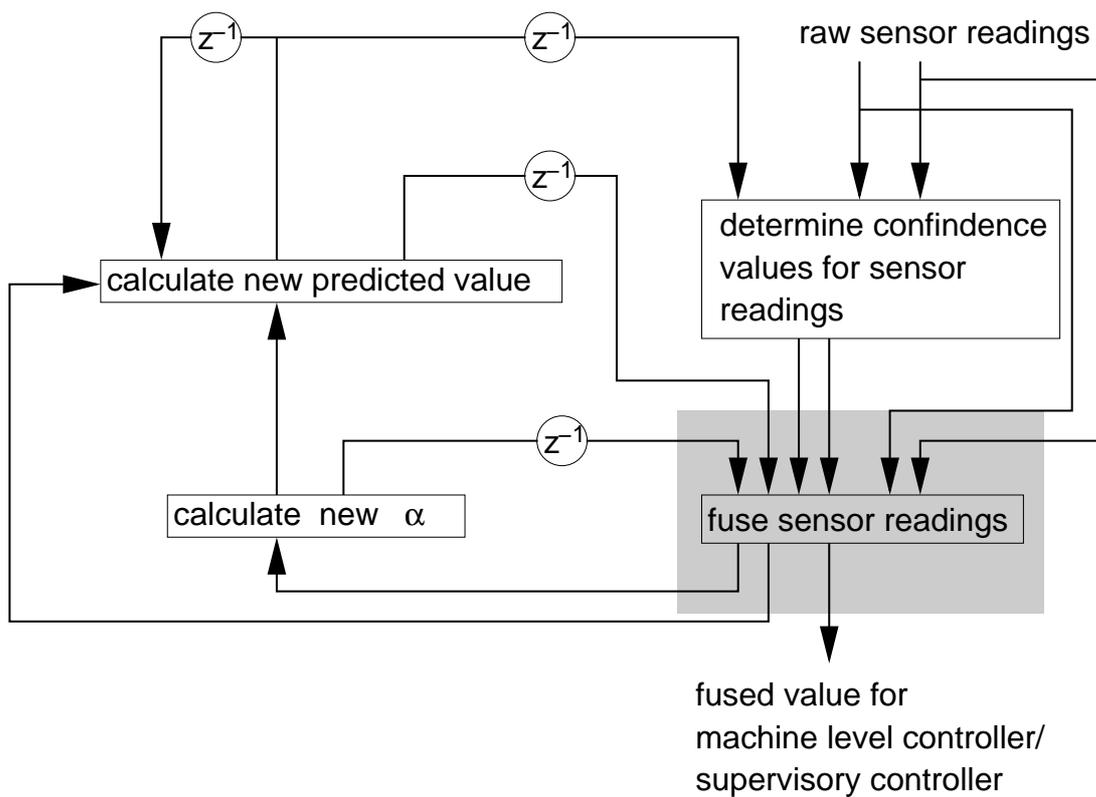
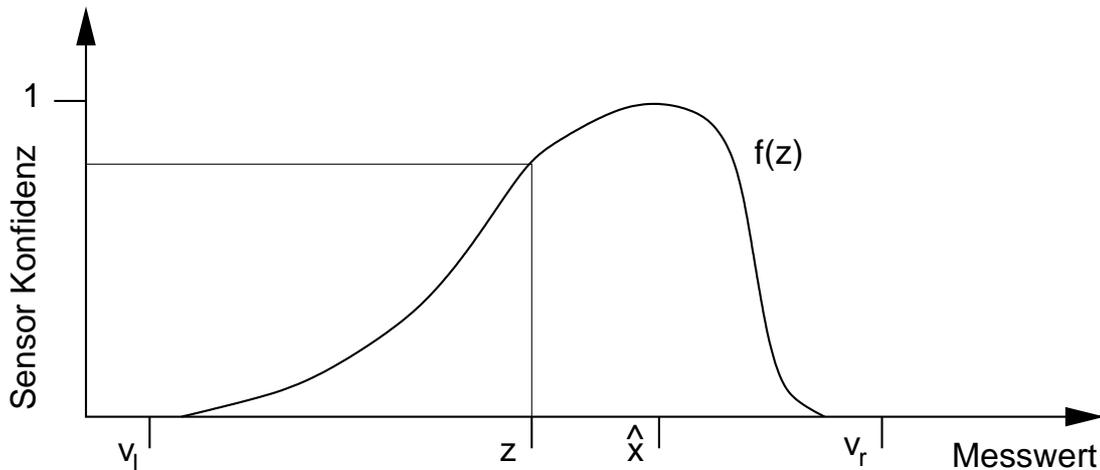


Abbildung 1.8: FUSVAF (Fuzzy Sensor validation and Fusion) Algorithmus für fuzzy Sensorplausibilitätsprüfung und -fusion. Die Sensordaten werden im grau unterlegten Fusionsteil integriert.



v_l, v_r physikalische untere und obere Grenzwerte für den Messwert
 \hat{x} vorhergesagter Wert
 z gemessener Wert

Abbildung 1.9: Plausibilitätsbereich für die Messwerte eines Sensors

verhalten eines Sensors über seinen ganzen Arbeitsbereich erfassen und entsprechend bewerten. Unzuverlässige Sensoren erhalten eine schmalere Kurvenform als zuverlässige Sensoren. Die Sensorcharakteristik kann aber auch z.B. durch Feuchtigkeits- oder Temperaturänderungen beeinflusst werden. Durch entsprechende Änderungen der Plausibilitätsfunktion können die Auswirkungen dieser Störfaktoren korrigiert werden.

Der fusionierte Wert berechnet sich aus einem gewichteten Mittel der Konfidenzwerte und den Messwerten nach der Formel:

$$\hat{x}_f = \frac{\sum_{i=1}^n z_i \sigma(z_i)}{\sum_{i=1}^n \sigma(z_i)}$$

Wobei \hat{x}_f der fusionierte Wert, z_i der Messwert und σ der Konfidenzwert ist.

Der Algorithmus wurde mit Daten von acht Sensoren einer realen Kraftwerksanlage getestet. Die Daten sind mit verschiedenen Rauscharten behaftet. Dabei zeigt sich z.B. ein periodisches, normalverteiltes Rauschen mit einem 24 Stunden Zyklus. Dies lässt sich auf unterschiedliche Temperaturen und Druckverhältnisse im Tagesablauf zurückführen. Störungen ergaben sich auch durch Quantisierungseffekte beim Runden der Messwerte. Hinzu kommen Sensorausfälle. Um trotzdem einen optimalen Betrieb zu gewährleisten und unnötige Abschaltungen des Kraftwerks zu vermeiden, wurde der vorher erklärte Algorithmus angewendet. Der Algorithmus soll eine genauere und robustere Kontrolle des Systems ermöglichen.

Zum Test wurden Sensorausfälle zu verschiedenen Zeitpunkten simuliert und die Stabilität des Algorithmus bewertet. Je nach Kombination und Zeitdauer der simulierten Sensorausfälle liefert der Algorithmus gute bis unbrauchbare Ergebnisse. Einerseits lieferte der Algorithmus bei fünf gleichzeitigen Sensorausfällen noch brauchbare Ergebnisse, andererseits führte der Ausfall von zwei anderen Sensoren bereits zu nicht verwertbaren Resultaten.

Robotik

In [HM99] wird eine Methode zur Datenfusion beschrieben, die sich nicht nur auf die Integration von Sensordaten beschränkt, sondern auch versucht, im System bereits vorhandene Information mit zu integrieren. Dadurch wird berücksichtigt, dass ein System nicht erst durch seine Sensorik zu Wissen gelangt, sondern dass ein System auch a priori Wissen über seine Umgebung hat. Grundlage dieses Multi-Sensor / Knowledge Fusion (MSKF) Ansatzes ist das Real-time Control System (RCS) [Alb97]. Der MSKF-Ansatz wird als Prozess definiert, der drei Forderungen erfüllt:

1. Selbständiges Sammeln von Informationen aus verschiedenen Sensoren.
2. Kombination der von den Sensoren gewonnenen Information mit a priori Wissen des Systems über seine Umgebung zu einer kohärenten Perzeption.
3. Effiziente Auswahl eines oder mehrerer Sensoren für eine bestimmte Aufgabe.

Zusätzliche Eigenschaften des Systems sollen Echtzeitverarbeitung, Erstellen oder Benutzen von Plänen für komplexe Aufgaben und Robustheit sein. Das System soll das von Albus [Alb97] klassifizierte reflexive und deliberative Verhalten eines Organismus integrieren. Mit reflexivem Verhalten ist dabei eine einfache und unverzügliche Reaktion auf einen Stimulus gemeint. Deliberatives Verhalten kommt durch komplexere Überlegungen zustande. Dabei werden internes Wissen, logische Schlüsse und zielgerichtete Strategien verwendet.

Die RCS-Architektur, siehe Abbildung 1.10, besteht in der Version 4 aus den Basiskomponenten: Value Judgement (VJ), Sensory Processing (SP), World Modelling (WM) und Behaviour Generating (BG). Hinzu kommen noch die Komponenten Knowledge Database (KD) und Operator Interface.

Das VJ-Modul enthält Funktionen, die ihre Entsprechung im limbischen System finden. Es werden also Berechnungen durchgeführt, welche die Kosten und den Nutzen, sowie das Risiko einer geplanten Aktion bewerten. Es sind Zustandsvariablen vorgesehen, die beschreiben, welche Ziele momentan am wichtigsten sind und welchen Regionen oder Objekten besondere Aufmerksamkeit gewidmet werden soll.

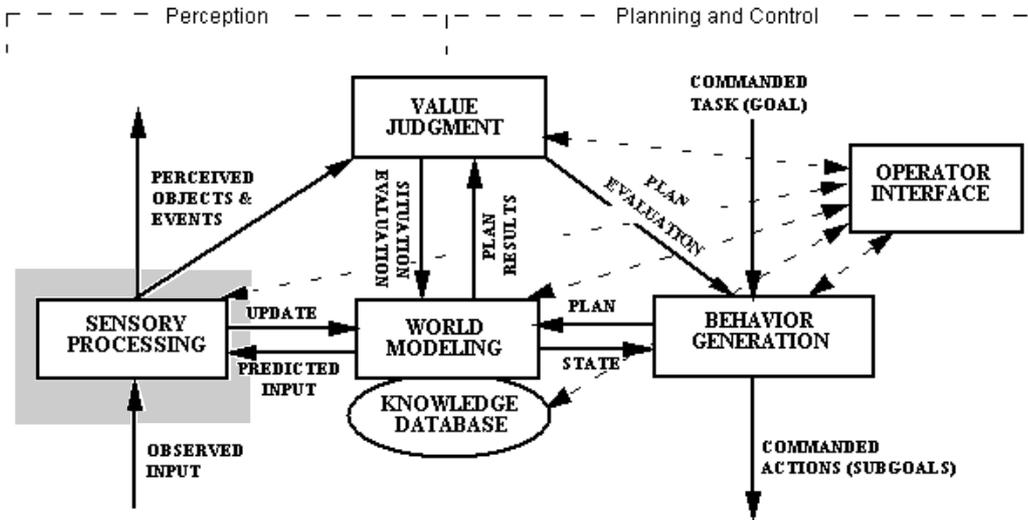


Abbildung 1.10: Basic Building Block des RCS-4 Modells. Im grau unterlegten Sensory Processing Modul werden die unterschiedlichen Sensordaten fusioniert.

Das SP-Modul erhält die Sensordaten. Es ist für die Filterung, Integration und Interpretation der Sensordaten zuständig. Das SP-Modul besteht aus den folgenden fünf Funktionen: Koordinatentransformation, Wertevergleich, zeitliche und räumliche Integration, sowie Objekterkennung bzw. Ereignisdetektion.

Das WM-Modul erfüllt für die Knowledge Database die Funktion eines Datenbank-Management-Systems und sorgt im KD-Modul für aktuellen und konsistenten Datenbestand.

Im BG-Modul findet die Planung statt. Dort wird das Hauptziel des Systems in Teilaufgaben zerlegt und zur Ausführung gebracht.

Abbildung 1.11 zeigt eine Beispielarchitektur des RCS-3 Modells. Bei diesem Modell ist das VJ-Modul noch nicht integriert. Es ist zu sehen, wie mit den Basic Building Blocks eine hierarchische Struktur aufgebaut werden kann. Dieses Beispiel zeigt verschiedene Granularitätsstufen eines Systems. Auf der rechten Seite sind Beispiele der funktionellen Charakteristik der BG-Module der einzelnen Ebenen zu sehen. Auf der linken Seite stehen Beispiele für die von den SP-Modulen erfassten Entitäten.

Um den Ansprüchen eines Multi-Sensor-Data-Fusion Systems gerecht zu werden, wurde die RCS-Architektur um drei Teile erweitert:

1. Innerhalb eines SP-Moduls können mehrere sensorische Systeme enthalten sein.
2. Im BG-Modul existieren mehrere Pläne zum Erledigen einer Aufgabe. Die

Kognitive Informatik

Im letzten Beispiel dieses Abschnitts wird gezeigt, wie die Datenfusion im Bereich der Kognitionswissenschaften behandelt wird. In [SSVW01] wird die biologische Bedeutung des Begriffs „Multisensory Integration“ erklärt: „Sensory systems have evolved to work in concert, and normally, different sensory cues that originate from the same event are concordant in both space and time.“ Der Autor stellt weiterhin fest, dass die sensorische Interaktion der durch räumlich und zeitlich kohärenten Stimuli angeregter, unterschiedlicher Modalitäten vermutlich zu einer Hervorhebung des auslösenden Ereignisses führt. Allerdings wird eingeräumt, dass momentan noch wenig über die integrativen Eigenschaften der multisensorischen Neurone im Kortex und ihren Einfluss auf die verschiedenen Aspekte der Wahrnehmung bekannt ist. Dieses Bild spiegelt sich auch in der Literatur zu kognitiven Systemen wieder. Dort lassen sich wenig Modelle und Verarbeitungsstrategien zur Datenfusion finden.

Eine theoretische Betrachtung zur Integration von Daten aus verschiedenen sensorischen Systemen ist in [PS99] zu finden. Beim Design von Roboteragenten kommen mehrere gleichartige oder verschiedene Systeme zum Einsatz. Damit können verschiedene Redundanzprinzipien realisiert werden. Durch überlegte, von der Aufgabe und der Umgebung des Agenten abhängige Anordnung und Auswahl der Sensoren, kann z.B. ein räumlicher Bereich mit verschiedenen Sensoren exploriert werden. Durch diese Überlappung kann die Unsicherheit der Information einzelner Sensorkanäle verringert werden. Eine Methode zur Kombination der Daten aus den unterschiedlichen Kanälen wird allerdings nicht erwähnt.

Eine Implementation eines Modells zur multisensorischen Integration findet sich in dem hierarchischen Weltmodell [Gar95]. Dieses Modell unterstützt die Einbindung von topologischen Informationen aus verschiedenen Sensorkanälen auf verschiedenen Abstraktionsstufen. Die hierarchische Weltrepräsentation besteht aus einer hierarchischen Anordnung von begrenzten Quadern (oriented bounding box, OBB) mit einer bestimmten Orientierung. Jede OBB wird durch eine größere OBB eingeschlossen. Dadurch entsteht eine Eltern-Kind-Beziehung, die durch eine baumartige Struktur dargestellt wird. Die Wurzel dieses Baumes repräsentiert dabei den ganzen vom Modell erfassten Raum. Die OBB besteht dabei aus folgenden vier Elementen:

1. Eine allgemeine Beschreibung der OBB (Identifikator, Ausmaße, Koordinaten, Orientierung, usw.)
2. Eine Liste der Nachfolgenden OBBs (Kinder)
3. Eine zur OBB gehörende Oberfläche
4. Globale Attribute über die Inhalte der OBB (Volumen, Seitenverhältnisse,

Temperatur, Farbe, Prozeduren zum Bewegen eines Objektes, usw.)

Die Oberflächenbeschreibung besteht aus einem trianguliertem Gitter von sogenannten Kontrollpunkten. Jeder Punkt ist definiert durch seine räumlichen Koordinaten und einer Liste von skalaren Variablen (Temperatur, Druck, usw.). Den Koordinaten und Variablen sind Konfidenzwerte zugeordnet. Diese drücken die Zuverlässigkeit der Information aus. Die Konfidenzwerte sind von der Zuverlässigkeit und Genauigkeit der Sensoren abhängig, welche den entsprechenden Kontrollpunkt generiert haben. In dieser Oberflächenbeschreibung sind somit in einer einzigen Repräsentation alle von den Sensoren erfassten Daten zu einem bestimmten Objekt enthalten. Die Aufgabe besteht nun darin, aus den Kontrollpunkten die geometrische Oberfläche des entsprechenden Objekts und aus den Variablen die Eigenschaften des Objektes zu schätzen. Die zu berechnende Oberfläche besteht aus einer Menge von Evaluierungspunkten $P = (x_i, y_i, z_i)$, die durch einen nur sehr oberflächlich erläuterten Approximationsalgorithmus angenähert werden. Die Berechnung erfolgt mit Hilfe eines triangulierten Gitters, welches aus den Kontrollpunkten besteht. Dabei werden zur Berechnung eines Evaluierungspunktes innerhalb eines Dreiecks nicht nur dessen Eckpunkte verwendet, sondern auch die Kontrollpunkte der angrenzenden Dreiecke mitsamt deren Gewichtung. Dadurch kommt eine stückweise „glatte“ Oberfläche zustande. Die Feinheit der Oberfläche ist von der Anzahl der Evaluierungspunkte abhängig. Durch den Gewichtungsfaktor, der mit jedem Kontrollpunkt assoziiert ist, wird die Struktur der Oberfläche ebenfalls beeinflusst. Da der Gewichtungsfaktor von der Sensorcharakteristik abhängig ist, beeinflusst diese somit direkt die Rekonstruktion der Oberfläche.

Zum Abschluss dieses Abschnitts werden einige mögliche Anwendungen der Datenfusion in technischen Systemen und ihre Methoden in Tabelle 1.1 dargestellt. In den Zellen finden sich einige Beispielanwendungen. Die Methoden werden in Neuronale Netze (NN), Fuzzy-Set-Theorie (FUZZY), Dempster-Shafer Theorie (DS) und in die klassische Wahrscheinlichkeitsrechnung (BAYES) unterteilt.

1.3 Datenfusion in biologischen Systemen

In diesem Abschnitt soll verdeutlicht werden, auf welche Art und Weise Rückschlüsse auf die Informationsintegration in biologischen Systemen gewonnen werden können. Dabei werden wichtige Erkenntnisse über die Datenfusion im Gehirn vorgestellt. Eine ausführliche Behandlung multisensorischer Prozesse ist in [CSS04] zu finden.

Anwendungsgebiete	Methoden			
	NN	FUZZY	DS	BAYES
Robotik			[Mur98]	[PS99] [HM99]
Anlagensteuerung	[PRD98] [NSSW97]	[GA99] [PRD98]		[GA99]
Kollisionsvermeidung	[GZ96]		([GZ96])	
Bildsequenzanalyse	[Hun92]			
Objekterkennung			[HK92]	[Gar95]
Geografische Informationssysteme	[BGM00]	[BLM+98] [BGM00]	[BLM+98] [BGM00]	
Mustererkennung		[Den00]	[Den00]	

Tabelle 1.1: Übersicht von Anwendungen und Methoden der Datenfusion in technischen Systemen

Neurophysiologische Untersuchungen

Neurologische Untersuchungen an Tieren liefern aufschlussreiche Erkenntnisse über die neuronale Informationsverknüpfung aus unterschiedlichen Modalitäten. Durch Einzelzelleableitungen an Säugetieren [SMW94] wurden verschiedene multimodale Neurone im Superior Colliculus (SC) gefunden. Bestimmte Neurone zeigten bei bimodalen Stimuli eine drastisch erhöhte Aktivität gegenüber unimodaler Reizung. Andere Neurone hingegen verminderten ihre Aktivität. Bimodale Neurone können damit das Signal bei bimodaler Reizung wesentlich verstärken. Durch systematische Untersuchungen wurde festgestellt, dass die prozentuale Verstärkung umso stärker ist, je schwächer die Signalstärke bei unimodaler Reizung ist. Aus den Untersuchungen ergaben sich die sogenannten integrativen Regeln, mit deren Hilfe multisensorische Neurone beschrieben werden. Da diese Regeln auch für eine spätere Modellierung biologischer Systeme von Bedeutung sind, werden sie nun kurz vorgestellt.

- Spatial rule

Im SC existieren für mehrere Sinnessysteme, wie z.B. den auditiven, visuellen und somatosensorischen Sinn, jeweils eine eigene räumliche Karte, welche den jeweiligen sensorischen Raum abbilden. Die Koordinatensysteme dieser Karten sind einander sehr ähnlich. Diese angeglichenen Karten werden deshalb zusammen als eine multisensorische Karte betrachtet. Räumlich korrespondierende Neurone der verschiedenen Karten leiten ihr Signal an das rezeptive Feld eines multisensorischen Neurons weiter. Falls nun Stimuli aus verschiedenen Modalitäten, aber vom gleichen Ort im Raum

stammen, wird so das exitatorische rezeptive Feld des multisensorischen Neurons aktiviert. Dies führt zu einer multiplikativen Signalzunahme dieses Neurons. Im Gegensatz dazu wird das inhibitorische Feld des multisensorischen Neurons aktiviert, falls die Stimuli sich räumlich von einander entfernen. Dies führt zu einer Signalabnahme des multisensorischen Neurons [MS86]. Diese von der räumlichen Beziehung der Stimuli aus unterschiedlichen Modalitäten abhängigen dynamischen Eigenschaften, werden unter dem Begriff *Spatial rule* zusammengefasst.

- Temporal rule

Die neuronalen Antworten auf die verschiedenen Stimuli eines Objekts stehen im SC nicht gleichzeitig zur Verfügung. Dies liegt zum einen an den unterschiedlichen Laufzeiten der physikalischen Reize vom Objekt zum Rezeptor, zum anderen an der unterschiedlichen Verarbeitungszeit der Reize vom Rezeptor zum SC. Durch diese beiden Faktoren entstehen unterschiedlichen Latenzen der Stimuli. Somit treffen die Signale aus den verschiedenen Modalitäten nicht gleichzeitig im SC ein. Dieser Umstand scheint eine Interaktion der Signale zu erschweren. Da die exitatorischen und inhibitorischen Ereignisse aber meistens länger anhalten, als die unterschiedlichen Latenzen dauern, entsteht ein sogenanntes interaktives Fenster für die Neurone im SC [MNS87]. Maximale multisensorische Interaktion ergibt sich, wenn sich die Spitzen der unimodalen Entladungssakkaden überlappen. Dies ist der Fall, wenn der zeitliche Abstand der Stimuli nicht mehr als 50-100 Millisekunden beträgt. Zur Veranschaulichung sind aus [PY02] in Abbildung 1.12 die Entladungssakkaden eines bimodalen Neurons gezeigt. Die Abbildung zeigt die aus einer Einzelzelleitung aus dem Superior Colliculus Inferior bei der Katze gewonnenen Daten. Auf der Abszisse bezeichnet der Zeitpunkt 0 den Beginn des jeweiligen Stimulus. Auf der Ordinate sind die kumulierten Antworten eines Neurons aufgetragen. Die Daten sind auf die Anzahl der Durchgänge pro Bedingung (auditiv, visuell, bimodal) normiert. Das bimodale Neuron zeigt auf den audio-visuellen Reiz eine deutlich stärkere Antwort als auf die unimodalen Reize. Würde man nun die beiden Stimuli zeitlich versetzt präsentieren, würde bei genügend großem zeitlichen Abstand der Verstärkungseffekt bei der bimodalen Reizung verschwinden.

- Magnitude, oder inverse effectiveness rule

Die Stärke der multisensorischen Interaktion wird nach folgender Formel berechnet:

$$\frac{100 * (CM - SM_{max})}{SM_{max}}$$

Dabei steht CM für die multimodale Antwort und SM_{max} für die stärkste unimodale Antwort (in Entladungen/Trial). Das Ergebnis gibt die prozentuale Verstärkung an. Bei Experimenten ist zu beobachten, dass die

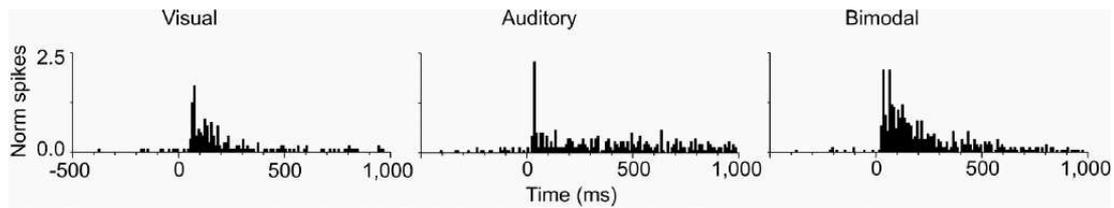


Abbildung 1.12: Normierte Spikerate eines bimodalen Neurons auf einen visuellen, auditiven und bimodalen Reiz.

prozentuale Verstärkung dann besonders hoch ist, wenn die unimodalen Antworten schwach sind. Umgekehrt ergibt sich nur eine geringe prozentuale Verstärkung, wenn die unimodalen Antworten hoch sind. Dies ist insoweit sinnvoll, da deutliche unimodale Reize keine Verstärkung erfordern. Bei schwachen Kombinationen unimodaler Reize hingegen, ist eine hohe Verstärkung von großem Nutzen.

- Receptive field preservation rule

Jedes multisensorische Neuron besitzt für jede Modalität ein eigenes rezeptives Feld. Die unimodalen Eigenschaften des rezeptiven Feldes eines multisensorischen Neurons im SC verändern sich nicht durch einen Stimulus in einer anderen Modalität.

Psychophysische Methoden

Bei der multisensorischen Forschung kann zwischen verschiedenen methodischen Ansätzen unterschieden werden. Sehr gängig sind, wohl auch historisch bedingt, psychophysische Methoden. Dabei wird versucht, einen quantitativen Zusammenhang zwischen einer physikalischen Reizgröße und der dazu gehörenden Sinnesempfindung zu ermitteln. Um zu untersuchen, wie sich mehrere Modalitäten gegenseitig beeinflussen, kann man z.B. die Wahrnehmung in einer Modalität verzerren. Um beispielsweise die Wahrnehmung des visuellen Systems zu verzerren, werden Linsen, Prismen oder Spiegel vor die Augen gesetzt, die dann das visuelle Feld verzerren, kippen oder invertieren. Über Kopfhörer kann die auditive Wahrnehmung beeinflusst werden und mit force-feedback-Geräten kann die haptische Wahrnehmung verändert werden. Somit kann ein Objekt in den verschiedenen Modalitäten z.B. an unterschiedlichen Orten oder in unterschiedlicher Größe präsentiert werden. Durch die Interaktion der verschiedenen Modalitäten kommt es zu einer gegenseitigen Beeinflussung (Bias). In [WW80] findet sich ein Überblick zu den drei klassischen Bias-Theorien, die im Folgenden kurz vorgestellt werden:

- Modality Precision Hypothese

Diese Theorie besagt, dass bei der Wahrnehmung eines Ereignisses, derjenigen Modalität der Vorzug gegeben wird, die das Ereignis am präzisesten erfassen kann. Diese Theorie wird vor allem durch die Ergebnisse von Lokalisationsaufgaben unterstützt.

- Directed-Attention Hypothese

Nach dieser Theorie kommt der intersensorische Bias durch die unterschiedliche Aufteilung an Aufmerksamkeit an die sensorischen Modalitäten zustande. Die dominante Modalität ist demnach diejenige, welcher mehr Aufmerksamkeit zugeteilt wird. Innerhalb dieser Theorie lassen sich zwei Fälle unterscheiden: zum einen kann die unterschiedliche Zuteilung der Aufmerksamkeit durch die Art des Stimulus ausgelöst werden, z.B. durch plötzliches Auftauchen des Stimulus. Zum anderen kann die Zuteilung der Aufmerksamkeit auch durch die Prädisposition des Beobachters beeinflusst werden, wenn z.B. der Beobachter ausdrücklich aufgefordert wird, nur auf eine Modalität zu achten.

- Modality Appropriateness Hypothese

Bei dieser Theorie wird davon ausgegangen, dass jede einzelne sensorische Modalität eine oder mehrere Funktionen besser erfüllen kann als eine andere Modalität. Dadurch wird bei komplexen Aufgaben diejenige Modalität bevorzugt, die für die Aufgabe am besten geeignet ist. Für räumliche Aufgaben ist demnach das visuelle System besser geeignet, für zeitliche Aufgaben wird dem auditiven System der Vorzug gegeben.

Beim Betrachten dieser drei Theorien wird deutlich, dass sie sich nicht gegenseitig ausschließen, sondern vielmehr ergänzen. Jede einzelne Hypothese betrachtet einen anderen Teilaspekt der Wahrnehmung. Die Modality Precision Hypothese kann gewissermaßen als Untermenge der Modality Appropriateness Hypothese betrachtet werden. Auch in der Summe ergeben die drei Theorien keine umfassende Beschreibung der gegenseitigen Beeinflussung von verschiedenen Modalitäten. In [WW80] wird ein weiteres Modell vorgeschlagen, welches die Annahme zur Grundlage hat, dass das Wahrnehmungssystem dazu neigt, die verschiedenen sensorischen Eindrücke einem einzelnen, einheitlichen Ereignis zuzuschreiben. Davon ausgehend werden vier Stufen definiert, in denen die verschiedenen Faktoren postuliert werden, die bei der Wahrnehmung einer widersprüchlichen Stimuluskombination eine Rolle spielen:

In der **Stimulus Situation** werden die experimentellen Bedingungen und die Eigenschaften der Stimuli wie Form, Intensität, Bewegung, usw., festgehalten. Bei der **Modality Characteristics** wird beschrieben, wie die Stimuli durch die verschiedenen Modalitäten verarbeitet werden. In der Stufe **Observer Processes** wird Vorwissen des Beobachters zur Wahrnehmung eines Ereignisses in die Betrachtung mit einbezogen. Hier werden die Annahmen des Beobachters über

die Einheit eines Ereignisses und Aufmerksamkeitsprozesse berücksichtigt. Die letzte Stufe des Modells bildet das **Perceptual Result**. Hier kommt schließlich der Kompromiss über das wahrgenommene Ereignis zustande. Das Ergebnis ist sowohl von der Diskrepanz der Information aus den verschiedenen Modalitäten, als auch von den Annahmen des Beobachters abhängig. Bei kleinem Bias ist die Wahrscheinlichkeit groß, dass ein einziges, einheitliches Ereignis wahrgenommen wird. Bei großen Unstimmigkeiten ist eher zu erwarten, dass der Beobachter die Wahrnehmung aus den einzelnen Modalitäten verschiedenen Ereignissen zuordnet.

Bildgebende Verfahren

Zunehmend werden zur Erfassung sensorischer Interaktion auch bildgebende Verfahren eingesetzt. Damit lassen sich in erster Linie räumliche Zusammenhänge erfassen. In einer FMRI-Studie [BHI⁺03] wurde ein Experiment durchgeführt, bei dem sich zwei visuelle Objekte aufeinander zu bewegen, sich überlappen und ihren Weg fortsetzen. Normalerweise entsteht der Eindruck, dass die Objekte übereinander gleiten. Wird nun beim Kontakt der beiden Objekte ein auditives Kollisionsgeräusch erzeugt, entsteht vermehrt der Eindruck, dass die Objekte aneinander abprallen. Für die verschiedenen Bedingungen kann nun jeweils ein FMRI-Bild erzeugt werden. Dabei wird deutlich, dass im Kollisionsfall deutlich vermehrte Aktivität in multimodalen Regionen zu beobachten ist. Diese erhöhte Aktivität ist unter anderem auch im Superior Colliculus festzustellen. Diese Region ist als Verbindungsstelle von visueller und auditiver Information zur Blickrichtungsteuerung, und damit zur Objekterfassung, bekannt.

In einer weiteren FMRI-Studie [CHIB01] wurde ebenfalls die audio-visuelle Interaktion untersucht. Die Präsentation der Stimuli erfolgte unimodal, als auch synchron und asynchron bimodal. Dadurch konnten nicht nur Regionen mit vermehrter Aktivität bei gleichzeitiger bimodaler Stimulierung festgestellt werden, sondern es wurden auch Regionen gefunden, die bei bimodaler asynchroner Reizung eine deutlich verminderte Aktivierung (Response Depression) zeigen. Bei Gehirnregionen die bei synchroner Stimulierung hohe Aktivität zeigen und bei asynchroner Aktivierung verminderte Aktivität, handelt es sich demnach um multimodale Areale, die für die zeitliche Integration multisensorischer Wahrnehmungen verantwortlich sind. Hierfür kommt vor allem der Superior Colliculus in Betracht. In den Kortexregionen fallen vor allem Bereiche in der rechten Gehirnhälfte auf, wie z.B. der inferior frontale Gyrus, der laterale Sulcus sowie der superiore parietale Lappen und inferiore parietale Sulcus. Es finden sich aber auch Regionen in der linken Gehirnhälfte (superior temporale Sulcus und superior occipitale Gyrus) und auch beidseitige Reaktionen in der Insula. Etliche dieser Areale werden, wie nachfolgend gezeigt, durch elektrophysiologische Un-

tersuchungen unterstützt.

Elektrophysiologische Untersuchungen

Zur Erforschung multimodaler Zusammenhänge werden auch evozierte Potentiale mit Hilfe des EEG gemessen. Im Gegensatz zu FMRI-Studien lassen sich mit EEG-Ableitungen zeitliche Zusammenhänge besser erfassen. Allerdings beschränken sich die EEG-Ergebnisse auf die Kortexregionen, da mit diesem Verfahren tieferliegende Strukturen, wie z.B. der Superior Colliculus nicht erfasst werden.

In [FMM⁺00] werden ereigniskorrelierte Potentiale (ERP's) gemessen und eine auditorisch-somatosensorische Interaktion festgestellt. Der auditive Stimulus bestand aus einem kurzen 1000 Hz Ton und der somatosensorische Stimulus aus einer kurzen elektrischen Reizung des nervus medianus. Die Aufnahme der ERP's erfolgte mit 64 Elektroden. Die Probanden wurden angehalten, nicht auf die Reize zu achten, sondern ihre Aufmerksamkeit auf ein vor ihnen präsentiertes Bild zu richten. Dadurch wurde versucht, störende Augenbewegungen zu minimieren. Aufnahmen mit Blinzeln und starken Augenbewegungen, sowie starken EMG-Ausschlägen wurden in der Auswertung nicht berücksichtigt. Die verschiedenen Reize waren nur auditiv, nur somatosensorisch und gleichzeitig auditorisch-somatosensorisch. Gemessen wurde zum einen die mittlere Spannungsänderung an den einzelnen Ableitungspunkten und die Stromdichteverteilung der Kopfhaut (sculp current density (SCD)) für die drei verschiedenen Bedingungen. Für den Vergleich der mittleren Spannungsänderung wurden die Messwerte der unimodalen Messungen aufsummiert und mit den bimodalen Messungen verglichen. Multisensorische Integration wurde dabei als ein signifikanter Unterschied zwischen beiden Messungen definiert. Dabei wird davon ausgegangen, dass sich die Antworten der unimodalen Komponenten linear summieren, solange sie von den unisensorischen, neuronalen Strukturen verarbeitet werden. Wenn die Antworten der bimodalen Messung diese lineare Annahme übertreffen, ist nach den Autoren der begründete Schluss zulässig, dass ein multisensorischer Bereich aktiviert wurde und eine Integration der auditiven und somatosensorischen Reize stattfindet. Abbildung 1.13 zeigt das Ergebnis einer Elektrode. Die Kurve mit dem größten Maximum zeigt das Ergebnis der bimodalen Messung. Die andere durchgezogene Kurve zeigt das aufsummierte Signal der unimodalen Messungen. Die gestrichelte Kurve zeigt das Differenzbild. Es wurden vier Bereiche (40-49, 50-59, 60-69, 70-79 ms) untersucht. Die drei letzten Bereiche zeigen einen signifikanten Unterschied.

Die SCD wird aus der zweiten räumlichen Ableitung der gemessenen Spannungen an allen 64 Elektroden berechnet. Die Auswertung der SCD erfolgt ebenfalls nach der Differenzbild-Methode. Dabei wird wiederum der Unterschied zwischen der Summe der unimodalen Ergebnisse und des bimodalen Ergebnisses berechnet.

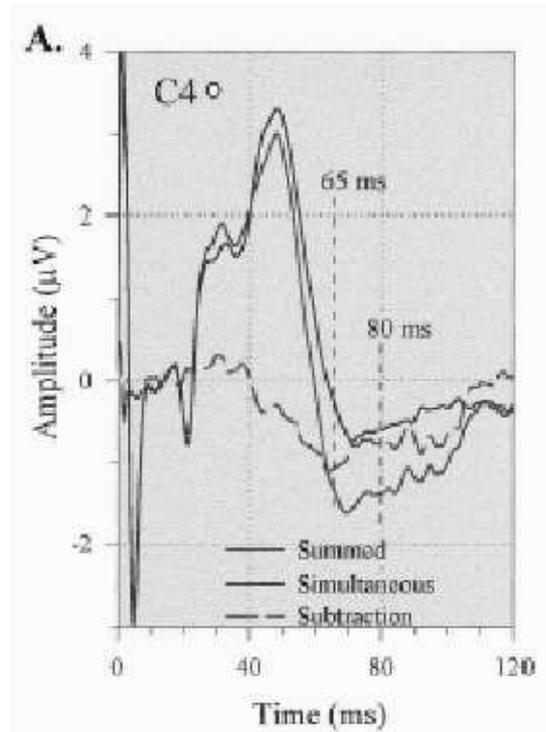


Abbildung 1.13: Gemittelter Spannungsverlauf einer rechts-zentralen Elektrode (C4). Dargestellt sind die summierten und die bimodalen Spannungsverläufe. Die gestrichelte Linie zeigt das Differenzbild.

Abbildung 1.14 zeigt zwei Differenzbilder zu den Zeitpunkten 65 ms und 80 ms. Der runde Punkt zeigt die Position der Elektrode C4. Nach 65 ms zeigt sich eine negative Dichte im rechten, lateralen, zentralen Kopfbereich. Diese Verteilung ist weitgehend übereinstimmend mit Interaktionen in somatosensorischen Hirnarealen. Bei 80 ms hat sich dieser Bereich nach anterior verschoben und überdeckt nun ein anterior, temporal gelegenes auditorisches Hirnareal.

Diese Ergebnisse deuten auf eine auditorisch-somatosensorische Integration in der frühen kortikalen Verarbeitung dieser Signale hin.

In einer anderen elektrophysiologischen Studie wurde die audio-visuelle Integration beim Menschen während einer Objekterkennungsaufgabe untersucht [GP99]. Gemessen wurden zuerst die Reaktionszeiten für die Erkennung der Objekte in den verschiedenen Modalitäten. Die Probanden waren im Durchschnitt 64 ms schneller bei der Erkennung eines bimodal präsentierten Objekts, gegenüber der nur auditiven oder nur visuellen Präsentation. Bei der Betrachtung der unimodalen Reaktionszeiten stellte sich heraus, dass sich die Probanden in zwei Gruppen einteilen lassen: Die eine Gruppe (VIS-Personen) war schneller in der Identifikation der visuell präsentierten Objekte, die andere Gruppe (AUD-Personen)

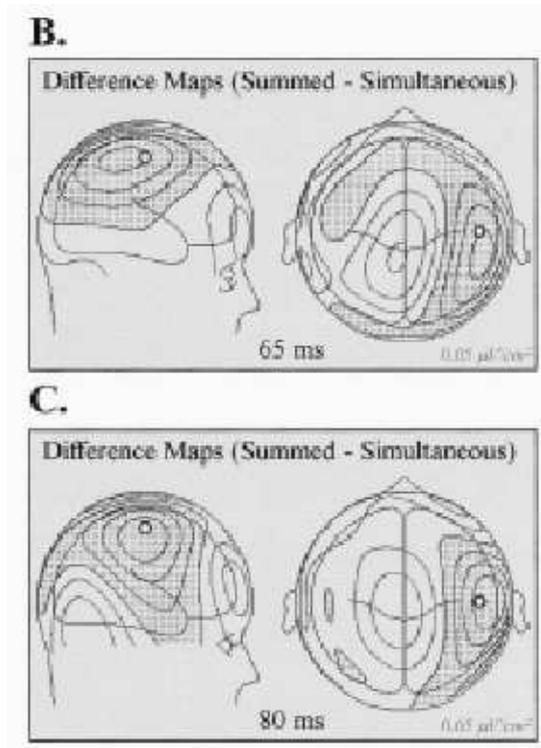


Abbildung 1.14: Stromdichteverteilung auf der Kopfhaut (SCD). Abbildung B zeigt das Differenzbild nach 65 ms. Abbildung C zeigt die Verteilung nach 80 ms. Die dunklen Bereiche zeigen erhöhte negative Werte.

war schneller, die auditiv präsentierten Objekte zu erkennen. Dementsprechend erfolgte die Analyse des SCD auch getrennt nach AUD- und VIS-Personen. Abbildung 1.15 zeigt die nach Gruppen getrennte Analyse der Differenzbilder. Eine weitere Untergliederung der Ergebnisse erfolgte nach räumlichen (occipital, temporal-central, rechts fronto-temporal) und zeitlichen (a-f) Gesichtspunkten. In den jeweils linken Abbildungen der zwei Gruppen sind die Potentialverteilungen der Interaktionsmuster zu sehen. Jeweils rechts daneben zeigt der t-Test signifikante Interaktionsmuster.

Demzufolge zeigt sich in der frühen Verarbeitung im visuellen Kortex (40-145 ms, Zeilen a und b) ein Interaktionsmuster nur für die AUD-Gruppe. In der VIS-Gruppe hingegen ist nur eine sehr geringe Interaktion zu finden. Für den Bereich zwischen 155 und 200 ms (c) zeigt sich ein ähnliches Interaktionsmuster für beide Gruppen. Für diesen Zeitbereich sind die Unterschiede in den Interaktionsmuster nicht signifikant.

Im temporalen/zentralen Bereich zeigt sich zwischen 90 ms und 105 ms (d) nur für die VIS-Gruppe ein Interaktionseffekt im Bereich der supratemporalen Ebene des auditiven Kortex. Für die zentralen Bereiche (e) sind die Interaktionsmuster

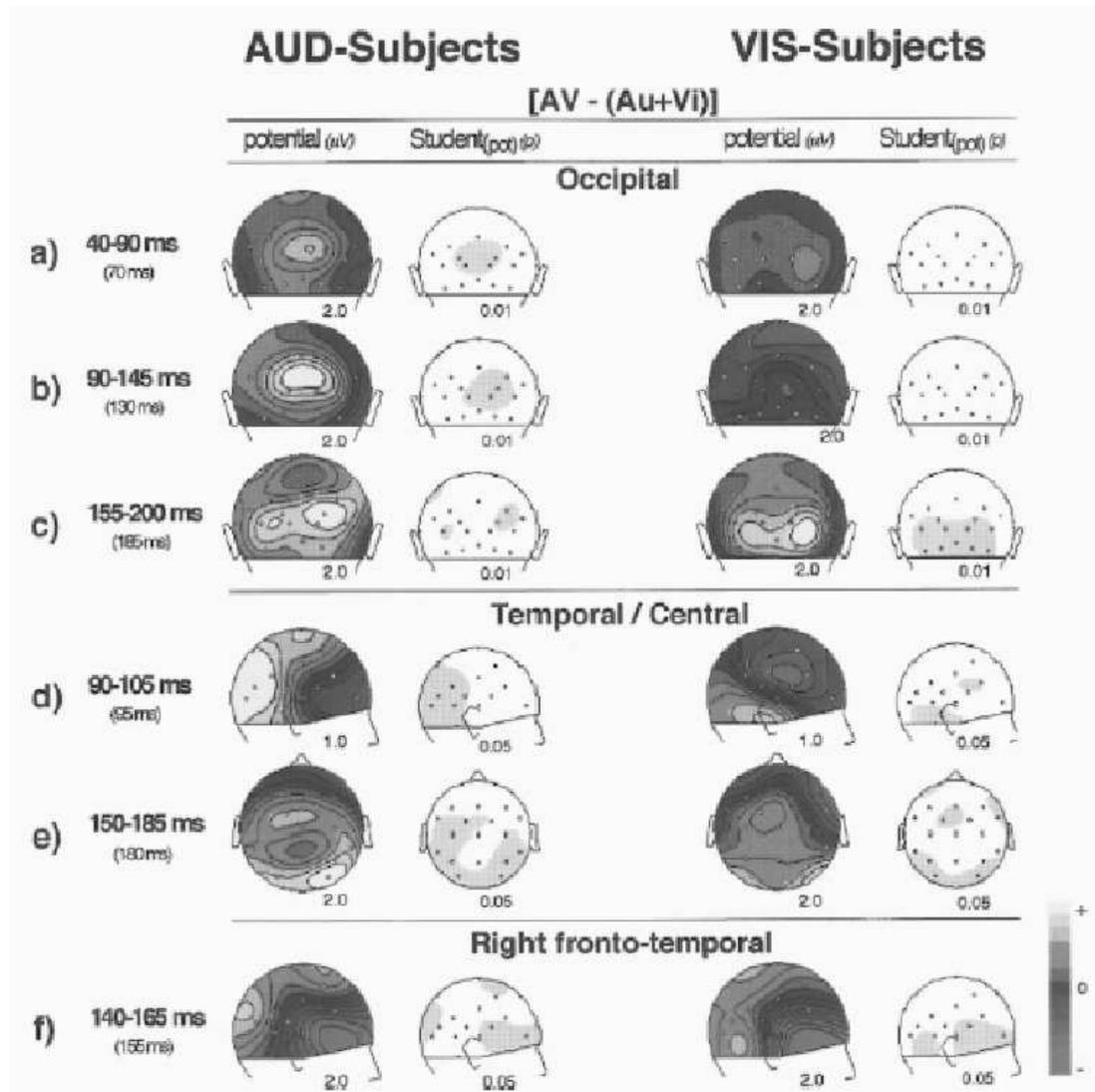


Abbildung 1.15: Differenzbilder der SCD und Darstellung der signifikant unterschiedlichen Bereiche. Die Einteilung erfolgt nach AUD- und VIS-Gruppe, sowie nach zeitlichen und räumlichen Aspekten.

in beiden Gruppen einander sehr ähnlich zwischen 150 ms und 185 ms.

Die Interaktionsmuster für die rechts fronto-temporalen Bereiche (f) sind sich ebenfalls sehr ähnlich für beide Gruppen und konzentrieren sich im anterioren Teil des Temporallappens.

In einer späteren audio-visuellen ERP-Studie [MRM⁺02] konnten trotz eines anderen Versuchsparadigmas, sowohl die Orte als auch die Zeiten der Interaktionsmuster bestätigt werden. Es fanden sich ebenfalls frühe Interaktionen (ca. 40-50 ms) im occipitalen Bereich, Interaktion mit mittlerer Latenz (ca. 165 ms) im Temporalbereich, als auch Interaktionsmuster bei ca. 185 ms im zentralen Bereich.

Während der frühen kortikalen Verarbeitung (unter 150 ms) bestehen in Abbildung 1.15 auffällige Unterschiede in den beiden Gruppen. Bei der VIS-Gruppe (Beobachter, die bei der unimodalen Objekterkennung in der visuellen Modalität besser sind als in der auditiven Modalität) führt der zusätzliche auditive Stimulus zum bereits vorhandenen visuellen Stimulus zu keiner Zunahme der neuronalen Aktivität im visuellen Kortex (Abbildung 1.15 a und b, VIS-Subjects). Eine Erhöhung der neuronalen Aktivität ist aber im auditorischen Kortex zu finden, wenn gleichzeitig zum auditiven Stimulus auch der visuelle Stimulus präsentiert wird (Abbildung 1.15 d, VIS-Subjects). Bei der AUD-Gruppe (Beobachter, die bei der unimodalen Objekterkennung in der auditiven Modalität besser sind als in der visuellen Modalität) verhält es sich genau andersherum. Wenn zum auditiven Stimulus zusätzlich der visuelle Stimulus präsentiert wird, führt dies zu keiner erhöhten neuronalen Aktivität im auditorischen Kortex (Abbildung 1.15 d, AUD-Subjects). Wenn aber parallel zum visuellen Stimulus auch der auditive Stimulus präsentiert wird, ist eine Erhöhung der neuronalen Aktivität im visuellen Kortex zu beobachten (Abbildung 1.15 a und b, AUD-Subjects). Visuell dominante Beobachter zeigen also nur schwache frühe Integrationseffekte im visuellen Kortex, aber starke frühe Integrationseffekte im auditorischen Kortex. Auditorisch dominante Beobachter zeigen frühe Integrationseffekte im visuellen Kortex, aber keinen frühen Integrationseffekt im auditorischen Kortex. Die Autoren sehen für diese Ergebnisse eine direkte Verbindung zur bereits erklärten „inverse effectiveness rule“ ([MS86]) und erklären, dass diese Regel auch auf globale neuronale Aktivitätsmuster im sensorischen Kortex des Menschen übertragbar ist. Desweiteren wird beschrieben, dass die multisensorische Verstärkung auf einer „subjective effectiveness“ der unimodalen Signale beruht, da für alle Beobachter dieselben physikalischen Stimuli verwendet wurden. Daraus kann man schließen, dass die multisensorischen Integrationsprozesse keinen starren Regeln unterworfen sind, sondern ein hoch adaptives System bilden, welches an die individuellen sensorischen und kognitiven Eigenschaften des Menschen angepasst ist.

Klinische Ergebnisse

Weitere Evidenz für die Existenz multimodaler Gehirnregionen im menschlichen Kortex kann gewonnen werden, indem elektrophysiologische Untersuchungen an Primaten mit den experimentellen Ergebnissen von Neglektpatienten nach Schlaganfällen kombiniert werden.

Untersuchungen an Primaten, z.B. [GG00], [DCG98], brachten verschiedene multimodale Regionen (z.B. am posterior parietalen Kortex oder präzentralen Gyrus) am Kortex von Primaten hervor. Obwohl die Ergebnisse auch innerhalb eines Experiments [GG00] nicht immer konsistent sind, geben sie doch Anhaltspunkte dafür, wo sich solche Regionen auch beim Menschen finden lassen. Bei Schlaganfallpatienten mit nachgewiesenem visuellen Neglekt konnten auch auditorische Defizite nachgewiesen werden, die nicht nur auf den Ausfall von auditorischen Hirnregionen zurückzuführen sind.

In [PLD03] sind eine Reihe von Patienten beschrieben, die neben visuellen Lokalisationsschwierigkeiten auch auditorische Lokalisationsschwierigkeiten haben. Darüberhinaus korrelieren die Defizite in ihrer Stärke; d.h. starke visuelle Defizite treten meist mit starken auditiven Defiziten zusammen auf. Die bei solchen Patienten typischerweise beschädigten Hirnregionen stimmen mit denen von Primaten als multisensorisch bekannten Regionen überein.

In einem Experiment mit Patienten nach einem Infarkt im insulären Kortex [MPS⁺99] konnten visuelle, auditive und taktile Defizite festgestellt werden. Der insuläre Kortex hat Verbindungen zu verschiedenen sensorischen kortikalen Arealen und dem limbischen System. Daher wird vermutet, dass die Beschädigung des insulären Kortex und der damit verbundenen Unterbrechung dieser Verbindung, eine tragende Rolle bei der Entstehung multimodaler Neglekte spielt.

Abschließend sollte noch zwischen Experimenten unterscheiden werden, bei denen der Einfluß einer Modalität auf eine andere untersucht wird und solchen, bei denen die Interaktion der Modalitäten im Vordergrund steht. Im ersten Fall wird eine Modalität als die dominante angesehen, die durch eine sekundäre Modalität beeinflusst wird. Dabei werden die Probanden oft auch angewiesen, die sekundäre Modalität zu ignorieren. Im Gegensatz dazu stehen Experimente bei denen die Probanden ausdrücklich aufgefordert werden, auf beide Modalitäten gleichzeitig zu achten.

Kapitel 2

Methodik

2.1 Psychophysik

Mit den Methoden der Psychophysik wird grundsätzlich versucht, eine Beziehung zwischen physikalischen Reizen und deren Wahrnehmung herzustellen. Das Ergebnis einer solchen Untersuchung ist dabei oft eine psychometrische Funktion, die den Zusammenhang zwischen der Intensität des physikalischen Reizes und der wahrgenommenen Intensität darstellt. Prinzipiell wird zwischen Diskriminations- und Detektionsexperimenten unterschieden.

- Diskriminationsexperimente

Hierbei geht es um die Ermittlung eines eben noch wahrnehmbaren Unterschieds zwischen zwei verschiedenen physikalischen Reizen, der Diskriminanzschwelle. Dieser Unterschied wird als „just noticeable difference“ (jnd) bezeichnet. Der jnd drückt jenen Unterschied zwischen zwei Reizen aus, bei dem in 50 % aller dargebotenen Fälle eine Differenzierung der beiden Reize nicht mehr möglich war.

- Detektionsexperimente

Bei diesen Untersuchungen soll eine absolute Schwelle ermittelt werden. Gesucht wird die kleinstmögliche Signalstärke, bei der der Reiz überhaupt noch wahrgenommen werden kann. Die absolute Schwelle ist dabei diejenige Signalstärke, bei welcher der Reiz noch in 50% aller Darbietungen wahrgenommen wird.

Bei psychophysischen Experimenten wird also versucht, die Grenzen eines Sinnessystems zu bestimmen. Ob diese Grenzen zuverlässig bestimmt werden, hängt zu einem großen Teil auch vom experimentellen Design ab. Wichtig sind hierbei auch Art und Umfang der Antwortmöglichkeiten eines Probanden. Gibt es

z.B. drei Antwortkategorien (z.B. größer, kleiner, gleich) kann dies das Ergebnis verfälschen. Der Proband tendiert leicht dazu, im Zweifelsfall eine neutrale Antwort („gleich“) zu geben. Diese Kategorie hat jedoch keinen Informationsgehalt im Hinblick auf eine zu ermittelnde Schwelle. Deshalb ist es oft besser, mit nur zwei Antwortkategorien (größer, kleiner) zu arbeiten und so vom Probanden nur nutzbringende Information zu erhalten. Diese Vorgehensweise wird *two-alternative-force-choice* (2afc) Design genannt.

Bei klassischen Experimenten dieser Art gibt es **eine** veränderliche Variable (der physikalische Reiz) und es wird versucht, alle anderen Umgebungsbedingungen konstant zu halten. Ansonsten besteht die Gefahr, dass das Messergebnis durch Störgrößen beeinflusst wird. Deshalb muss dafür gesorgt werden, dass während eines Experiments der Proband nicht von seiner Aufgabe abgelenkt wird und er sich mit seiner ganzen Aufmerksamkeit der Aufgabenstellung widmen kann. Andererseits entsteht durch das Festsetzen von Umgebungsbedingungen eine künstliche Umgebung, in der wir uns anders verhalten als in einer natürlichen Umgebung. In auditiven oder visuellen Lokalisationsexperimenten darf in der Regel der Kopf nicht bewegt werden, um Einflüsse, die durch Kopfbewegungen entstehen könnten, auszuschließen. Dies entspricht aber nicht unserem normalen Verhalten. In [Eas83] wird dieses Problem dadurch umgangen, dass Kopfbewegungen ausdrücklich in die Untersuchung mit einbezogen werden.

Die Aufmerksamkeit ist ein schwierig zu kontrollierender Faktor und kann durch verschiedenste Faktoren beeinflusst werden, wie z.B. bereits vorhandene Übermüdung, Überbeanspruchung des Probanden während des Experiments oder aber auch durch ein für den Probanden uninteressantes Experiment an sich.

Bei Experimenten, mit denen mehrere Sinnessysteme gleichzeitig untersucht werden, treten noch mehr zu kontrollierende Faktoren auf. Um Integrationseffekte zwischen zwei Modalitäten zu finden, ist es wichtig, dass der Proband auch auf beide Modalitäten gleichzeitig achtet. Ob diese Bedingung eingehalten wird, ist während des Experiments nur schwer zu kontrollieren und kann meist erst bei der Auswertung der Daten festgestellt werden. Deshalb ist es bei der Planung des Experiments wichtig, darauf zu achten, dass die Aufgabe nur gelöst werden kann, wenn der Proband wirklich auf alle Modalitäten achtet. Dazu ist es unter Umständen aber auch notwendig, den Probanden entsprechend zu motivieren. Wobei natürlich nicht vergessen werden darf, dass eigentlich Integrationseffekte und nicht „Motivationseffekte“ gemessen werden sollen. Eine weitere Schwierigkeit ist die „visuelle Dominanz“ unseres Gesichtssinns. Etliche Experimente (siehe [WW80] für eine Übersicht) zeigen das Phänomen, dass die Wahrnehmung eines Reizes in einer bestimmten Modalität durch einen gleichzeitig präsentierten visuell, inkonsistenten Reiz wesentlich in Richtung des visuellen Reizes verschoben wird. Ähnliche Ergebnisse gibt es auch für die „auditive Dominanz“ des Hörsinns (siehe auch [WW80]). Eine mögliche Erklärung für diese gegenseitige Beeinflussung (Bias) der Modalitäten wurde bereits in Kapitel 1.3 vorgestellt.

In [War79] wird versucht die visuelle Dominanz durch verschiedene Methoden abzuschwächen.

2.2 Statistik

Bei der Versuchsplanung muss vor allem beachtet werden, ob der zu erklärende Zusammenhang durch die zu messenden Variablen hinreichend dargestellt werden kann. Dazu muss geklärt werden, auf welche Art und Weise die zu messenden Daten statistisch aufbereitet werden müssen, um die vor dem Experiment aufgestellten Hypothesen zu testen. Zuerst werden die zu messenden Faktoren festgelegt. Anschließend sollte überprüft werden, ob es sich bei den zu erhebenden Daten um abhängige oder unabhängige Stichproben handelt. Bei Messwiederholungen treten möglicherweise beide Stichprobenarten gleichzeitig auf. Hier ist zu überlegen, wie erhobene Daten sinnvoll zusammengefasst werden, um Konflikte bei der Auswertung zu vermeiden. Es sollte auch bedacht werden, welche Verteilungsform für die erhobenen Daten in Frage kommt, damit aussagekräftige Hypothesentests angewandt werden können. Zur Lösung der statistischen Fragen wurde [Ebn92] und [Sac84] verwendet.

2.3 Modelle

In diesem Abschnitt werden die Modelle vorgestellt, die zur Modellierung der Ergebnisse aus den Wahrnehmungsexperimenten verwendet wurden. In Abschnitt 2.3.1, Abschnitt 2.3.2 und Abschnitt 2.3.3 werden Modelle vorgestellt, die alle als Basis die Wahrscheinlichkeitstheorie haben. Abschnitt 2.3.4 behandelt eine andere Unsicherheitstheorie, die versucht, die in der Wahrscheinlichkeitstheorie vorhandenen Grenzen zu erweitern.

2.3.1 Maximum-Likelihood-Integrator

Mit Hilfe der Maximum-Likelihood-Methode (MLE) ergeben sich für die Berechnung des Mittelwerts μ und der Varianz σ^2 einer normalverteilten Variablen X folgende Lösungen:

$$\tilde{\mu} = \frac{1}{n} \sum_{k=1}^n x_k = \bar{x} \quad (2.1)$$

$$\tilde{\sigma}^2 = \frac{1}{n} \sum_{k=1}^n (x_k - \bar{x})^2 \quad (2.2)$$

Die nicht erwartungstreue Schätzung für die Varianz σ^2 entspricht somit dem mittleren quadratischen Fehler (MSE) von $\tilde{\mu}$. Mit der MLE wird im Folgenden ein optimaler Schätzer für einen Wert x gesucht, der mit verschiedenen Sensoren gemessen wird. Die Messungen der Sensoren ergeben:

$$z_1 = x + v_1 \quad \text{und} \quad z_2 = x + v_2 \quad (2.3)$$

Die Messfehler v_1 und v_2 mit den Erwartungswerten $E[v_1] = E[v_2] = 0$ und den Varianzen σ_1^2 und σ_2^2 sind dabei zufällig, unabhängig und ohne Bias (Verzerrung). Der Wert x ist konstant aber unbekannt.

Ohne Bedenken der Allgemeinheit kann angenommen werden, dass sich der Schätzer für x als Linearkombination aus z_i schreiben lässt:

$$\hat{x} = k_1 z_1 + k_2 z_2 \quad (2.4)$$

Die Gewichte k_1 und k_2 sind unabhängig von x . Der Schätzfehler \tilde{x} ergibt sich aus:

$$\tilde{x} = \hat{x} - x \quad (2.5)$$

Die Schätzung ist optimal, wenn der MSE minimal ist. Damit folgt:

$$E[\tilde{x}] = E[k_1(x + v_1) + k_2(x + v_2) - x] \quad (2.6)$$

Mit $E[v_1] = E[v_2] = 0$ und $E[x] = x$ folgt

$$k_2 = 1 - k_1 \quad (2.7)$$

Der MSE ergibt sich damit zu

$$E[\tilde{x}^2] = k_1^2 \sigma_1^2 + (1 - k_1)^2 \sigma_2^2 \quad (2.8)$$

Wobei σ_1^2 und σ_2^2 die Varianzen von v_1 und v_2 sind. Differenzieren nach k_1 und Nullsetzen des rechten Ausdrucks führt zu:

$$2k_1\sigma_1^2 - 2(1 - k_1)\sigma_2^2 = 0 \quad \text{bzw.} \quad k_1 = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2} \quad (2.9)$$

Mit der MSE, bzw. der MLE, ergibt sich die Varianz des Schätzers zu:

$$\sigma_{12}^2 = E[\tilde{x}^2] = \left(\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2} \right)^{-1} \quad \text{oder} \quad \sigma_{12}^2 = \frac{\sigma_1^2 \sigma_2^2}{\sigma_1^2 + \sigma_2^2} \quad (2.10)$$

Der entsprechende Mittelwert ergibt sich durch Einsetzen in 2.4 zu:

$$\hat{x} = \left(\frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2} \right) z_1 + \left(\frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2} \right) z_2 \quad (2.11)$$

Der Gewichtungsfaktor für die Messungen z_1 und z_2 ergeben sich aus den Varianzen der beiden Messungen. Dabei wird demjenigen Messwert mit der kleineren Varianz mehr Vertrauen gegeben und entsprechend höher gewichtet.

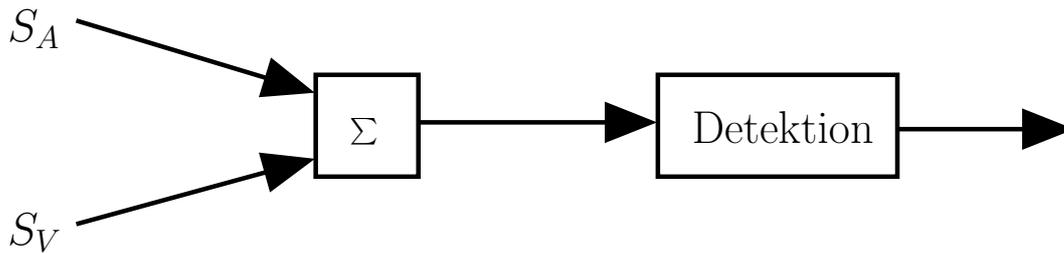


Abbildung 2.1: Das Linear-Summation-Modell lässt sich als „analoges UND-Modell“ beschreiben. Die beiden Kanäle werden zuerst integriert, dann findet die Detektion des aufsummierten Signals statt.

Anwendungsbeispiel zum Maximum-Likelihood-Integrator

Eine Anwendung dieses Verfahrens ist in [EB02] zu finden. Bei einem visuell-haptischen Integrationsexperiment wird versucht, die empirischen Daten mit obiger Methode zu modellieren. In einem Diskriminationsexperiment bestand die Aufgabe der Probanden darin, die Höhe eines Quaders zu schätzen. Dazu wurde der Quader entweder nur visuell oder nur haptisch oder bimodal präsentiert. Mittels einer Random Dot Kinematograms war es möglich, den visuellen Stimulus mit Rauschen zu überlagern.

Mit dem Maximum-Likelihood-Integrator ist eine Vorhersage der experimentellen bimodalen Daten möglich.

2.3.2 Linear Summation

Dieses Modell beruht auf der Annahme, dass die Signale mehrerer Kanäle zuerst mittels einer Linearkombination verknüpft werden, bevor eine Detektion des Gesamtsignals stattfindet. Diese Art der Informationsintegration kann somit durch eine „analoge UND-Verknüpfung“ dargestellt werden (siehe Abbildung 2.1). Dabei wird die Information aus den Sensorkanälen S_A und S_V zuerst addiert, bevor die Schwellendetektion stattfindet.

Bei einer psychophysischen Schwellenmessung kann zur Approximation an die empirischen Daten die Weibull-Funktion verwendet werden:

$$P(x) = 1 - 0.5 * e^{-\left(\frac{x}{\alpha}\right)^\beta} \quad (2.12)$$

Die Wahrscheinlichkeit für eine korrekte Antwort $P(x)$ für ein 2afc-Experiment liegt für $x = 0$ bei 0.5 (Ratewahrscheinlichkeit) und nähert sich mit größer werdenden x asymptotisch gegen 1. Die freien Parameter in der Gleichung sind α und β . Mit α wird die Schwelle bestimmt. Der Parameter β gibt die Steigung der

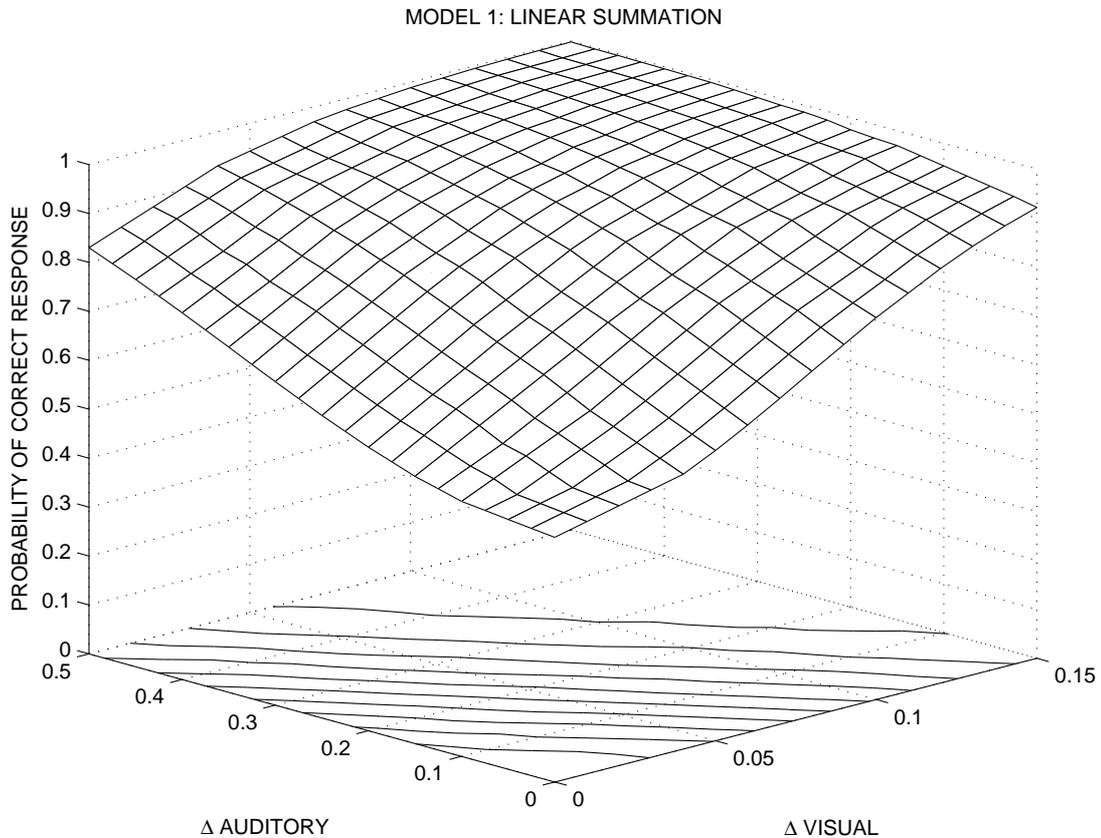


Abbildung 2.2: Weibull-Funktion für das Linear-Summation-Modell. Die Isolinien sind gerade.

Funktion an.

Bei der linearen Summation von zwei Kanälen, im Beispiel ein visueller und ein auditiver Kanal, ergibt sich die folgende Weibull-Funktion:

$$P(x_{vis}, x_{aud}) = 1 - 0.5 * e^{-\left(\frac{x_{vis}}{\alpha_{vis}} + \frac{x_{aud}}{\alpha_{aud}}\right)^\beta} \quad (2.13)$$

Damit sind insgesamt drei Parameter (α_{vis} , α_{aud} und β) zu schätzen. Beispielhaft ist in Abbildung 2.2 eine solche Weibull-Funktion zu sehen. An den X- und Y-Achsen sind die Änderungen der jeweiligen physikalischen Größen angetragen. An der Z-Achse ist die Wahrscheinlichkeit für eine korrekte Antwort bei einem 2afc-Experiment angetragen. Bei dieser Berechnung ergeben sich gerade Isolinien. Diejenige Reizstärke, bei der sich ein 81%ige Wahrscheinlichkeit für eine korrekte Antwort ergibt, wird als Schwelle definiert.

Für die Schwellen in beispielsweise einem Diskriminanzschwellenexperiment ergibt sich daraus der in Abbildung 2.3 gezeigte Zusammenhang.

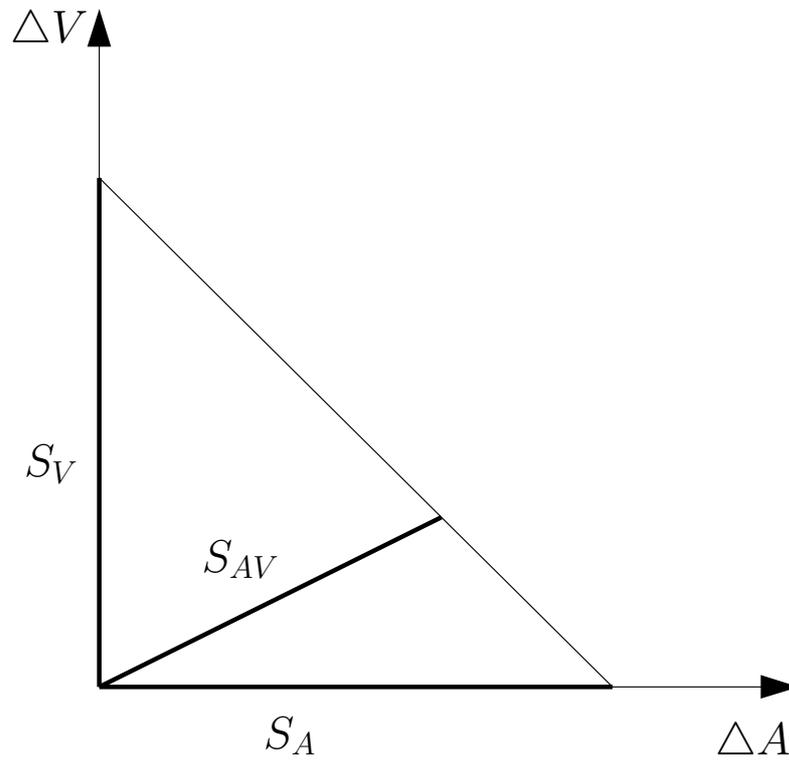


Abbildung 2.3: Beim Linear-Summation-Modell zeigt sich für die bimodale Schwelle S_{AV} gegenüber den unimodalen Schwellen S_A und S_V eine wesentliche Verbesserung. An den Achsen sind die Änderungen der physikalischen Größen der Stimuli angetragen. ΔA kann z.B. für die Änderung der Amplitude eines auditiven Stimulus und ΔV die Änderung der Fläche eines visuellen Stimulus stehen.

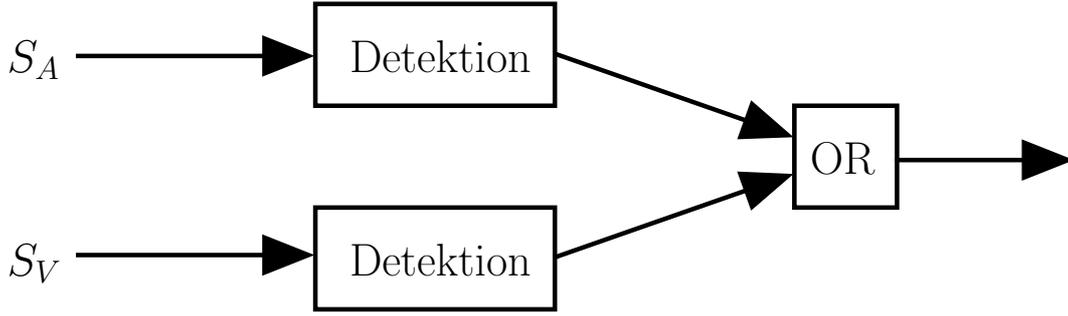


Abbildung 2.4: Das Probability-Summation-Modell kann als ODER-Verknüpfung dargestellt werden. Die Detektion findet in jedem einzelnen Kanal statt.

An den Achsen ist die Änderung der physikalischen Größe für die jeweilige Modalität (ΔA und ΔV) aufgetragen. Dabei bezeichnen S_A und S_V die unimodalen Schwellen. Die Größe der bimodalen Schwelle S_{AV} ändert sich entlang einer Geraden, abhängig von den prozentualen Anteilen der unimodalen Signaländerungen.

2.3.3 Probability Summation

Dieses Modell kann, im Vergleich zum Modell in Kapitel 2.3.2, als ODER-Verknüpfung, wie in Abbildung 2.4, visualisiert werden. Während beim Linear-Summation-Modell die Eingangssignale summiert werden, findet bei diesem Modell eine Summation von Wahrscheinlichkeiten statt.

Die Detektion des Signals findet unabhängig voneinander in jedem Kanal einzeln statt und wird deshalb auch „Independent Decisions Model“ genannt. Das Modell stellt somit eine Detektion fest, sobald in mindestens einem einzelnen Kanal eine Detektion stattgefunden hat. Bei mehr als einem Kanal steigt somit die Detektionswahrscheinlichkeit des Modells allein aufgrund der Tatsache, dass das Signal von mehreren Kanälen detektiert werden kann. Die Detektionswahrscheinlichkeit $P(av)$ für das kombinierte Signal berechnet sich aus 1 abzüglich dem Produkt der Wahrscheinlichkeiten, dass das Signal in den einzelnen Sensoren nicht detektiert wird:

$$P(av) = 1 - [1 - P(a)][1 - P(v)] \quad (2.14)$$

Mit der Weibull-Funktion aus Gleichung 2.12 ergibt sich die obige Gleichung zu:

$$P(x_{vis}, x_{aud}) = 1 - 0.5 * e^{-\left[\left(\frac{x_{vis}}{\alpha_{vis}}\right)^{\beta_{vis}} + \left(\frac{x_{aud}}{\alpha_{aud}}\right)^{\beta_{aud}}\right]} \quad (2.15)$$

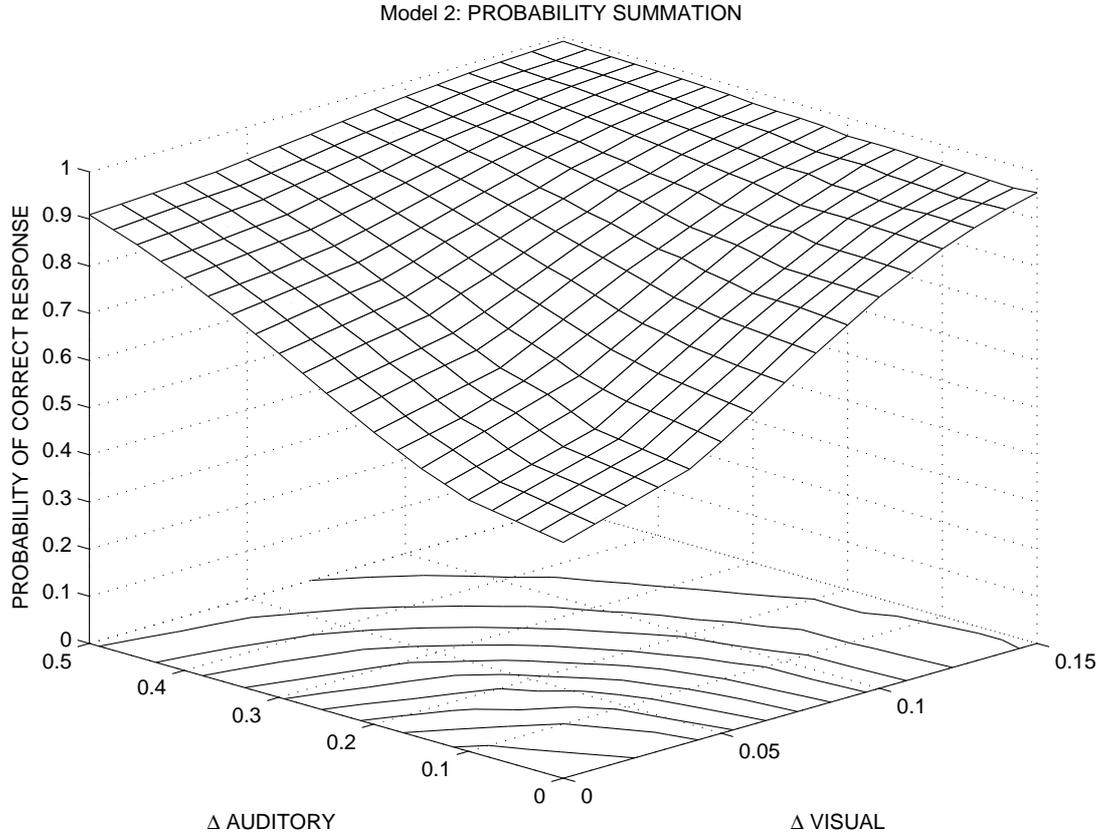


Abbildung 2.5: Weibull-Funktion für das Probability-Summation-Modell. Die Isolinien ergeben einen gekrümmten Kurvenverlauf.

Zur besseren Vergleichbarkeit zwischen Linear-Summation-Modell und Probability-Summation-Modell und weil das vier-parametrische Modell in der Anwendung in Kapitel 5.1 keine Verbesserung erzielt, wurde $\beta_{aud} = \beta_{vis} = \beta$ gleichgesetzt. Daraus resultiert die folgende Weibull-Funktion:

$$P(x_{vis}, x_{aud}) = 1 - 0.5 * e^{-\left[\left(\frac{x_{vis}}{\alpha_{vis}}\right)^\beta + \left(\frac{x_{aud}}{\alpha_{aud}}\right)^\beta\right]} \quad (2.16)$$

Abbildung 2.5 zeigt eine solche Weibull-Funktion. Die Isolinien sind gekrümmt. Die Krümmung ist abhängig vom Parameter β . Ein größeres β ergibt eine stärker gekrümmte Kurve. Die Parameter α_{aud} und α_{vis} bestimmen die Steigung der Funktion. Ein kleineres α führt zu einer stärkeren Steigung.

Abbildung 2.6 zeigt den entsprechenden Schwellenverlauf, analog zu Abbildung 2.3. Die Verbesserung der bimodalen Schwelle S_{AV} gegenüber den unimodalen Schwellen S_A und S_V ist nun von der Krümmung der Kurve, bzw. von der Größe des Parameters β der Weibull Funktion, abhängig.

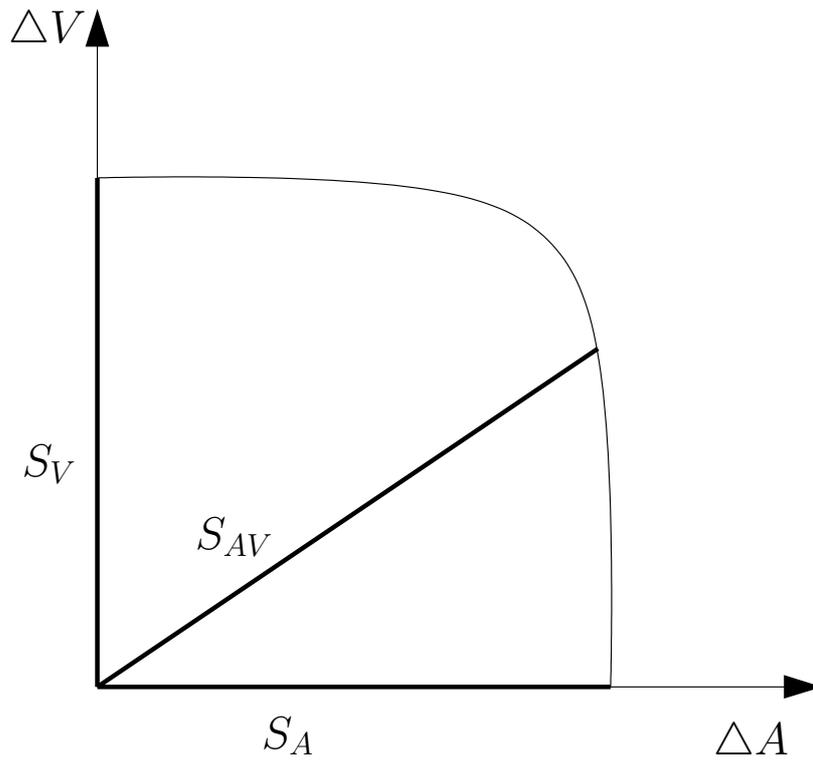


Abbildung 2.6: Für das Probability-Summation-Modell ergibt sich für die bimodale Schwelle S_{AV} gegenüber den unimodalen Schwellen S_A und S_V eine von der Kurvenkrümmung abhängige Verbesserung. An den Achsen sind die Änderungen der physikalischen Größen der Stimuli angetragen. ΔA kann z.B. für die Änderung der Amplitude eines auditiven Stimulus und ΔV die Änderung der Fläche eines visuellen Stimulus stehen.

Vergleich zwischen Linear- und Probability-Summation-Modell

Die beiden Modelle können gut miteinander verglichen werden, da sie die beiden Extremfälle möglicher Integrationsmechanismen darstellen. Wird beim Probability-Summation-Modell $\beta = \inf$, ergibt die Kurve in Abbildung 2.6 ein Rechteck. Dies bedeutet keinerlei Interaktion und somit keine Verbesserung der bimodalen Schwelle. Für $\beta = 1$ vereinfacht sich Gleichung 2.16 zu Gleichung 2.13. Das Probability-Summation-Modell kann somit in das Linear-Summation-Modell verwandelt werden. Beide Modelle machen aber ganz unterschiedliche Annahmen über die Art der Informationsverarbeitung. Beim Probability-Summation-Modell steigt die Wahrscheinlichkeit der Detektion eines Signals aufgrund des statistischen Vorteils, der durch das Vorhandensein mehrerer Kanäle zustande kommt. Das Linear-Summation-Modell hat zur Grundlage, dass die Eingangssignale aus den unterschiedlichen Kanälen bereits vor der Detektion zusammengeführt und summiert werden.

Anwendungsbeispiel zum Linear- und Probability-Summation-Modell

In [MS80] findet man eine Gegenüberstellung beider Modelle für ein audiovisuelles Detektionsexperiment. Zusätzlich findet man dort noch zwei Abwandlungen des Probability-Summation-Modells. Der visuelle Stimulus bei diesem Experiment bestand aus einer pulsierenden Lampe. Das visuelle Rauschen war ein Dauerlicht. Der auditive Stimulus bestand aus einem pulsierenden 500 Hz Ton, eingebettet in weißes Hintergrundrauschen. Die Präsentation der Stimuli erfolgte räumlich kongruent in der Peripherie (40 Grad Azimuth) des Beobachters. Durch die Präsentation des auditiven Stimulus über einen Lautsprecher nahe des visuellen Stimulus konnte eine räumliche Diskrepanz, wie sie bei der Verwendung von Kopfhörern entstehen würde, vermieden werden. Durch die Präsentation des visuellen Stimulus außerhalb des fovealen Bereichs, wird besonders der auf Detektion spezialisierte periphere Bereich des visuellen Systems angesprochen.

Die experimentellen Ergebnisse zeigen keine Unterstützung für das Linear-Summation-Modell. Die Ergebnisse lassen sich mit dem Probability Summation-Modell besser vorhersagen.

Die beiden Modelle finden sinngemäß auch ihre Anwendung in der statistischen Entscheidungstheorie [Hau74]. Im Falle von Mehrfachbeobachtungen stellt sich dort die Frage, wie mehrere zeitlich und räumlich wiederholt auftretende Signale zu behandeln sind, bevor der Beobachter eine Entscheidung trifft. Es werden verschiedene Möglichkeiten betrachtet, wie unabhängige Signale zu einem Kanal und damit zu einer einzigen Entscheidung zusammengeführt werden. Dabei wird auch die additive Zusammenfassung, die dem Grundprinzip des Linear-Summation-Modells entspricht behandelt. Eine andere Variation in [Hau74] ist die Kom-

bination zweier aufgrund der Einzelbeobachtungen getroffenen Entscheidungen zu einer endgültigen Entscheidung und entspricht dem Prinzip des Probability-Summation-Modells.

2.3.4 Dempster-Shafer Evidenztheorie

Diese Theorie wurde ursprünglich von A. Dempster entwickelt und von G. Shafer [Sha76] weiter entwickelt. Später wurden bestimmte Axiome noch von Smets [Sme88] modifiziert.

Diese Theorie möchte sich deutlich von der Wahrscheinlichkeitstheorie abgrenzen. Der Vorteil dieser Methode ist, dass Unsicherheit und der Umgang mit unsicherem Wissen ein elementarer Bestandteil der Theorie ist. Unsicheres Wissen ist mit der Bayes'schen Wahrscheinlichkeitstheorie etwas schwieriger zu handhaben, wie im Anwendungsbeispiel am Ende dieses Abschnitts zu sehen ist. Nachteilig ist zu erwähnen, dass die verwendeten Begriffe teilweise nicht exakt definiert sind [Umk00]. Es bleibt auch etwas unklar, wie die Zuweisung eines Wertes an eine Hypothese zustande kommt. Zum Begriff „degree of belief“ , bzw. „body of evidence“ schreibt Shafer [Sha76] z. B.:

„Having surveyed the sometimes vague and sometimes confused perception and understanding that constitutes a given body of evidence, he can announce a number that represents the degree to which he judges that evidence to support a given proposition and, hence, the degree of belief he wishes to accord the proposition.“

Durch diese Unschärfe bleibt aber auch Raum zur Interpretation, der von Shafer durchaus beabsichtigt ist.

Die Definitionsmenge, die als „frame of discernment“ (foc) bezeichnet wird, setzt sich aus einer Menge von Elementaraussagen zusammen. Nach Shafer muss die Definitionsmenge alle möglichen Aussagen zu einem Problem oder einer Fragestellung enthalten. Die Aussagen müssen sich gegenseitig ausschließen und es muss genau eine Aussage wahr sein.

Auf der Potenzmenge der Definitionsmenge wird eine Funktion $m : 2^\Theta \rightarrow [0, 1]$ definiert. Diese Funktion wird basic probability assignment (bpa) genannt. Mit Hilfe dieser Funktion werden nun Werte an die Elemente aus der Potenzmenge der Definitionsmenge vergeben. Für diese Funktion gelten folgende drei Axiome:

1.
$$\forall A \subseteq \Theta : 0 \leq m(A) \leq 1 \tag{2.17}$$

2.
$$m(\emptyset) = 0 \tag{2.18}$$

3.

$$\sum_{A \subseteq \Theta} m(A) = 1 \quad (2.19)$$

Das erste Axiom legt den Wertebereich für die Funktion m fest. Das zweite Axiom setzt den Wert für die leere Menge und das dritte Axiom definiert die Beziehung zwischen den Werten für m . Die von der m-Funktion vergebenen Werte werden als eine bestimmte Menge an „Glauben“ verstanden, der exakt an die einzelnen Aussagen verteilt wird. Eine m-Funktion stellt somit eine Evidenz dar. Die Bedeutung eines einzelnen m-Wertes hängt laut Shafer vom aktuellen Kontext ab und wird somit über die zu lösende Problemstellung definiert.

Des Weiteren wird eine Glaubensfunktion $Bel : 2^\Theta \rightarrow [0, 1]$ (belief function) wie folgt definiert:

$$\forall A, B_i \subseteq \Theta : Bel(A) = \sum_{B_i \subseteq A} m(B_i) \quad (2.20)$$

Diese Funktion drückt nun den Glauben aus, den man in eine bestimmte Aussage hat. In der Glaubensfunktion $Bel(A)$ werden nun die m-Werte aller Teilmengen von A auf summiert. $Bel(A)$ gibt somit den Glauben an, der mindestens in der Aussage A enthalten ist.

Den Kern der Dempster-Shafer Evidenztheorie bildet die Kombinationsregel (Dempster's rule of combination). Mit ihr können zwei m-Funktionen m_1 und m_2 auf eine m-Funktion $m_{12} = m_1 \otimes m_2$ abgebildet werden. Dabei gilt

$$m_{12}(\emptyset) = 0 \quad (2.21)$$

m_1 und m_2 müssen dabei dieselbe Definitionsmenge haben. Außerdem gilt:

$$\forall A_i, B_j \subseteq \Theta : \sum_{A_i \cap B_j = \emptyset} m_1(A_i) * m_2(B_j) < 1 \quad (2.22)$$

und für die Kombinationsregel:

$$\forall A_i, B_j, C \subseteq \Theta, C \neq \emptyset : m_{12}(C) = \frac{\sum_{A_i \cap B_j = C} m_1(A_i) * m_2(B_j)}{1 - \sum_{A_i \cap B_j = \emptyset} m_1(A_i) * m_2(B_j)} \quad (2.23)$$

Zu Begründung dieser Regel schreibt Shafer in [Sha76]:

„Though this essay provides no conclusive a priori argument for Dempster's rule, we will see in the following chapters that the rule does seem to reflect the pooling of evidence, provided only that the belief functions to be combined are actually based on entirely distinct bodies of evidence and that the frame of discernment discerns the relevant interactions of those bodies of evidence.“

Modifikationen nach Smets

Smets nimmt an der Theorie einige Veränderungen vor. Das zweite Axiom, Gleichung 2.18, wird gestrichen. Auch die Bedingung 2.21 fällt weg. Gleichung 2.23 vereinfacht sich deshalb zu

$$\forall A_i, B_j, C \subseteq \Theta : m_{12}(C) = \sum_{A_i \cap B_j = C} m_1(A_i) * m_2(B_j) \quad (2.24)$$

Dadurch ist zu beachten, dass jetzt auch der leeren Menge \emptyset Werte zugewiesen werden können. Für die Definitionsmenge ergibt sich somit, dass nun höchstens eine Aussage wahr sein darf.

Das Verteilen von Evidenz an die leere Menge bedeutet, dass man sich nicht für eine der Hypothesen aus der Grundmenge Θ entscheiden kann. Somit kann Unwissen über die vorhandenen Hypothesen ausgedrückt werden. Evidenz für die leere Menge kann aber auch bedeuten, dass die gegebene Fragestellung mit den vorhandenen Hypothesen nicht beantwortet werden kann.

Bei Teilmengen von Θ handelt es sich mengentheoretisch um Vereinigungen aus Elementaraussagen. In einer Teilmenge kann die eine oder die andere Hypothese wahr sein. Dadurch ergeben sich für die Teilmengen neue Interpretationsmöglichkeiten, die in Abschnitt 4.1.3 angewendet werden.

Anwendungsbeispiel zur Dempster-Shafer-Theorie

Ein anschauliches Beispiel zur Anwendung der Dempster-Shafer-Theorie ist in [Sha76] zu finden. Dort wird die Frage behandelt, ob Lebewesen um den Stern Sirius existieren. Dazu wird eine Menge von Hypothesen $\Theta = \{\theta_1, \theta_2\}$ mit

θ_1 : Es gibt Leben um den Stern Sirius

θ_2 : Es gibt kein Leben um den Stern Sirius

aufgestellt. Bei einer vollständigen Unwissenheit darüber, ob um den Stern Sirius Leben existiert oder nicht, kann folgende Glaubensfunktion aufgestellt werden:

$$Bel(X) = \begin{cases} 0 & \text{if } X \neq \Theta \\ 1 & \text{if } X = \Theta \end{cases} \quad (2.25)$$

Zu beachten ist dabei, dass die komplette Unwissenheit ausgedrückt wird, ohne dem Komplement von X irgendeinen Wert zu zuweisen. Dies ist mit der Bayes-Theorie nicht ohne weiteres möglich, da dort gilt:

$$Bel(X) + Bel(\bar{X}) = 1 \quad (2.26)$$

Es ist somit mit der Bayes-Theorie nicht möglich, einer Menge X und gleichzeitig dessen Negation \bar{X} kleine Werte zuzuordnen. Man kann sich aber, um auch in der Bayes-Theorie Unwissenheit auszudrücken, mit folgender Zuweisung helfen:

$$Bel(X) = Bel(\bar{X}) = \frac{1}{2} \quad (2.27)$$

Vollkommene Unwissenheit wird also dadurch ausgedrückt, indem einer Aussage und dessen Negation jeweils die Hälfte des zu verteilenden Glaubens zugeteilt wird. Dies funktioniert aber nur, bei der Zweiteilung der Aussagen in X und \bar{X} . Eine Hypothesenmenge Θ kann aber aus durchaus mehreren Hypothesen bestehen. Auf einer feineren Granularitätsstufe unserer Ausgangsfrage kann man z.B. folgende drei Hypothesen (A,B,C) aufstellen:

A: Es gibt Leben um den Stern Sirius

B: Es gibt Planeten, aber kein Leben um den Stern Sirius

C: Es gibt keine Planeten um den Stern Sirius

Daraus ergibt sich die Hypothesenmenge $\Omega = \{A, B, C\}$. Nun wird die erste Hypothesenmenge $\Theta = \{\theta_1, \theta_2\}$ mit $\theta_1 = A$ und $\theta_2 = \{A, B\}$ zur zweiten Hypothesenmenge (Ω) in Beziehung gesetzt. Wenn nun mit der Bayes-Methode für die Menge Θ komplette Unwissenheit über die Hypothesen herrscht, kann dies durch folgende Gleichung repräsentiert werden:

$$Bel(\{\theta_1\}) = Bel(\{\theta_2\}) = \frac{1}{2} \quad (2.28)$$

Wenn nun die Menge Ω betrachtet wird, muss folgende Gleichung erfüllt werden:

$$Bel(\{A\}) + Bel(\{B\}) + Bel(\{C\}) = 1 \quad (2.29)$$

Um nun in dieser Hypothesenmenge vollkommene Unwissenheit auszudrücken muss der Glaube auf alle drei Hypothesen gleich verteilt werden:

$$Bel(\{A\}) = Bel(\{B\}) = Bel(\{C\}) = \frac{1}{3} \quad (2.30)$$

Dies führt aber zu

$$Bel(\{A\}) = \frac{1}{3} \quad \text{und} \quad Bel(\{B, C\}) = \frac{2}{3} \quad (2.31)$$

und somit zu einem Widerspruch zu Gleichung 2.28, da θ_1 die gleiche Bedeutung wie A und θ_2 die gleiche Bedeutung wie $\{B, C\}$ haben.

Dieser Widerspruch resultiert aus der Tatsache, dass in der Bayes-Theorie die Restmenge an Glauben, welche nicht an eine Aussage X vergeben wird, automatisch der Negation \bar{X} zugeteilt wird. Dies ist bei der Dempster-Shafer-Theorie nicht notwendig. Eine Menge an Glauben, die nicht auf einzelne Hypothesen

verteilt werden möchte, kann dort einfach der Definitionsmenge Θ zugewiesen werden (siehe Gleichung 2.25). Dadurch ist es möglich, jeder echten Teilmenge aus Θ den Wert Null zuzuweisen. Damit lässt sich auf intuitive Weise Unsicherheit ausdrücken, da gilt: Jede Teilmenge θ_i aus Θ bekommt den Wert Null, aber die ganze Definitionsmenge Θ erhält den Wert 1. Dies bedeutet, dass die wahre Aussage in Θ enthalten ist, aber nicht näher bestimmt werden kann.

2.3.5 Minimum-Modell

Das Modell wurde im Rahmen dieser Dissertation entwickelt. Dabei wird aus einer Reihe von Messwerten x_1, \dots, x_n , derjenige Wert x ausgewählt, der möglichst Nahe an Null liegt. Auf diesen Wert x kann noch eine Konstante addiert werden. Die Berechnung von x lässt sich mit folgender Formel ausdrücken:

$$x = MIN(X) + t \quad (2.32)$$

Die Menge X enthält die Messwerte x_1, \dots, x_n . Die Funktion $MIN()$ berechnet das absolute Minimum aus X liefert aber den gefundenen Wert mit dem Vorzeichen zurück. Der Parameter t ist eine Konstante.

Anwendungsbeispiel zum Minimum-Modell

Wie in Kapitel 5.2 zu sehen ist, kann mit diesem Modell eine Informationsintegration aus verschiedenen sensorischen Kanälen beschrieben werden.

Kapitel 3

Computersimulation eines multisensorischen Inputs

Die typischen Objekt- und Beobachterbewegungen unter natürlichen Bedingungen bewirken systematische statistische Kovariationen in den sensorischen Kanälen, die eine mögliche Grundlage für intermodale Interaktionen darstellen. Diese statistischen Kovariationen wurden mit Hilfe von Computersimulationen untersucht [ZRHS02].

Die Simulation von typischen Relativbewegungen zwischen einem Beobachter und bewegten Objekten mit verschiedenen Bewegungseigenschaften ergab ein systematisches statistisches Muster in Bezug auf die resultierenden Kombinationen von auditiven und visuellen Signalen. Dieses Muster zeigte, dass bestimmte Kovariationen, wie z.B. eine gleichsinnige Vergrößerung von Objekt-Blickwinkel und wahrgenommener Lautstärke, statistisch häufiger auftraten als andere Kombinationen.

Das Programm simulierte folgendes Szenario: Ein Beobachter, siehe Abbildung 3.1, der entweder stillsteht oder sich in eine wählbare Richtung bewegte, hatte zwei sich überlappende Wahrnehmungsbereiche. Einen für sein visuelles und einen für sein auditives System. Im Wahrnehmungsbereich des Beobachters befanden sich eine Anzahl von Objekten mit jeweils in einem bestimmten Intervall statistisch gleichverteilter Größe, Schallintensität, Bewegungsrichtung und Geschwindigkeit.

3.1 Methode

Die beim Beobachter ankommenden physikalischen Größen der Objekte waren die Fläche des Abbildes des Objekts auf der Retina und die Schallintensität des

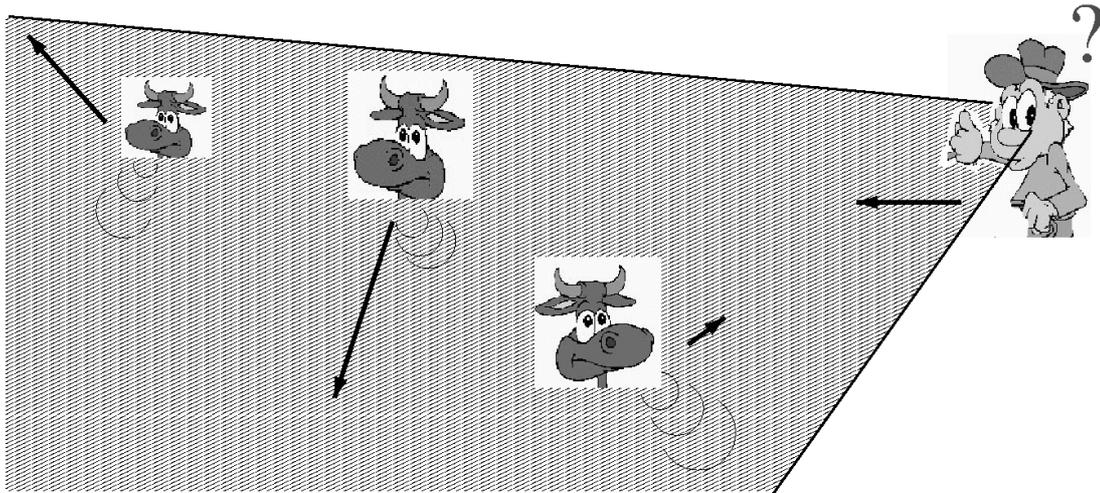


Abbildung 3.1: Bildliche Veranschaulichung der Simulation

Objekts. Die Fläche des Abbildes A_0 ändert sich nach Formel $A_d = A_0/d^2$ quadratisch mit dem Abstand d .

Für das vom Objekt ausgestrahlte auditive Signal wurde angenommen, dass es sich um eine punktförmige Schallquelle mit isotroper Abstrahlung handelt. Die am Ohr des Beobachters messbare Leistung der Schallquelle wurde durch die Schallintensität I (=Leistung pro Fläche) ermittelt. Die Schallintensität I_0 ändert sich linear mit der durchstrahlten Fläche und damit quadratisch mit dem Abstand d zur Schallquelle nach der Formel: $I_d = I_0/d^2$.

Der Wahrnehmungsbereich des Beobachters betrug 360° für beide Modalitäten. Der Wahrnehmungsbereich war für die zwei Modalitäten unterschiedlich groß. Er bestand aus jeweils einer Kreisscheibe, deren äußerer Radius durch die minimalen Wahrnehmungsgrenzen (z.B. Hörschwelle) und deren innerer Radius durch die maximalen Wahrnehmungsgrenzen (z.B. Schmerzschwelle) bestimmt wurde. Bewegte sich ein Objekt aus dem auditiven und visuellen Wahrnehmungsbereich, tauchte ein neues Objekt an der Wahrnehmungsgrenze wieder auf. Damit wurde die Anzahl der wahrgenommenen Objekte über die Zeit konstant gehalten, da ansonsten die meisten Objekte durch die geradlinige Bewegung von Objekten und Beobachter aus dem Wahrnehmungsbereich des Beobachters verschwinden würden. Berührte ein Objekt die innere Grenze des Wahrnehmungsbereichs ändert entweder der Beobachter oder das Objekt seine aktuelle Bewegungsrichtung. Die Richtungsänderung wurde dabei vom Schnelleren der beiden vollzogen.

Nach der Initialisierung wurde für einen gewissen Zeitraum in diskreten Abständen die Entfernung des Beobachters zu jedem Objekt ermittelt. Daraufhin wurde für das jeweilige Objekt die relative Größe, sowie die relative Schallintensität in Abhängigkeit der Entfernung zum Beobachter berechnet. Die zwei berechneten Werte ergaben einen Punkt in einem Koordinatensystem mit dem

auditiven Wert als Abszisse und dem visuellen Wert als Ordinate. In Abbildung 3.2 wird das Entstehen der Punktwolke deutlich. Ein Punkt rechts oben stellt ein groß und laut erscheinendes Objekt dar. Ein Punkt im linken unteren Bereich bedeutet ein klein und leise erscheinendes Objekt. Die Relativbewegung wird durch die auf einer Geraden liegenden Punkte sichtbar. Kleine Punktabstände markieren eine langsame, große Abstände eine schnelle Relativbewegung. Abbildung 3.3 zeigt die vier verschiedenen Wahrnehmungsbereiche des Beobachters. Die beiden gestrichelten Linien zeigen die Grenzen des Wahrnehmungsbereichs. Links der vertikalen Linie ist das Objekt zu leise und unterhalb der horizontalen Linie ist das Objekt zu klein, um vom Beobachter wahrgenommen zu werden. Im rechten oberen Bereich ist das Objekt in beiden Modalitäten wahrnehmbar. Im rechten unteren Bereich ist das Objekt unterhalb der visuellen Wahrnehmungsschwelle und deshalb nur hörbar. Im linken oberen Bereich ist das Objekt unterhalb der auditiven Wahrnehmungsschwelle und daher nur sichtbar. Im linken unteren Bereich ist das Objekt gleichzeitig unterhalb beider Wahrnehmungsschwellen und deshalb weder hör- noch sichtbar.

3.2 Ergebnis

In Abbildung 3.4 ist der stationäre Fall zu sehen. Es bewegten sich weder der Beobachter, noch die Objekte. In diesem Fall hatten alle Objekte immer den selben Abstand zum Beobachter. Jeder Punkt repräsentiert dabei ein Objekt. Die Objekte waren über die vom Beobachter wahrzunehmende Fläche gleich verteilt.

In Abbildung 3.5 bewegte sich der Beobachter mit konstanter Geschwindigkeit geradlinig in eine festgelegte, ausgewählte Richtung. Die Objekte standen still.

In Abbildung 3.6 bewegte sich der Beobachter nicht. Die Objekte bewegten sich mit verschiedenen, aber jeweils konstanten, Geschwindigkeiten geradlinig in eine statistisch ausgewählte Richtung.

In Abbildung 3.7 bewegte sich der Beobachter mit festgelegter Geschwindigkeit in eine bestimmte Richtung. Die Objekte bewegten sich jeweils in statistisch ausgewählte Richtungen mit konstanter, statistisch ausgewählter Geschwindigkeit.

3.3 Auswertung

Das Muster in den gezeigten Grafiken macht deutlich, dass bestimmte Kovariationen häufiger auftraten als andere. Gleichsinnige Veränderungen von Größeneindruck und Schallintensität überwogen dabei sehr deutlich. Die Struktur der Muster war dabei von der Art der Bewegung zwischen Beobachter und Objekt

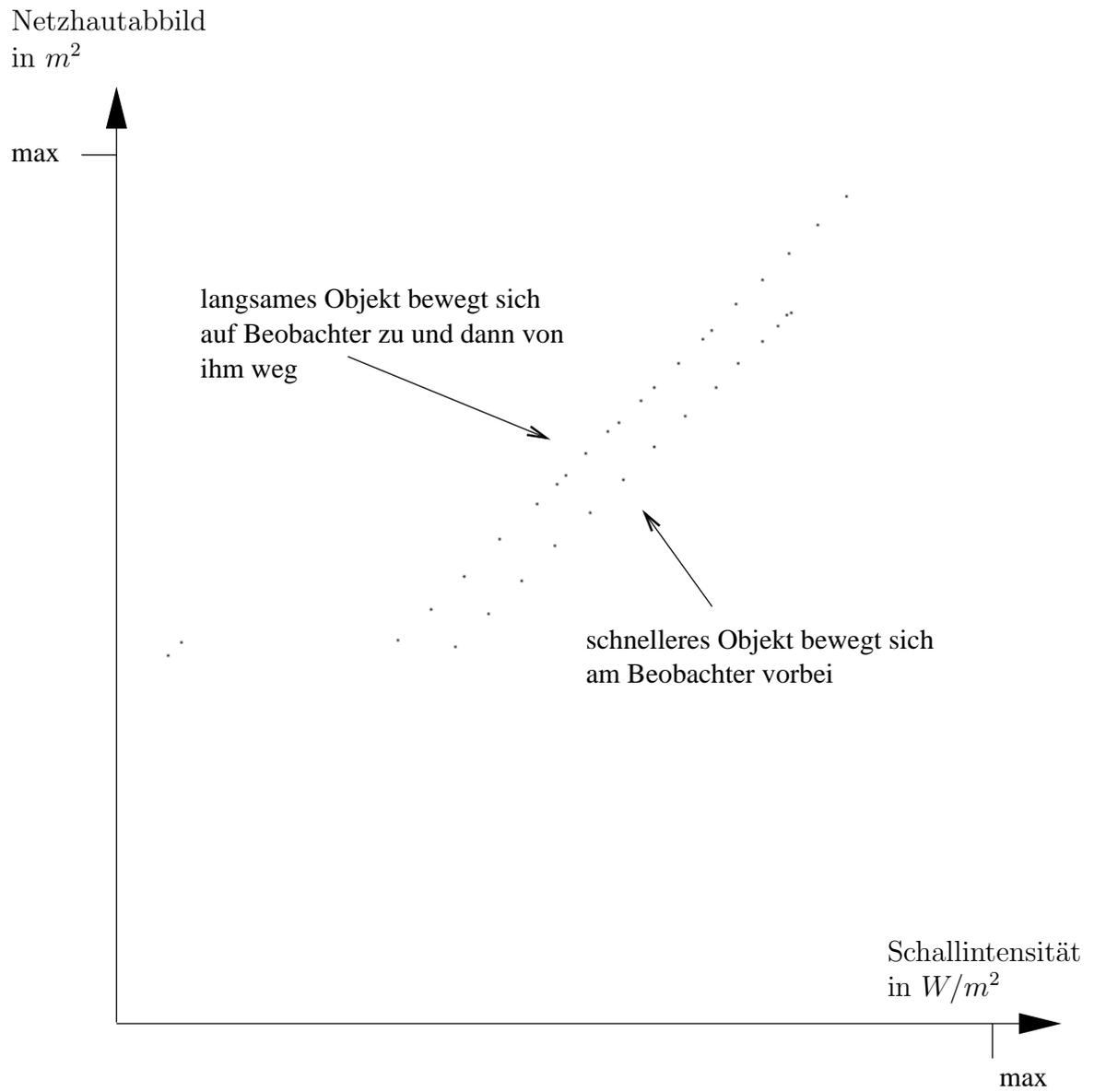


Abbildung 3.2: Die Entstehung der Punktwolke

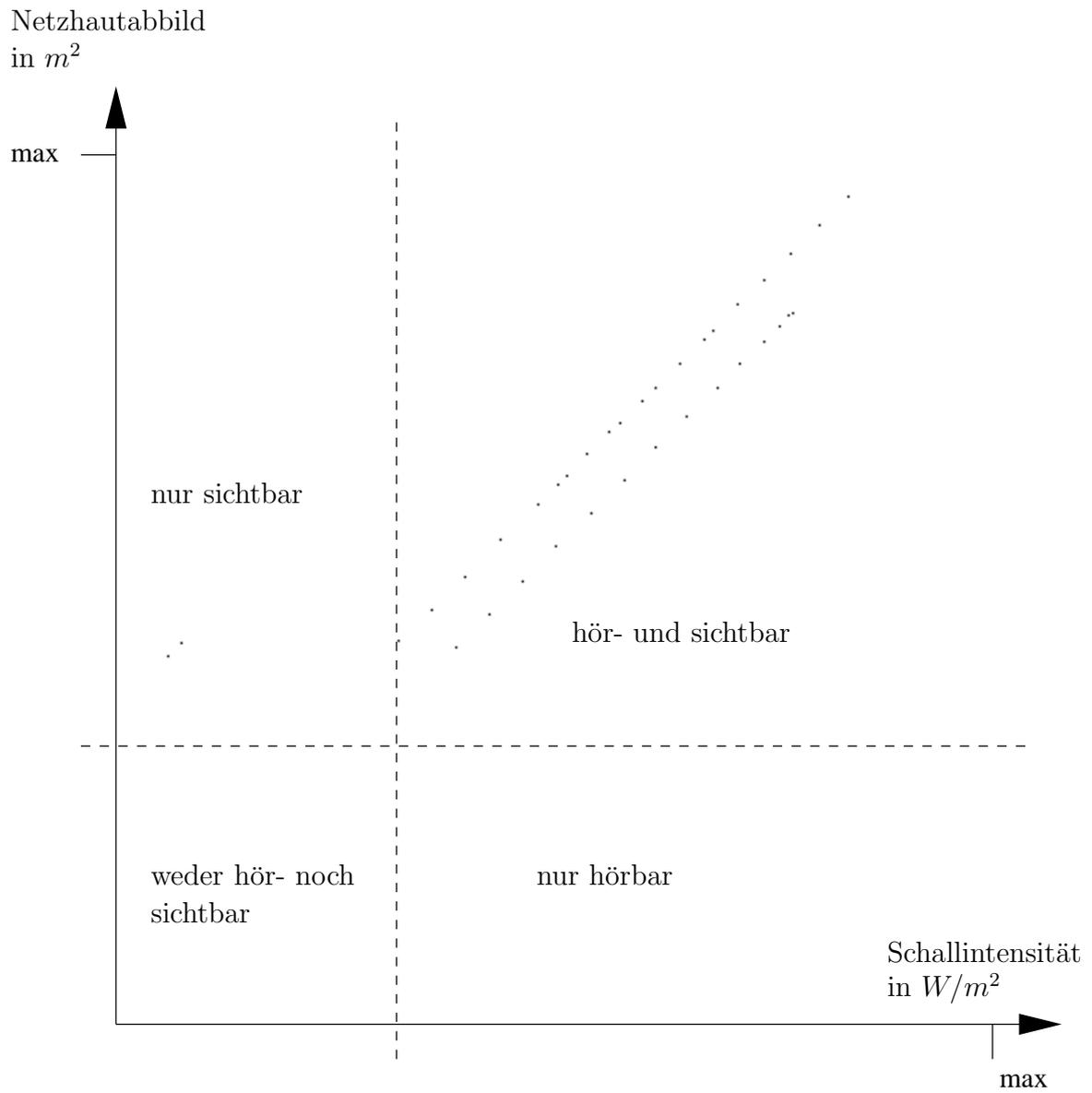


Abbildung 3.3: Die vier unterschiedlichen Wahrnehmungsbereiche des Beobachters

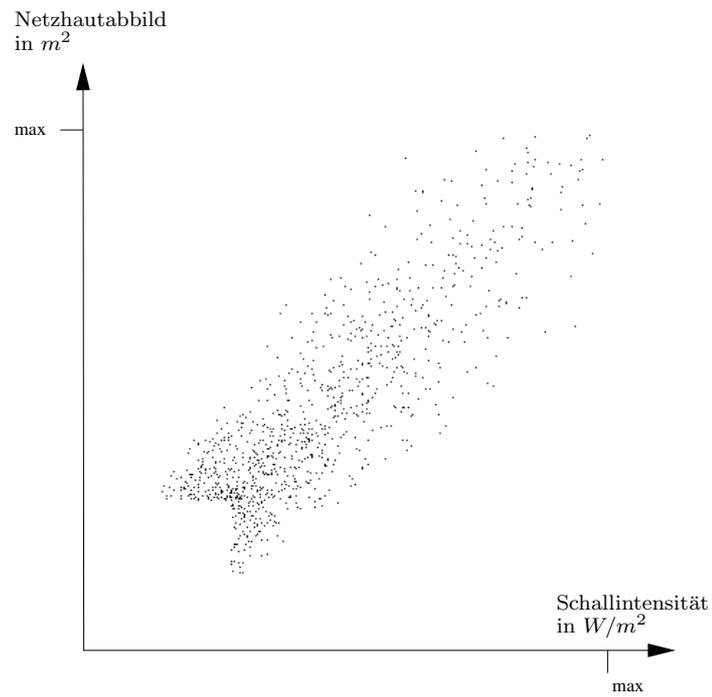


Abbildung 3.4: Der Beobachter und die Objekte standen still. Jeder Punkt repräsentiert ein Objekt.

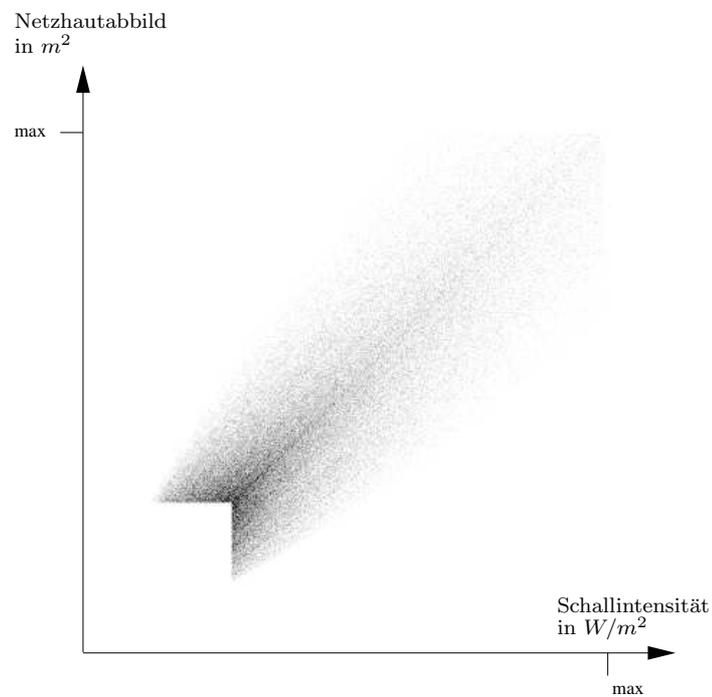


Abbildung 3.5: Der Beobachter bewegte sich. Die Objekte standen still.

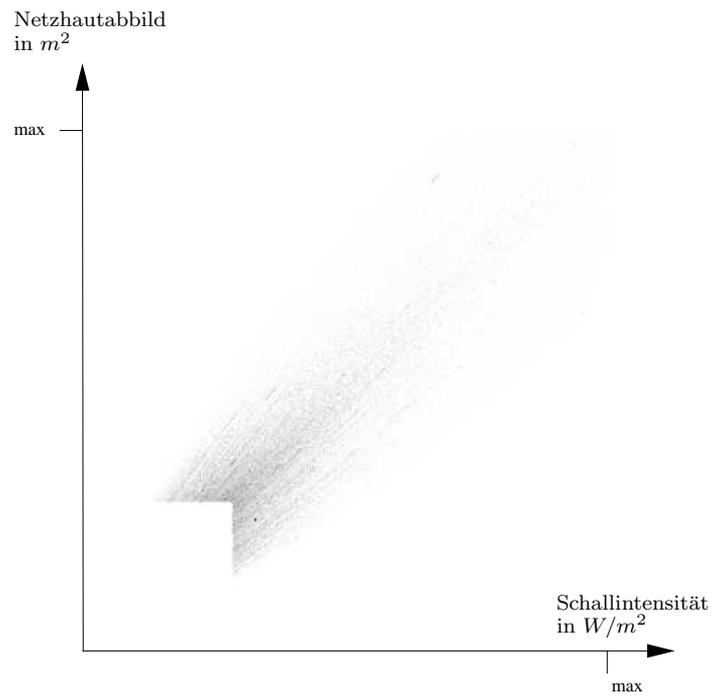


Abbildung 3.6: Der Beobachter stand still und die Objekte bewegten sich mit verschiedenen Geschwindigkeiten in unterschiedliche Richtungen.

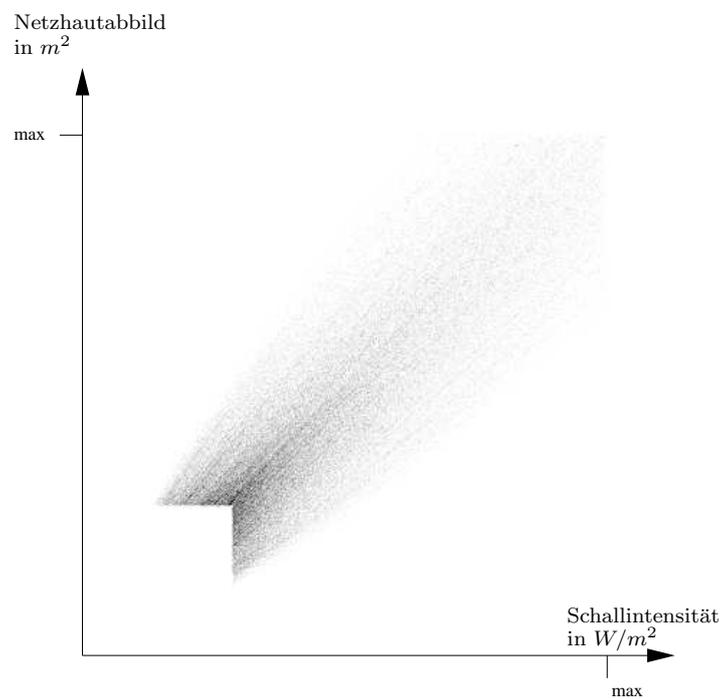


Abbildung 3.7: Der Beobachter bewegte sich mit konstanter Geschwindigkeit. Die Objekte bewegten sich mit unterschiedlichen Geschwindigkeiten in verschiedene Richtungen.

unabhängig. Da die Ausgangsposition der Objekte auf der vom Beobachter wahrnehmbaren Fläche gleich verteilt war (siehe Abbildung 3.4), ist dieses Ergebnis durchaus überraschend. Durch die geradlinige Bewegung des Beobachters und der Objekte ist leicht nachvollziehbar, dass die vom Beobachter wahrgenommenen Objekte mit der Zeit kleiner und leiser erscheinen. Letztlich würden sich alle Objekte, die einen anderen Bewegungsvektor als der Beobachter haben, aus dessen Wahrnehmungsbereich entfernen. Dies wurde aber dadurch verhindert, dass solche Objekte an einer zufälligen Stelle am Rande des Wahrnehmungsbereichs wieder auftauchten.

Die statistischen Redundanzen der natürlichen Umgebung sind nach unserer Ansicht die Voraussetzung für die Mechanismen der sensorischen Informationsverarbeitung. Die beobachtete statistische Kovariation zwischen visuellem und auditivem Input sollte daher zu einer spezifischen Form der intersensorischen Interaktion führen.

Es wird erwartet, dass die Schwellen für die Wahrnehmung statistisch häufiger auftretender Signalkombinationen geringer sind, als die Schwellen für statistisch seltenere Signalkombinationen. Abbildung 3.8 zeigt eine Vorhersage für die erwarteten Diskriminanzschwellen. Die vier unimodalen Schwellen sind dabei auf eins normiert. Die statistisch häufiger auftretenden Stimuluskombinationen sind jene, bei denen sich beide Stimuli gleichsinnig verändern. In der Abbildung entspricht dies dem 1. und 3. Quadranten. Idealerweise werden hier Schwellen erwartet, die der optimalen Linearkombination aus den unimodalen Werten entsprechen. Für die Schwellen im 2. und 4. Quadrant wird keine Interaktion der Signale erwartet. Die Schwellen bleiben gegenüber den unimodalen Schwellen konstant. Die Abrundung an den Ecken ergibt sich aus einem statistischen Effekt.

Außer dem aus der Simulation erhaltenen, wäre z.B. auch noch ein ganz anderes Kovariationsmuster denkbar. Abbildung 3.9 zeigt eine Kovariation, bei der die beim Beobachter auftretenden Kombinationen aus Schallintensität und Netzhautabbild zwischen einem minimalen und maximalen Wert gleichverteilt sind. Für eine Umgebung, bei der alle Signalkombinationen gleich häufig auftreten, ist auch keine Veränderung der Schwellen zu erwarten. Das erwartete Ergebnis zeigt Abbildung 3.10. Die Schwelle bleibt für alle Stimuluskombinationen unverändert und gleich den unimodalen Schwellen.

Die aus der Kovariation der Simulation vorhergesagten Wahrnehmungsschwellen (Abbildung 3.8) werden im nachfolgenden Wahrnehmungsexperiment überprüft.

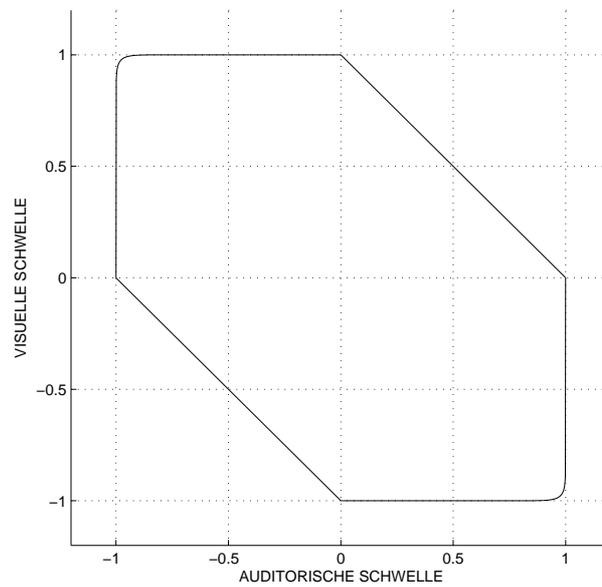


Abbildung 3.8: Die Vorhersage der Wahrnehmungsschwellen für die aus der Simulation erhaltene Kovariation der Signale. Im 1. und 3. Quadranten wird eine geringere Schwelle erwartet als im 2. und 4. Quadranten.

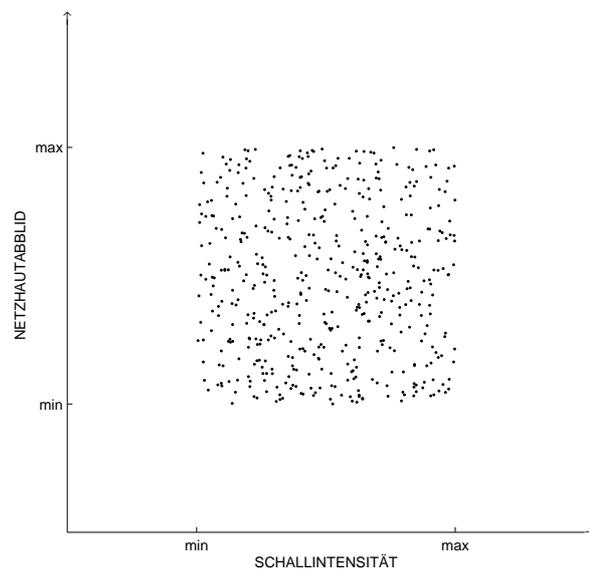


Abbildung 3.9: Beispiel eines möglichen Kovariationsmusters. Die Kombinationen aus Schallintensität und Netzhautabbild sind gleichverteilt.

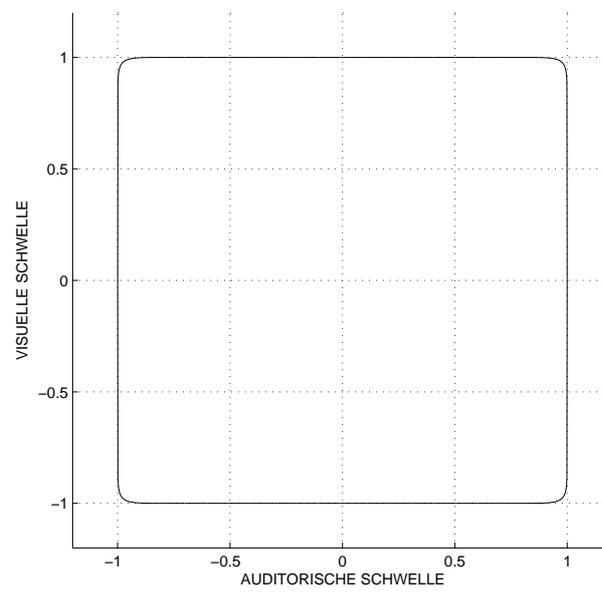


Abbildung 3.10: Vorhersage der Schwellen für ein gleichverteiltes Kovariationsmuster der Eingangssignale. Die Schwellen sind in jeder Bedingung gleich groß

Kapitel 4

Experimente zur Bewegung in Blickrichtung

In diesem Kapitel werden zwei Wahrnehmungsexperimente behandelt, bei denen der visuelle Stimulus seine Größe und der auditive Stimulus seine Amplitude verändert. Es entsteht somit der Eindruck, dass der Stimulus näher kommt oder sich entfernt. Im Abschnitt 4.1 wurden dabei sehr abstrakte Stimuli verwendet. Im nachfolgenden Experiment in Abschnitt 4.2 wurden die Stimuli aus dem ersten Experiment gegen bedeutungsvolle Stimuli ausgetauscht. Damit konnte untersucht werden, welche Auswirkung die Stimulusart auf die Wahrnehmungsleistung hat.

4.1 Abstrakte Objektbewegung

Aufbauend auf frühere Experimente [RZ00] wurde im folgenden Experiment eine audio-visuelle Schwellenmessung durchgeführt [HRSZ01]. Zum einen wurden dabei statistisch häufiger auftretende Stimuluskombinationen untersucht. Zum anderen wurden auch in einer natürlichen Umgebung eher ungewöhnliche Stimuluskombinationen verwendet. Die vier grundsätzlich aufgetretenen Kombinationen sind in Abbildung 4.1 zu sehen. Ein Objekt konnte lauter und größer werden (1.Quadrant), leiser und größer (2. Quadrant), leiser und kleiner (3. Quadrant) und schließlich kleiner und lauter (4.Quadrant). Der Pfeil kennzeichnet die in einer natürlichen Umgebung häufiger auftretenden Stimuluskombinationen. In Abbildung 4.2 wird die Vorgehensweise verdeutlicht: von einem audio-visuellen Referenzsignal aus wurde ein in einer bestimmten Richtung verändertes Testsignal erzeugt. Damit wurde für eine bestimmte Richtung der „just noticeable difference (jnd)“ ermittelt.

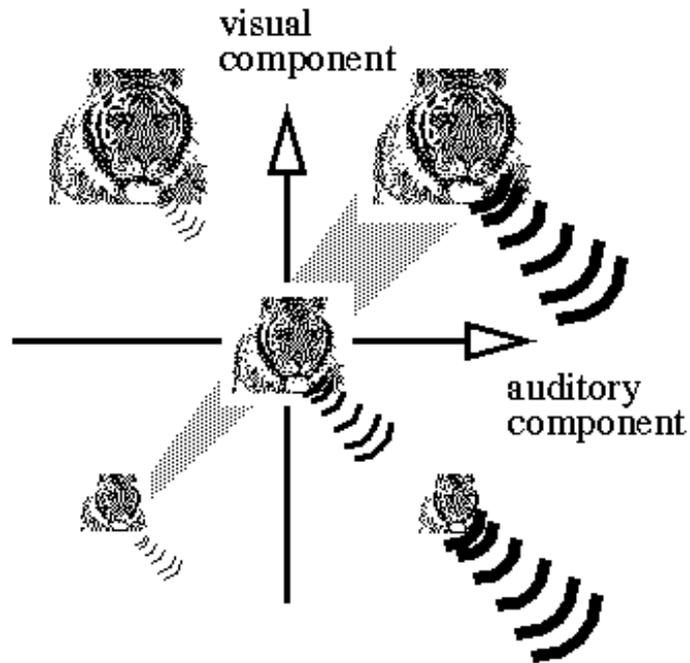


Abbildung 4.1: Die Stimuluskombination konnte sich auf vier verschiedene Arten verändern (Abbildung aus [RZ00]).

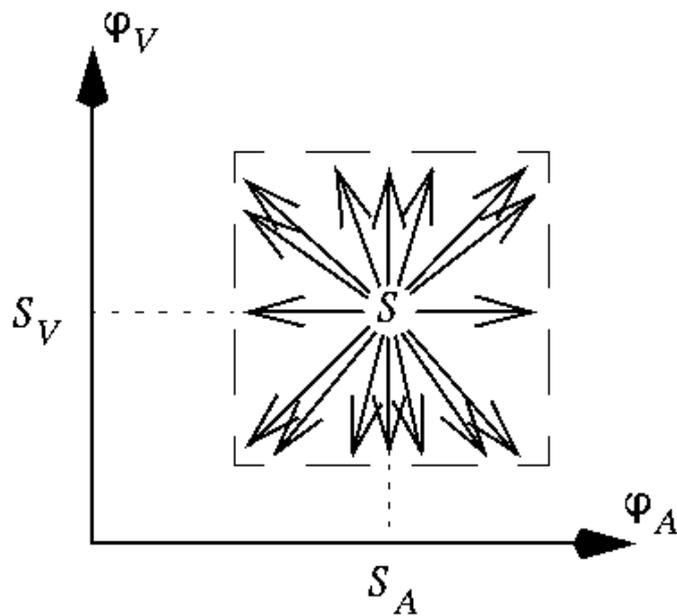


Abbildung 4.2: Ausgehend von einem audio-visuellen Referenzsignal (S) konnte das Signal in jede Richtung verändert werden und somit der kleinste wahrnehmbare Unterschied ermittelt werden.

4.1.1 Methode

Versuchsaufbau

Der Versuch fand in einem abgedunkelten Raum statt. Die Versuchsperson saß 80 cm vom einem 15 Zoll Monitor entfernt und trug Kopfhörer.

Auditiver Stimulus

Der auditive Stimulus wurde über den Kopfhörer präsentiert und bestand aus einem 1kHz Ton, dessen Amplitude variiert wurde.

Visueller Stimulus

Der visuelle Reiz wurde auf dem Monitor präsentiert und bestand aus einem grauen Quadrat, dessen Größe verändert wurde.

Durchführung

In dem „two-alternative forced-choice (2afc)“ Experiment wurden in jedem Intervall jeweils eine Kombination aus auditivem und visuellem Stimulus präsentiert. In einem Intervall blieb die Stimuluskombination über die Zeit konstant, während im anderen Intervall eine Stimulusänderung stattfand. Aufgabe des Beobachters war es, per Tastendruck zu entscheiden, in welchem Intervall eine Veränderung eines Reizes stattgefunden hat. Die Stimulusänderung konnte nur auditiv oder nur visuell oder bimodal sein. Pro Durchgang wurden zwei Schwellen mit einer adaptiven Prozedur [Tre97] ermittelt. Ein Durchgang dauerte ca. 15 Minuten. Bei 14 Probanden wurden jeweils 28 Schwellen gemessen.

4.1.2 Ergebnisse

Die ermittelten Schwellen wurden über alle Probanden gemittelt und in einem normierten Koordinatensystem aufgetragen. Die Darstellung erfolgt mit dem auditiven Wert als Abszisse und dem visuellen Wert als Ordinate. In Abbildung 4.3 sind die gemittelten und normierten Schwellen eingetragen. Die äußere Linie zeigt die gemittelte empirische Schwelle aller Probanden. Die schattierten Flächen zeigen den Abstand zur optimalen linearen Datenkombination. Die Schwellen für die inkonsistenten Stimuluskombinationen (Objekt wird größer und leiser, bzw. kleiner und lauter) im 2. und 4. Quadranten waren deutlich größer als für die

konsistenten Stimuluskombinationen (Objekt wird größer und lauter, bzw. kleiner und leiser) im 1. und 3. Quadranten.

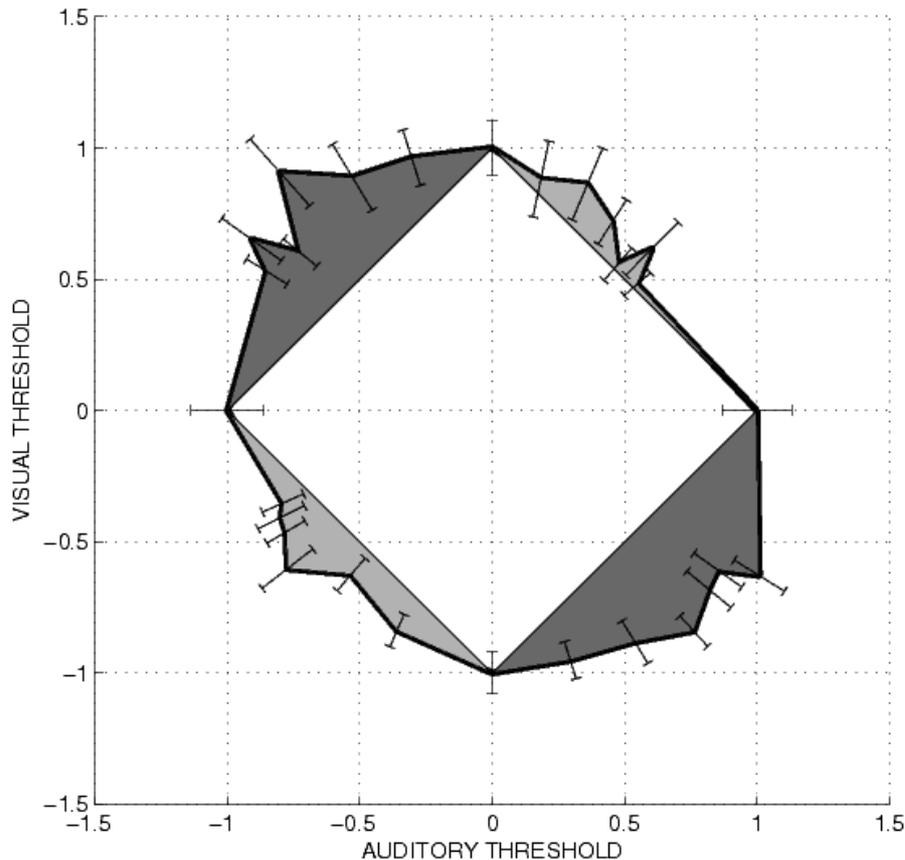


Abbildung 4.3: Die äußere Linie zeigt die gemittelten und normierten Schwellen aller 14 Probanden. Die schattierten Flächen zeigen den Unterschied zur optimalen linearen Summation. Für jeden Messpunkt sind die Standardfehler abgebildet.

In Abbildung 4.4 ist für jeden Messpunkt die Standardabweichung dargestellt. Messpunkt 1 liegt auf der positiven Ordinate. Die Messpunkte 2-7 liegen im 1. Quadranten. Messpunkt 8 liegt auf der positiven Abszisse. Die Messpunkte 9-14 liegen im 2. Quadranten. Der Messpunkt 15 liegt auf der positiven Ordinate. Die Messpunkte 16-20 liegen im 3. Quadranten. Messpunkt 21 liegt auf der negativen Abszisse und die Messpunkte 22-28 liegen im 4. Quadranten. Die Standardabweichungen waren für die Bedingungen, in denen das Objekt kleiner wurde, im Mittel etwas kleiner. Eine weitere Systematik war für die Standardabweichungen nicht erkennbar.

Die Ergebnisse wurden auch auf einen möglichen Lerneffekt untersucht. Dazu wurden die über alle Probanden gemittelten, prozentualen Abweichungen vom entsprechenden Mittelwert ermittelt und nach Sitzungen sortiert (Abbildung 4.5).

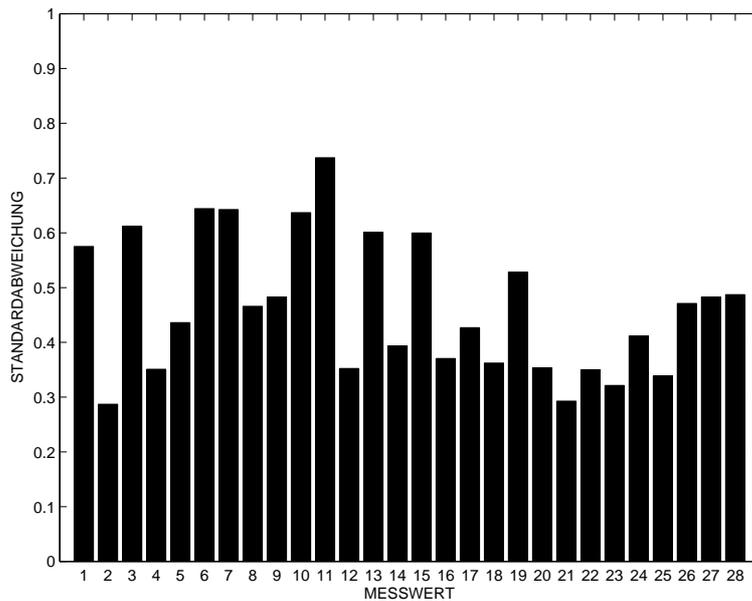


Abbildung 4.4: Die Standardabweichungen zu jedem Messwert. Für kleiner werdende Objekte (Messpunkte 16-28) zeigt sich im Mittel eine etwas kleinere Standardabweichung.

Mit diesem Verfahren konnte kein Lerneffekt festgestellt werden.

4.1.3 Modellierung

Es wurden ein neuronaler und ein evidenzbasierter Ansatz untersucht.

Neuronaler Modellansatz

Es gibt zwei Basismodelle zur Integration multisensorischer Daten. Das Modell mit Wahrscheinlichkeitssummation, sowie das Modell mit neuronaler Summation, kann jeweils nur einen Teilaspekt des Gesamtsystems beschreiben. Dieses Problem kann mit einem hybriden Modell gelöst werden, das auf einer ON/OFF-Repräsentation aufbaut [RZ00]. Die Informationsintegration wird damit vorzeichenabhängig, womit das unterschiedliche Systemverhalten in den vier Quadranten dargestellt werden kann. Die Funktion des hybriden Modells, Abbildung 4.6, läßt sich mit einem neuronalen Netz nachbilden.

Bei positiven Vorzeichen beider Signale wird der ON-Signalweg aktiviert. Beide Signale werden im oberen Zweig zuerst aufsummiert und anschließend detektiert. Bei negativen Vorzeichen beider Signale wird der OFF-Signalweg aktiviert und

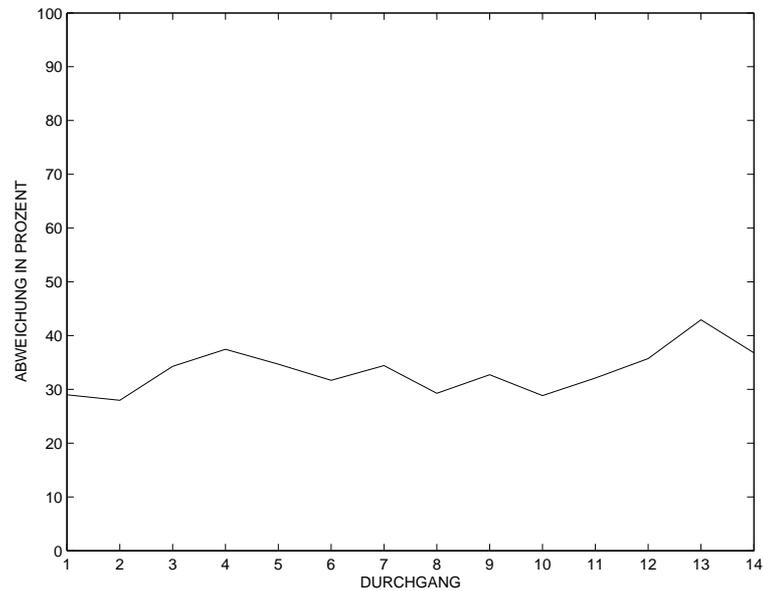


Abbildung 4.5: Die prozentualen gemittelten Abweichungen aller Probanden von den jeweiligen Mittelwerten aller Schwellen sortiert nach Sitzungen. Falls die Abweichungen im Laufe der Sitzungen kleiner werden, könnte dies auf einen Lerneffekt hindeuten. Dies war jedoch nicht der Fall.

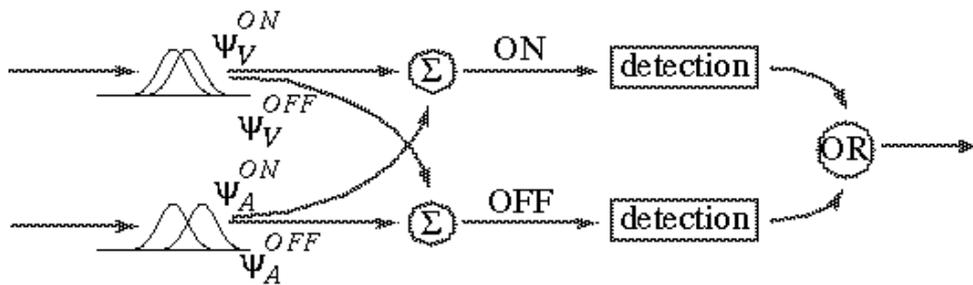


Abbildung 4.6: Hybrides Modell. Eingangssignale mit gleichen Vorzeichen (beide Signale im ON- oder OFF-Zweig) führen zu neuronaler Summation. Signale mit unterschiedlichen Vorzeichen (ein Signal im ON-Zweig, das andere Signal im OFF-Zweig) führen zur Wahrscheinlichkeitssummation.

beide Signale im unteren OFF-Zweig vor der Detektion aufsummiert. Bei gleichen Vorzeichen der Eingangssignale findet daher eine neuronale Summation der Werte statt. Bei unterschiedlichen Vorzeichen aktiviert eine Modalität das ON-System und die andere Modalität das OFF-System. Die Detektion des Signals findet für jeden Kanal einzeln statt. Damit ergibt sich eine Wahrscheinlichkeitssummation der Werte. Für die Vorhersage der Schwellen in diesem Experiment ergibt sich, je nach Quadrant, entweder eine gerade (wie in Abbildung 2.3) oder gekrümmte Linie (wie in Abbildung 2.6).

Dempster-Shafer Evidenztheorie

Alternativ zum vorherigen „neuronalen“ Ansatz wurde untersucht, wie die Reaktionen der Probanden im Rahmen eines Formalismus zur Evidenzkombination modelliert werden können. Das folgende Modell wurde in Kapitel 2.3.4 vorgestellt. Das Ergebnis des Modells bezieht sich dabei auf die beiden Aussagen: *es wurde eine Veränderung bemerkt* und *es wurde keine Veränderung bemerkt*. Im Kontext des Dempster-Shafer Ansatzes [Sha76] sind damit die zwei folgenden „internen“ Hypothesen denkbar: *Objekt kommt näher* (H^+) und *Objekt entfernt sich* (H^-). Aus der Potenzmenge ergeben sich noch H^{+-} (Vereinigungsmenge) und H^\emptyset , die leere Menge. Die Bedeutung dieser beiden Mengen wird wie folgt definiert:

- H^{+-} : Das Objekt kommt näher oder entfernt sich. Also: **Das Objekt bewegt sich.**
- H^\emptyset : **Ich kann die Frage nicht beantworten.**

Somit bezieht sich H^{+-} auf die Aussage *es wurde eine Veränderung bemerkt* und die leere Menge auf die Aussage *es wurde keine Veränderung bemerkt*. Dies entspricht somit der Fragestellung im Experiment. Dort musste sich der Beobachter ebenfalls entscheiden, in welcher der beiden Reizdarbietungen er eine Veränderung bemerkt hat. Das Modell besitzt einen auditiven und visuellen Sensor. Die Sensoren registrieren eine Veränderung der physikalischen Größen Schallintensität, bzw. projizierte Fläche des Objekts. Die möglichen Sensorwerte bewegen sich zwischen -1 und $+1$. Für den auditiven Sensor bedeutet ein kleiner positiver Wert eine geringe Zunahme der Schallintensität und ein großer positiver Wert eine starke Zunahme der Schallintensität. Negative Werte kennzeichnen eine Abnahme der Schallintensität. In Abbildung 4.7 sind die möglichen Sensorwerte für den auditorischen Sensor auf der Abszisse angetragen. Abhängig vom Sensorwert werden nun mittels der m-Funktionen den vier Mengen jeweils ein Wert zugewiesen. Dieser Wert ist das „basic probability assignment“ (bpa) auf der Ordinate und dient der Erfassung der Elementarevidenzen des Systems aufgrund des auditiven Eindrucks. Die Summe der bpa's addiert sich an jeder Stelle der Abszisse zu eins. Die Evidenzwerte für H^+ steigen für positive Sensorwerte. Analog dazu

steigen die Evidenzwerte für H^- für negative Sensorwerte. Für extreme Sensorwerte steigt die Evidenz für H^{+-} und die leere Menge bekommt für geringen Sensorauschlag hohe Evidenzwerte.

In gleicher Weise werden die Evidenzwerte des visuellen Größenänderung für den visuellen Sensor vergeben.

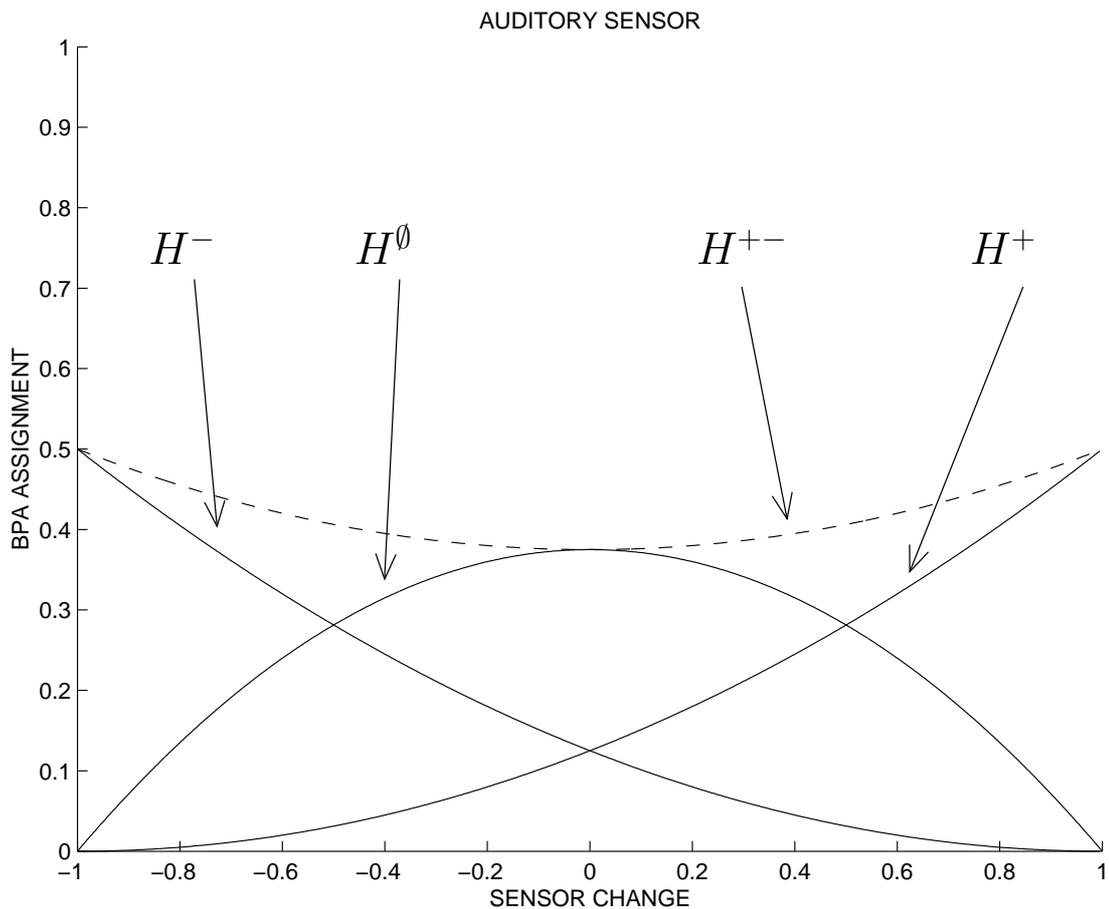


Abbildung 4.7: Die Werte des bpa wurden für jede Hypothese (H^0 , H^+ , H^- , H^{+-}) abhängig vom Sensorwert vergeben. Positive Werte auf der Abszisse stehen für eine Zunahme, negative Werte für eine Abnahme der vom Sensor registrierten Schallintensität.

Werden nun die bpa's beider Sensoren für alle möglichen Sensorwerte kombiniert und die Glaubensfunktion $Bel()$ berechnet, ergibt sich pro Hypothese eine dreidimensionale Figur. Abbildung 4.8 zeigt die Werte der Glaubensfunktion für alle vier Teilhypothesen. Die Werte für $Bel(H^0)$ sind am Höchsten, wenn beide Sensorwerte nahe Null sind. Für positiven Ausschlag beider Sensoren steigt $Bel(H^+)$. Analog steigt $Bel(H^-)$ für negative Werte beider Sensoren. Und für große gleichsinnige Werte der Sensoren erhält man hohe Werte für $Bel(H^{+-})$.

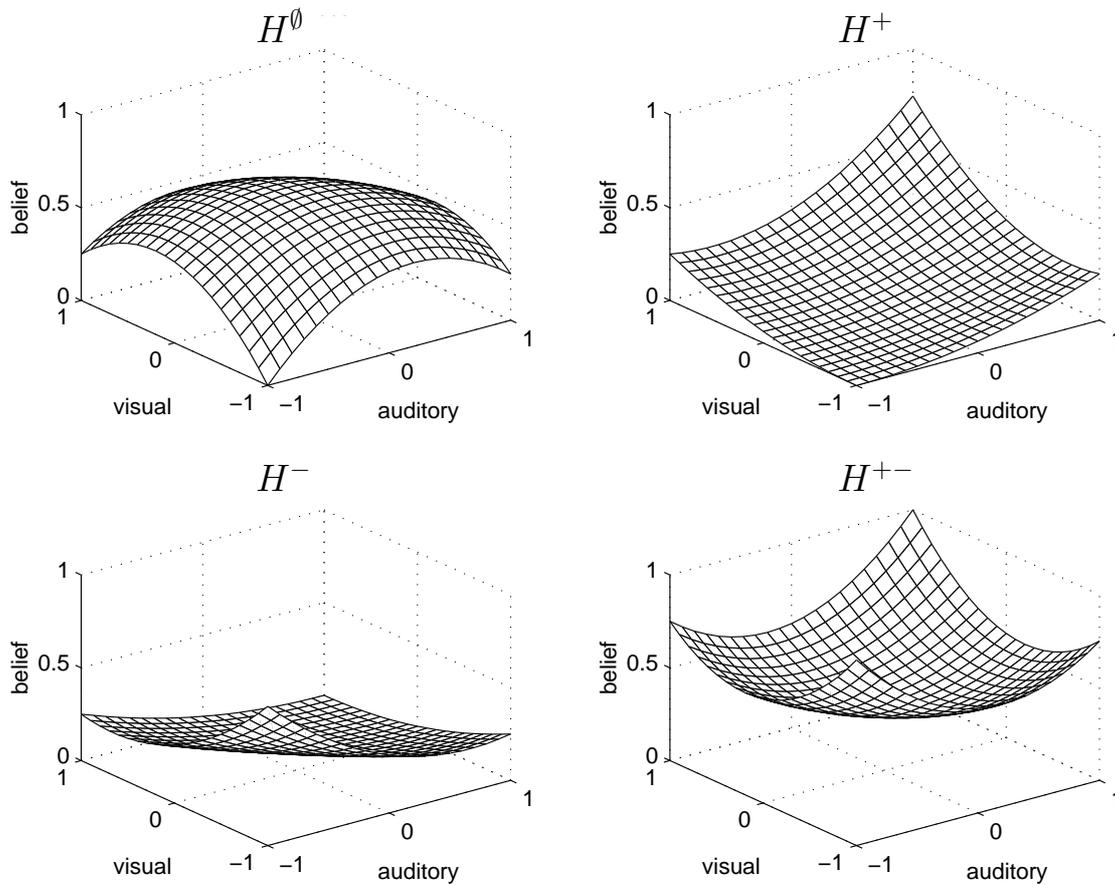


Abbildung 4.8: Die Belief-Funktionen für die vier Hypothesen (H^0, H^+, H^-, H^{+-}).

In Abbildung 4.9 sind die Kurven für H^{+-} und H^0 übereinander gelegt. Damit erhält man für jedes auditiv-visuelles Wertepaar $(a; v)$ mit $a, v \in [-1; 1]$ jeweils einen Wert für $Bel(H^0)$ und $Bel(H^{+-})$. Mit der Funktion $\max(Bel(H^{+-}), Bel(H^0))$ kann nun für jede Kombination der auditiven und visuellen Sensorwerte der höchste Glaubenswert ermittelt werden. Für große Wertepaare $(a; v)$ erhält man als Lösung einen Wert der Funktion H^{+-} und für kleine Wertepaare erhält man einen Wert aus der Funktion H^0 . Die Funktion $\max(Bel(H^{+-}), Bel(H^0))$ liefert z.B. für das Wertepaar $(a; v) = (0.3; 0.2)$ einen Wert aus H^0 zurück und für $(a; v) = (0.8; 0.9)$ einen Wert aus $Bel(H^{+-})$. Dementsprechend kann man für starke Sensorauschläge die Aussage *es wurde eine Veränderung bemerkt* unterstützen und für schwache Sensorauschläge die Aussage *es wurde keine Veränderung bemerkt*. Die Schnittlinie der beiden Ebenen bildet somit die Schwelle, an der keiner Aussage der Vorzug gegeben werden kann.

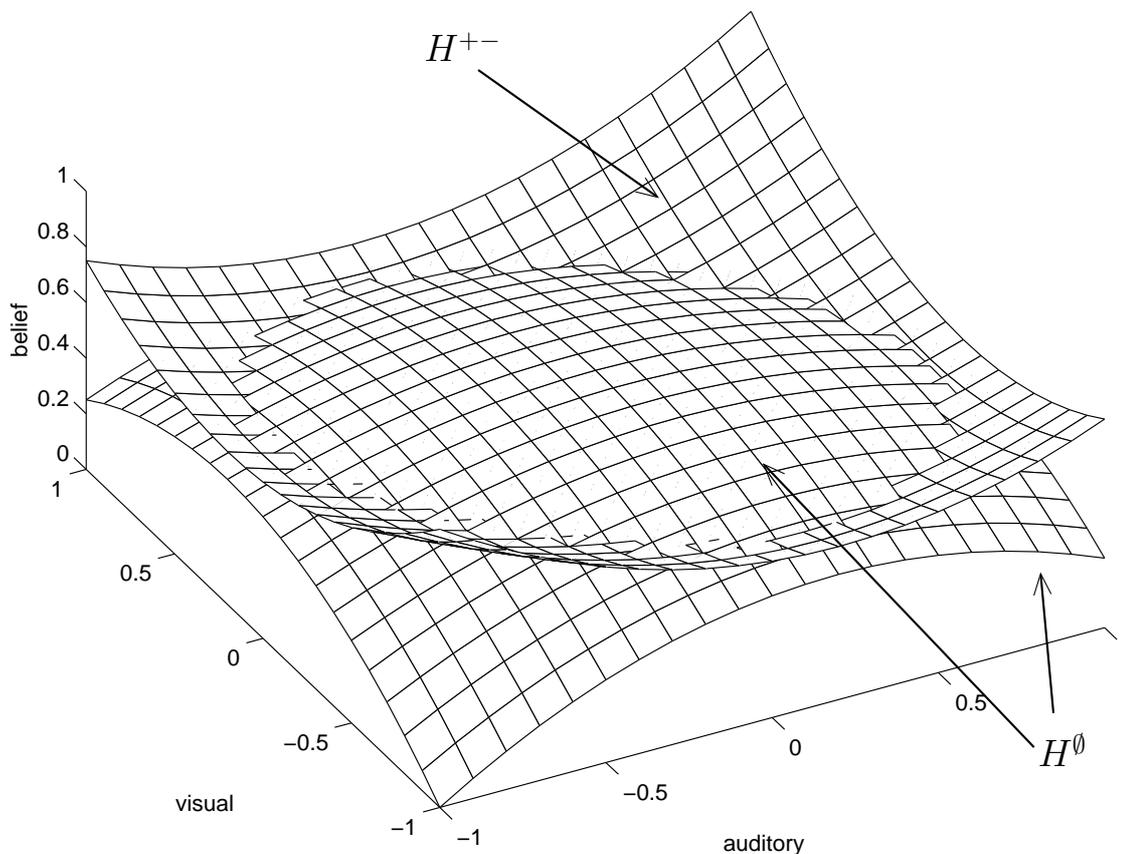


Abbildung 4.9: Für große Werte der Paare $(a;v)$ erhält man mit der Maximum-Operation als Lösung einen Wert der Funktion H^{+-} und für kleine Werte der Paare erhält man einen Wert aus der Funktion H^{\emptyset} . Die Schnittlinie zwischen den beiden Belief-Funktionen $Bel(H^{+-})$ und $Bel(H^{\emptyset})$ bildet die Schwelle.

Die Projektion der Schnittlinie aus der Dempster-Shafer-Theorie ist in Abbildung 4.10 als durchgezogene Linie zu sehen. Die Sternchen zeigen die gemessenen Schwellen aus dem Experiment. Die gestrichelte Linie zeigt den Schwellenverlauf für das hybride Modell. Die Werte auf den Achsen wurden auf das Modell normiert.

Die Form der Kurve für das Dempster-Shafer-Modell ist abhängig von der Funktion, welche die Sensorwerte auf die Evidenzwerte abbildet (Abbildung 4.7). In diesem Fall wurden dafür Parabeln verwendet. Eine quadratische Funktion beschreibt auch die Zusammenhänge der physikalischen Parameter: Sowohl die Größe des Netzhautbildes, als auch die Schallintensität eines von einer Person beobachteten Objekts, ändern sich quadratisch mit der Entfernung zwischen Beobachter und Objekt.

Für das hybride Modell wurde im 2. und 4. Quadranten die in Kapitel 2.3.3 vorgestellte Weibull-Funktion verwendet. Zum Erreichen der Krümmung wurde der Exponent $\beta = 4$ gesetzt. Der gerade Kurvenverlauf im 1. und 3. Quadranten ist durch das Linear-Summation-Modell aus Kapitel 2.3.2 festgelegt. Der Fehler zwischen Modell und Daten ist für das Dempster-Shafer-Modell signifikant kleiner (Wilcoxon Rangsummentest, $p < 0.01$) als für das hybride Modell.

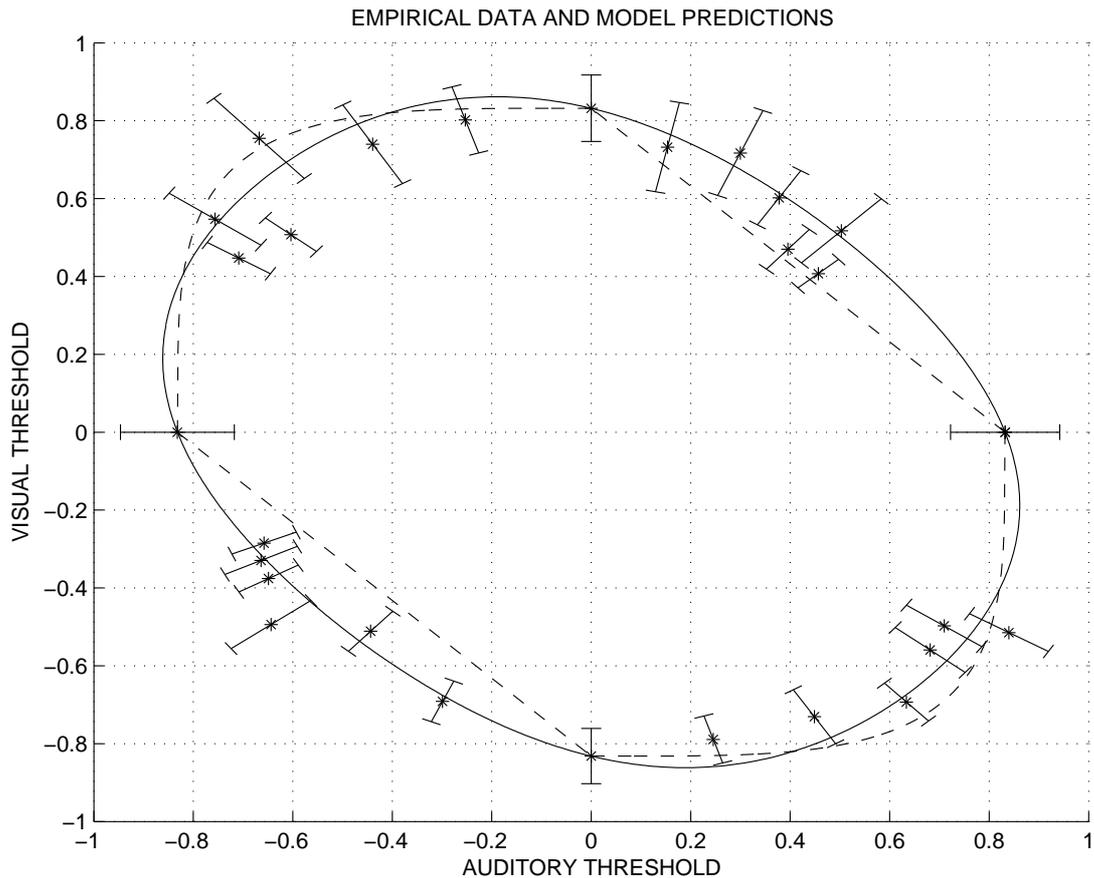


Abbildung 4.10: Die durchgezogene Kurve zeigt das Ergebnis des Dempster-Shafer-Modells. Die Sternchen zeigen die normierten Ergebnisse des Experiments. Die Vorhersage des hybriden Modells ist als gestrichelte Linie dargestellt. Zu jedem Messpunkt sind die Standardfehler abgebildet.

4.1.4 Folgerungen

Die Modellannahme der linearen Summation des hybriden Modells für die gleichsinnige Veränderung der Stimuli (1. und 3. Quadrant, Abbildung 4.10) konnte durch das Experiment nicht bestätigt werden. Mit dem Dempster-Shafer-Modell

kann in diesem Experiment ein Bezug zwischen messbaren physikalischen Größen (Änderung der Schallintensität und Änderung der Größe des Netzhautabbilds) und verschiedenen Hypothesen (H^0 : Keine Antwort möglich, H^+ : Objekt kommt näher, H^- : Objekt entfernt sich, H^{+-} : Objekt bewegt sich) hergestellt werden. Die Beziehung zwischen den physikalischen Größen und den Hypothesen mittels Parabel-Funktionen ist unmittelbar physikalisch nachvollziehbar. Durch die Verknüpfung der Hypothesen mit der Kombinationsregel (Abschnitt 2.3.4) entstehen zweidimensionale Hypothesenfunktionen. Die Definition der Schwelle erfolgt intuitiv als Schnittlinie zwischen der Hypothese H^{+-} (Objekt bewegt sich) und H^0 (Keine Antwort möglich). Durch unterschiedliche Parametrisierung der Hypothesenfunktionen kann die Form der Schwelle in jedem einzelnen Quadranten beeinflusst werden. Mit dem hybriden Modell hingegen wird die Schwellenkurve für jeweils zwei Quadranten festgelegt. Der Dempster-Shafer-Ansatz kann die empirisch gewonnenen Daten in allen vier Bedingungen deutlich besser abbilden als das hybride Modell. Die Dempster-Shafer-Theorie ist demnach ein adäquates Modell zur Abbildung der empirischen Daten dieses Experiments.

4.2 Natürliche Objektbewegung

Im folgenden Experiment wurde untersucht, ob die Art der Stimuli die Wahrnehmungsleistung beeinflusst. Mit der forced-choice Bedingung wird das Wahrnehmungssystem der Probanden bereits nahe an seinen Auflösungsgrenzen untersucht. Bei diesem Experiment wurde durch eine zusätzliche Semantik der Stimuli versucht, das Wahrnehmungssystem weiter auszureizen und die Grenzen des vorangegangenen Experiments zu erweitern.

4.2.1 Methode

Versuchsaufbau

Der Versuchsaufbau hat sich gegenüber Versuch 4.1 nicht geändert.

Auditiver Stimulus

Der auditive Stimulus bestand nun aus einem Tigergebrüll, dessen Amplitude variiert wurde.

Visueller Stimulus

Der visuelle Reiz wurde zu einem Tigerkopf verändert, dessen Größe verändert wurde.

Durchführung

Die Durchführung orientierte sich ebenfalls an Versuch 4.1

Bei 11 Probanden wurden jeweils 16 Schwellen gemessen. Dabei wurden diejenigen Schwellen nicht mehr gemessen, welche nahe an den unimodalen Achsen lagen. Aus Versuch 4.1 geht hervor, dass diese Schwellen den geringsten Einfluss auf die Form der Ergebniskurve haben.

4.2.2 Ergebnisse

Abbildung 4.11 zeigt die gemessenen Schwellen. Es wurde über alle 11 Probanden gemittelt. Die schattierten Flächen zeigen den Unterschied zu einer optimalen linearen Datenintegration. Qualitativ ähnlich zu Experiment 4.1 zeigte sich auch hier ein erheblicher Unterschied zwischen den konsistenten (1. und 3. Quadrant) und inkonsistenten Bedingungen (2. und 4. Quadrant). Die Probanden hatten in den konsistenten Bedingungen deutlich geringere Schwellen.

4.2.3 Modellierung

Die Ergebnisse wurden wieder mit der Dempster-Shafer-Methode modelliert. Abbildung 4.12 zeigt die entsprechenden Kurven. Die durchgezogene Linie zeigt das Ergebnis der Dempster-Shafer-Methode. Die Sternchen zeigen die jeweils auf die Achsen normierten Ergebnisse des Versuchs. Die gestrichelte Linie zeigt das Ergebnis des hybriden Modells.

Wie in Experiment 4.1 ist der Fehler zwischen Modell und Daten für das Dempster-Shafer-Modell signifikant kleiner ($p < 0.01$) als für das hybride Modell.

Auffallend ist die asymmetrische Verteilung der Schwellenwerte. Diese kann durch die Verwendung unterschiedlich parametrisierter Eingangsfunktionen für den Dempster-Shafer-Ansatz nachgebildet werden. Abbildung 4.13 zeigt die bpa-Funktionen für den auditiven Sensor. Die Hypothese H^+ (Objekt kommt näher) erhält dabei etwas weniger Gewicht als die Hypothese H^- (Objekt entfernt sich). Dadurch bildet sich in Abbildung 4.12 eine entsprechend höhere Schwelle für eine positive Änderung des auditiven Signals aus. Entsprechend wird die Hypothese

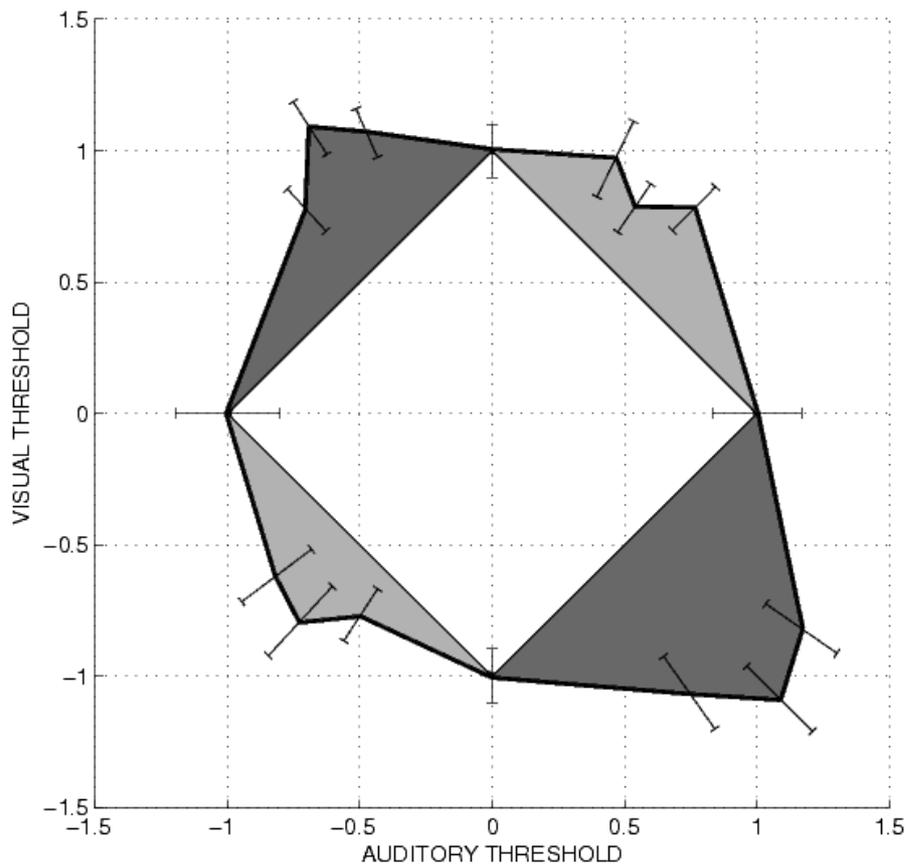


Abbildung 4.11: Die äußere Linie zeigt die gemittelten und normierten Schwellen aller 11 Probanden. Die schattierten Flächen zeigen den Unterschied zur linearen Summation.

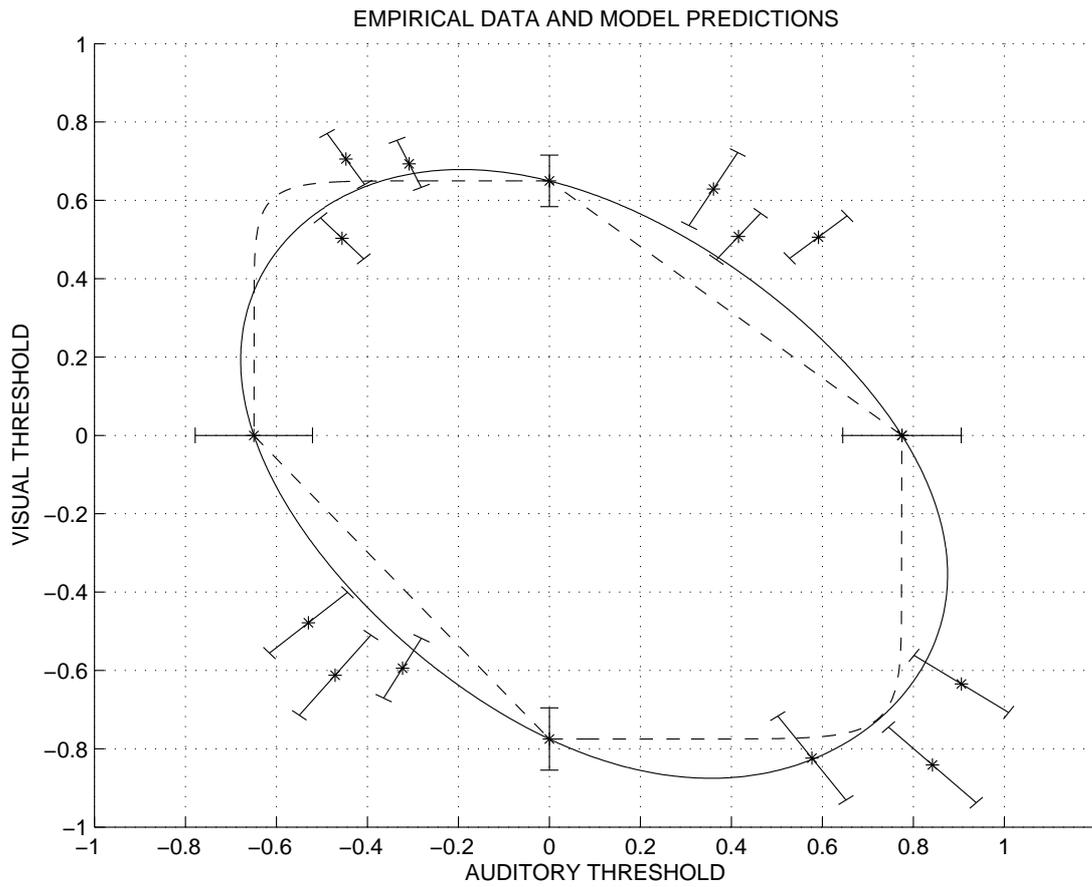


Abbildung 4.12: Die durchgezogene Linie zeigt das Ergebnis des Dempster-Shafer-Modells. Die Sternchen zeigen die normierten Ergebnisse des Experiments. Die Vorhersage des hybriden Modells ist als gestrichelte Linie dargestellt.

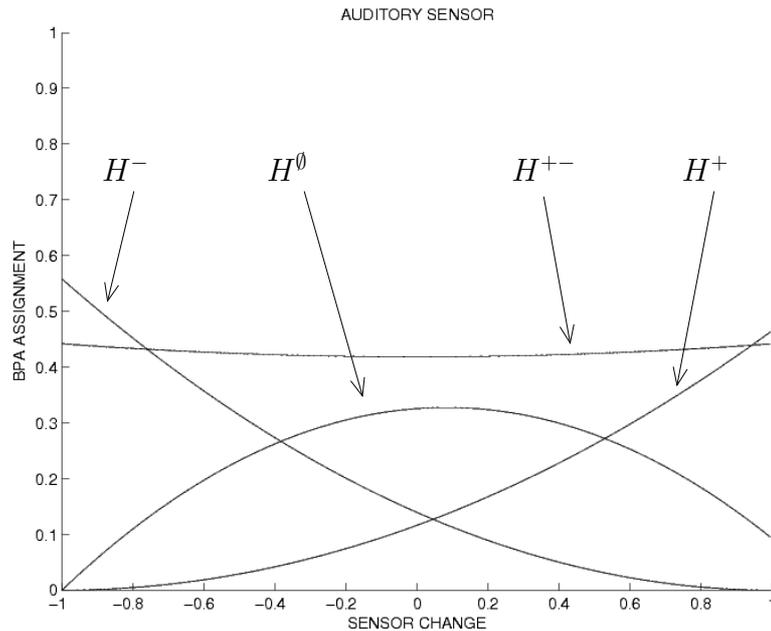


Abbildung 4.13: Die Werte des bpa werden für jede Hypothese (H^0, H^+, H^-, H^{+-}) abhängig vom Sensorwert vergeben. Für den auditiven Sensor bekommt die Hypothese H^- mehr Gewicht als H^+ .

H^- für den visuellen Sensor (Abbildung 4.14) etwas schwächer gewichtet als die Hypothese H^+ . Dadurch wird die negative visuelle Schwelle in Abbildung 4.12 erhöht. Dies führt insgesamt zu einer Zunahme der Schwelle im 4. Quadranten und somit zu einer besseren Anpassung des Dempster-Shafer-Modells an die empirischen Daten.

4.2.4 Folgerungen

Die gemessenen Schwellen lagen insgesamt etwas höher als im vorherigen Experiment (vgl. Abbildung 4.3 und 4.11). Durch die Verwendung von biologisch relevanten Stimuli (Tigerkopf statt Quadrat und Tigergebrüll statt Sinuston) konnte keine Verbesserung der Schwellen erzielt werden. Die Anpassung des Dempster-Shafer-Modells an die empirischen Daten gelingt nicht so gut wie in Versuch 4.1, da sich das Modell den hohen bimodalen Schwellen im 1. und 3. Quadranten nicht optimal nähern kann. Trotzdem nähert sich das Dempster-Shafer-Modell insgesamt besser an die Daten, als das hybride Modell (Abbildung 4.12). Zudem kann mit der Dempster-Shafer-Methode auch die Asymmetrie der empirischen Daten im 2. und 4. Quadranten durch die entsprechende Parametrisierung der Funktionen zur Evidenzbelegung der Hypothesen (Abbildung 4.13 und 4.14) intuitiv abgebildet werden.

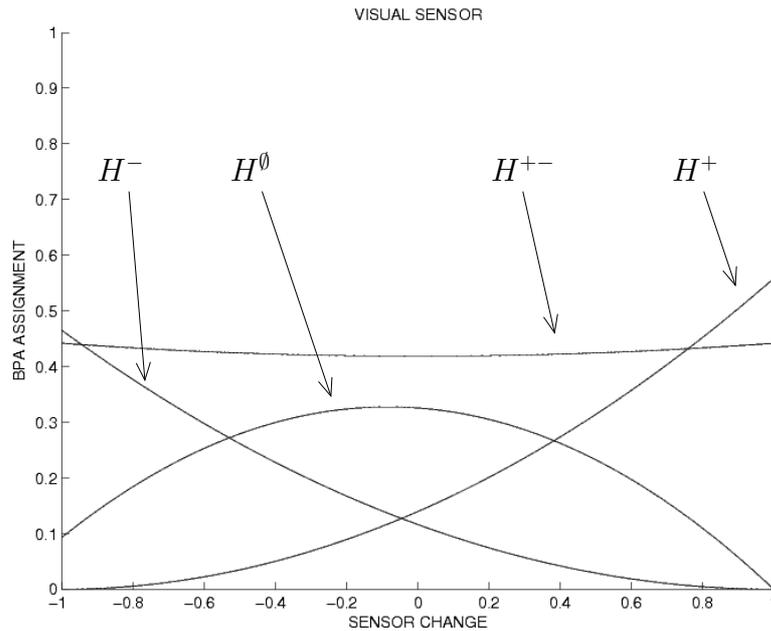


Abbildung 4.14: Die Werte des bpa werden für jede Hypothese (H^0, H^+, H^-, H^{+-}) abhängig vom Sensorwert vergeben. Für den visuellen Sensor bekommt nun die Hypothese H^+ mehr Gewicht als H^- .

Als Alternative zu den beiden Modellen würde sich das Probability-Summation-Modell für alle 4 Quadranten anbieten. Damit wäre zumindest im 1. und 3. Quadranten eine bessere Anpassung an die Daten möglich. Allerdings bietet das Probability-Summation-Modell keine Erklärung für die Asymmetrie im 4. Quadranten.

Kapitel 5

Experimente zur Bewegung in frontparalleler Ebene

In den folgenden Abschnitten werden Wahrnehmungsexperimente vorgestellt, bei denen sich die Stimuli entweder nach links oder nach rechts vor dem Beobachter bewegten. In Abschnitt 5.1 handelt es sich dabei um eine strukturlose Bewegung. Darunter ist eine Bewegung zu verstehen, die keinem Objekt zu ordnen ist. Im Gegensatz dazu stehen die Stimuli aus dem Experiment in Abschnitt 5.2, deren Ort eindeutig bestimmt werden konnte. Untersucht wurde, ob die „Objekthaftigkeit“ der Stimuli die Wahrnehmungsleistung beeinflusst.

5.1 Strukturlose Bewegung

Neurophysiologische Studien zeigen, dass im inferior und superior Colliculus Neurone zu finden sind, die selektiv auf Bewegungssignale reagieren [WMS98]. Es gibt auch Verhaltensstudien, welche zeigen, dass für überschwellige auditive und visuelle Bewegungsreize eine Interaktion der Signale stattfindet [SSL97], [ABH00]. Überschwellige auditive Bewegung [MW01] beeinflusst die Detektion visueller Bewegungswahrnehmung. Es ist unklar, ob diese Interaktion die Sensitivität für Bewegungserkennung erhöht oder nur beeinflusst. In diesem Experiment ([WHM03], [WHM⁺02]) wurde untersucht, ob audio-visuelle Integration auch die Sensitivität der Bewegungsdetektion erhöht und ob der Integrationsmechanismus richtungsselektiv ist.

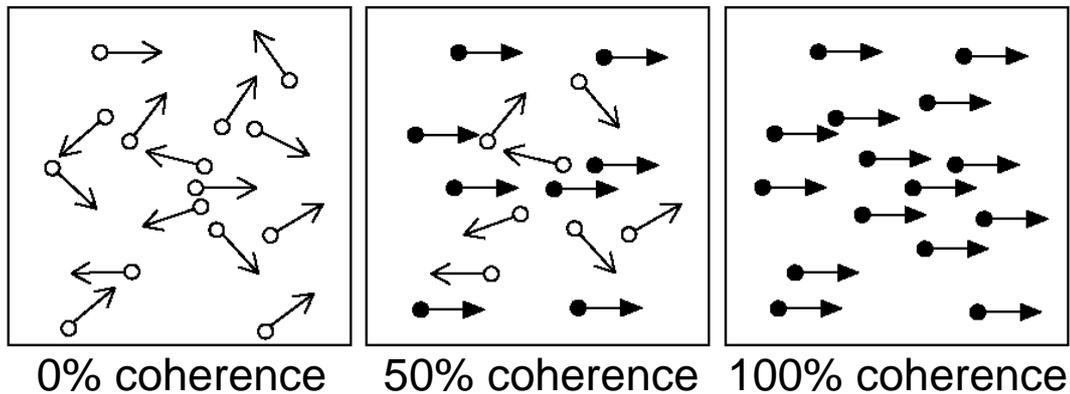


Abbildung 5.1: Der visuelle Stimulus wurde durch einen bestimmten Prozentsatz an Punkten erzeugt, die sich entweder nach rechts oder links bewegen. Die anderen Punkte bewegten sich in zufällige Richtungen.

5.1.1 Methode

Versuchsaufbau

Der visuelle Stimulus wurde mit einem Visual Stimulus Generator (VSG) 2/5 (Cambridge Research Systems), der von einem PC aus gesteuert wurde, erzeugt. Der Stimulus wurde mit einem Hitachi LCD Projektor mit einer Auflösung von 800x600 Pixeln auf eine Leinwand projiziert. Das VSG System war mit einem Tucker-Davis RP2.1 Signal Processor und HB7 Headphone Buffer (Tucker Davis) verbunden, der den auditiven Stimulus generierte. Präsentiert wurde der auditive Stimulus über zwei Lautsprecher, die sich jeweils rechts und links hinter der Leinwand befanden. Alle Experimente wurden in einem abgedunkelten, nach IAC 1204A schalldichten Raum durchgeführt. Der Abstand des Beobachters zur Leinwand betrug 200 cm.

Der visuelle Stimulus

Der visuelle Stimulus bestand aus einem Random Dot Kinematogram (RDK), Abbildung 5.1. Die relative Anzahl der Punkte, die sich in eine Richtung (nach links oder rechts) bewegten, konnte zwischen 0 (0% Kohärenz, alle Punkte bewegen sich in zufällige Richtungen) und 1 (100% Kohärenz, alle Punkte bewegen sich in die gleiche Richtung) variiert werden.

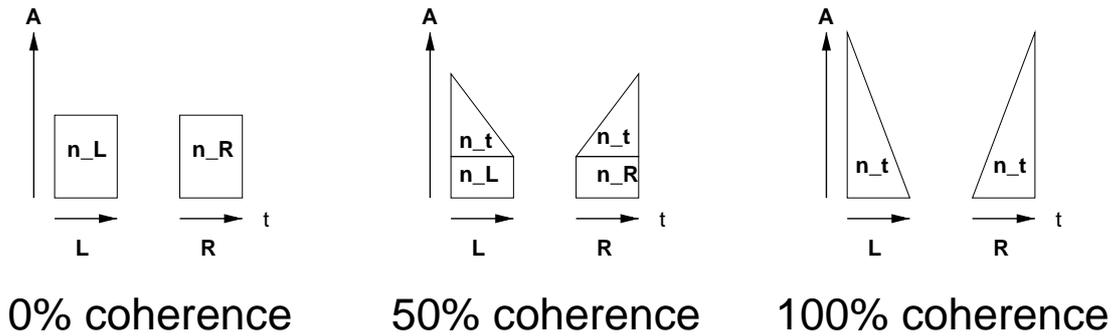


Abbildung 5.2: Durch entgegengesetztes Verändern der Amplitude des Signals über die Zeit in den beiden Lautsprechern wurde eine auditive Bewegung simuliert.

Auditiver Stimulus

Der auditive Bewegungsstimulus bestand aus weißem Rauschen, welches zwischen den beiden Lautsprechern überblendet werden konnte. Die auditive Kohärenz konnte ebenfalls zwischen 0 und 1 variiert werden. Das linke Bild in Abbildung 5.2 zeigt keine auditive Bewegung (0% Amplitudenkohärenz) und das rechte Bild zeigt maximale auditive Bewegung (100% Amplitudenkohärenz).

Durchführung

In jedem Experiment wurden alle acht möglichen Kombinationen, siehe Abbildung 5.3, von auditiven und visuellen Kohärenzleveln präsentiert. Für jede Modalität wurden sieben verschiedene Kohärenzlevel benutzt (drei positive, drei negative und keine Kohärenz). Dadurch ergaben sich 49 Stimuluskombinationen. Jede Kombination wurde jedem Beobachter pro Sitzung viermal präsentiert, wodurch sich 196 Durchgänge pro Sitzung ergaben. Jeder Beobachter absolvierte acht Sitzungen. Daraus resultierte eine Gesamtanzahl von 1568 Durchgängen pro Beobachter. Die möglichen Stimuluskombinationen wurden in konsistente (1. und 3. Quadrant), inkonsistente (2. und 4. Quadrant) und unimodale Bedingungen (auditive und visuelle Achse) aufgeteilt.

5.1.2 Ergebnisse

Die Abbildungen 5.4 und 5.5 zeigen die über fünf Beobachter gemittelten Ergebnisse. Gezeigt ist jeweils für die verschiedenen Kohärenzlevel der prozentuale Anteil korrekter Antworten. Der auditive Kohärenzlevel bewegte sich zwischen 0-0.50 (0%-50%) und die visuelle Kohärenz zwischen 0-0.15 (0%-15%). Die Er-

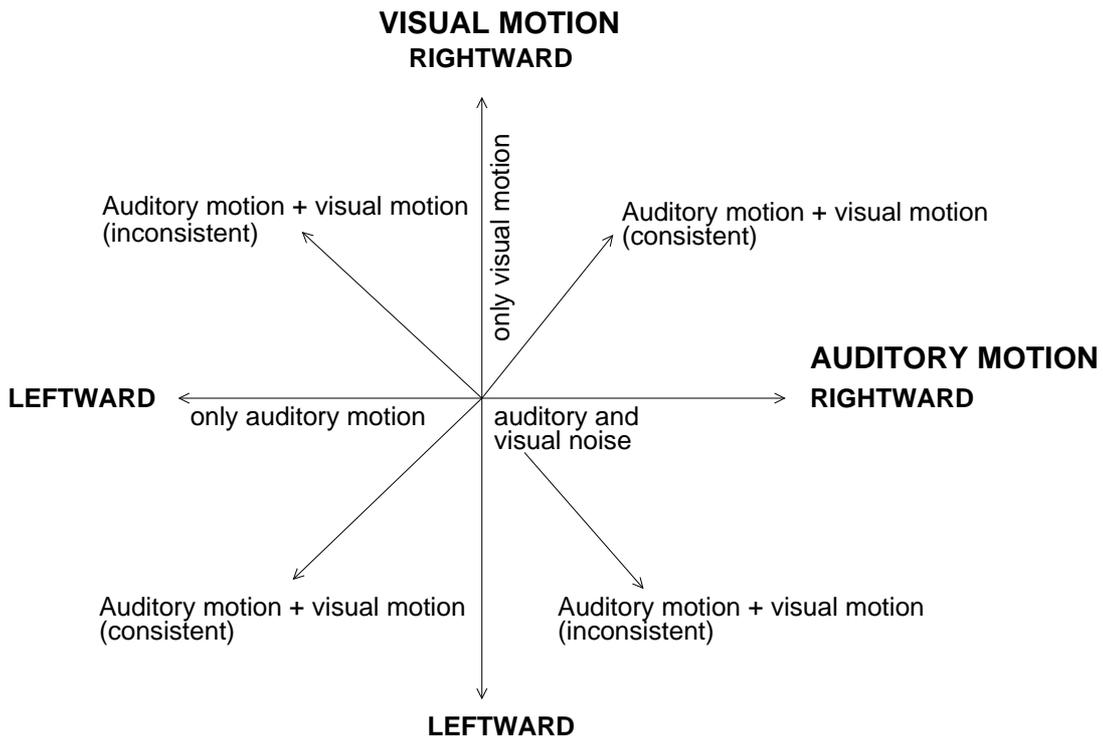


Abbildung 5.3: Mögliche Stimuluskombinationen

gebnisse bewegten sich zwischen 0.5 (50% korrekt) bei kleinen Kohärenzen und 1 (100% korrekt) bei großen Kohärenzen.

In Abbildung 5.4 ist die konsistente Bedingung zu sehen. Auditiver und visueller Stimulus bewegten sich dabei in dieselbe Richtung.

Abbildung 5.5 zeigt die inkonsistente Bedingung, bei der sich auditiver und visueller Stimulus in entgegengesetzte Richtungen bewegten.

Im nächsten Abschnitt wird nun untersucht, ob sich die empirisch ermittelten Kurven durch die in Kapitel 2.3.2 und 2.3.3 vorgestellten Modelle ausdrücken lassen und ob bei den beiden Versuchsbedingungen unterschiedliche Integrationsmechanismen zur Erklärung herangezogen werden können.

5.1.3 Modellierung

Um die Hypothesen zur audio-visuellen Integration und der Richtungsselektivität zu testen, wurden zwei verschiedene Modelle [MS80] für die Vorhersage der empirischen Daten verwendet. Die zwei Modelle unterscheiden sich in der Position, an der die audio-visuelle Interaktion angenommen wird.

Zum Anpassen der Kurven wurden als psychometrische Funktionen übliche, zweidimensionale Weibull-Funktionen [Gra89] verwendet. Für eine 2AFC-Aufgabe sa-

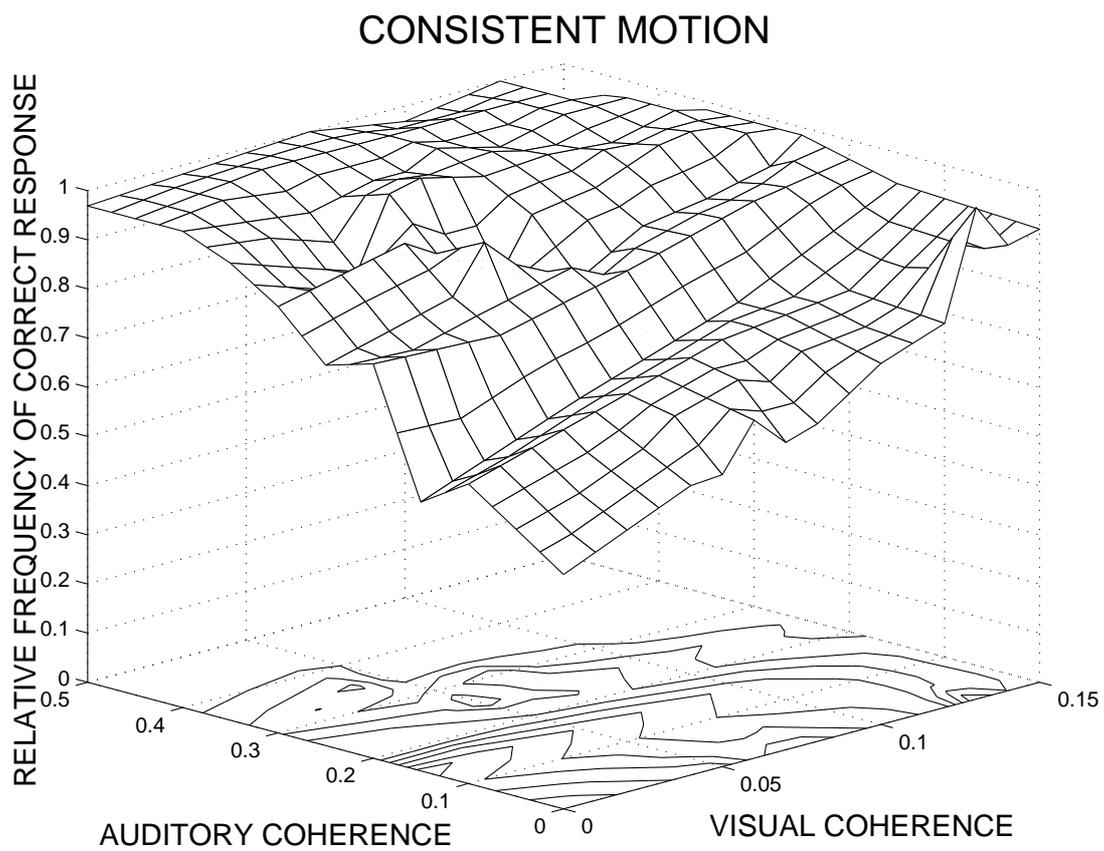


Abbildung 5.4: Psychometrische Ergebnisfunktionen für konsistente Bewegungssignale.

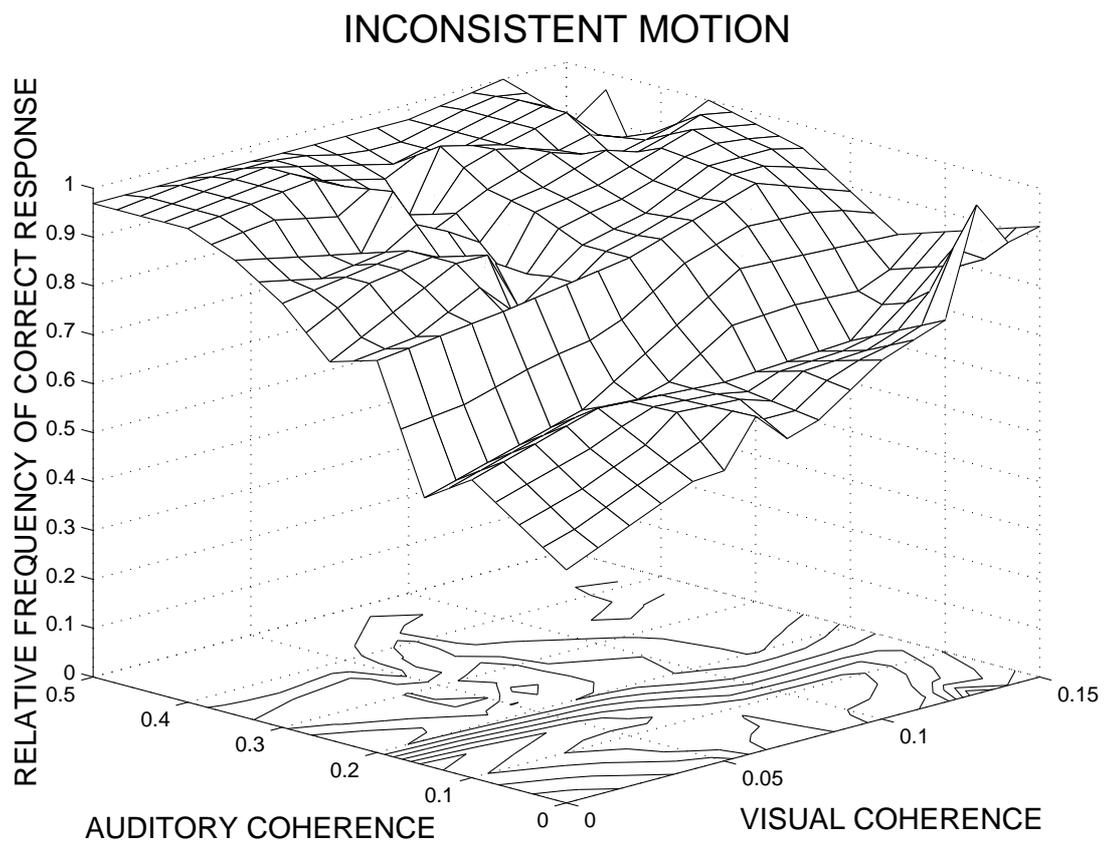


Abbildung 5.5: Psychometrische Ergebnisfunktionen für inkonsistente Bewegungssignale.

gen diese Funktionen eine Ratewahrscheinlichkeit von 0.5 und eine asymptotische maximale Leistung von 1 voraus.

Modell 1 „Linear-Summation-Modell“

Bei diesem Modell wird, wie in Abschnitt 2.3.2 gezeigt, angenommen, dass die auditiven und visuellen Signale zuerst aufsummiert werden, bevor eine Entscheidung über die An- oder Abwesenheit des Stimulus getroffen wird. Zur Modellierung wird die in Abschnitt 2.3.2 vorgestellte Weibull-Funktion verwendet:

$$P(x_{vis}, x_{aud}) = 1 - 0.5 * e^{-\left(\frac{x_{vis}}{\alpha_{vis}} + \frac{x_{aud}}{\alpha_{aud}}\right)^\beta} \quad (5.1)$$

Modell 2 „Probability-Summation-Modell“

Dieses Modell beruht, wie in Abschnitt 2.3.3 bereits dargelegt, auf der Annahme, dass beide Signale unabhängig voneinander detektiert werden und die Wahrnehmungsleistung des kombinierten Stimulus auf den unabhängigen Entscheidungen der einzelnen Kanäle beruht. Es wird die in Abschnitt 2.3.3 hergeleitete Weibull-Funktion verwendet:

$$P(x_{vis}, x_{aud}) = 1 - 0.5 * e^{-\left[\left(\frac{x_{vis}}{\alpha_{vis}}\right)^{\beta_{vis}} + \left(\frac{x_{aud}}{\alpha_{aud}}\right)^{\beta_{aud}}\right]} \quad (5.2)$$

Vorhersagen für Modell 1 und Modell 2

Das Linear-Summation-Modell (Abschnitt 2.3.2, Abbildung 2.2) sagt voraus, dass die Schwellen für den bimodalen Stimulus auf geraden Linien liegen. Das Probability-Summation-Modell (Abschnitt 2.3.3, Abbildung 2.5) sagt gekrümmte Kurven und damit größere Schwellen für den bimodalen Bewegungsstimulus voraus.

Güte der Anpassung für beide Modelle und beide Bedingungen

Die Güte der Anpassung ist in Tabelle 5.1 dargestellt. Das Probability-Summation-Modell liefert eine gute bis mäßige Anpassung an die Daten ($p=0.51$ für konsistente und $p=0.39$ für inkonsistente Bewegung). Das Linear-Summation-Modell muss hingegen sowohl für den konsistenten Fall ($p=0.08$) als auch für den inkonsistenten Fall ($p=0.01$) zurückgewiesen werden.

VP	KONSISTENTE BEWEGUNG						INKONSISTENTE BEWEGUNG					
	Subthreshold summation			Probability summation			Subthreshold summation			Probability summation		
	χ^2	df	p	χ^2	df	p	χ^2	df	p	χ^2	df	p
1	33.3	24	0.10	32.2	24	0.13	41.2	24	0.02	39.8	23	0.02
2	45.4	22	0.01	9.8	18	0.94	54.3	22	0.00	19.3	19	0.44
3	35.9	21	0.02	32.7	22	0.07	47.1	24	0.01	30.8	22	0.1
4	36.9	26	0.08	22.3	23	0.51	38.9	25	0.04	24.2	23	0.39
5	11.9	9	0.22	11.7	14	0.63	35.7	19	0.01	11.7	15	0.70
Median			0.08			0.51			0.01			0.39

Tabelle 5.1: Anpassung der Modelle an die empirischen Daten. Für beide Versuchsbedingungen (konsistente und inkonsistente Bedingungen) und beide Modelle (Subthreshold und Probability Summation) sind jeweils für alle fünf Beobachter die χ^2 -Werte (χ^2), die Freiheitsgrade (df) und der Signifikanzlevel (p) angegeben.

Geschätzte Parameter für beide Modelle und beide Bedingungen

In Tabelle 5.2 sind die geschätzten Modellparameter zu sehen. Die Modellparameter setzen sich aus auditiver (α_{aud}) und visueller (α_{vis}) Schwelle, sowie aus dem Exponenten (β) zusammen. Die Parameter α_{aud} und α_{vis} markieren die unimodalen Kohärenzlevel, bei denen mit 81 %iger Wahrscheinlichkeit eine korrekte Antwort gegeben wird. Der Exponent β ist der Parameter für die Steigung der Weibull-Funktion. Ein größeres β ergibt eine stärkere Steigung der Funktion. Es ergibt sich kein signifikanter Unterschied in den geschätzten Parametern zwischen den konsistenten und inkonsistenten Bedingungen. Dies unterstreicht die Ähnlichkeit der Daten in den beiden Bedingungen, die sich anhand eines Vergleichs der Abbildungen 5.4 und 5.5 bereits vermuten lässt.

5.1.4 Folgerungen

Ziel dieses Experiments war es, den Ort und die Art des audio-visuellen Integrationsmechanismus zu charakterisieren. Aus den vorangegangenen Experimenten in Kapitel 4 wird erwartet, dass sich in diesem Experiment für die konsistente Bedingung eine deutlich niedrigere Schwelle, als für die inkonsistente Bedingung ergibt. Neurologische Untersuchungen im Superior Colliculus [SM93] lassen für eine frühe Interaktion des auditorischen und visuellen Signals ebenfalls einen starken Effekt erwarten. Die experimentellen Ergebnisse deuten aber darauf hin, dass die Signale unabhängig von einander detektiert werden und erst auf späterer Ebene kombiniert werden. Desweiteren stellt sich heraus, dass der Integrationsmechanismus

VP	KONSISTENTE BEWEGUNG						INKONSISTENTE BEWEGUNG					
	Subthreshold summation			Probability summation			Subthreshold summation			Probability summation		
	α_{aud}	α_{vis}	β	α_{aud}	α_{vis}	β	α_{aud}	α_{vis}	β	α_{aud}	α_{vis}	β
1	0.26	0.13	1.10	0.25	0.12	1.27	0.26	0.13	1.14	0.24	0.12	1.31
2	0.31	0.14	1.78	0.26	0.11	3.34	0.30	0.13	1.80	0.26	0.11	2.82
3	0.23	0.17	2.00	0.19	0.15	2.03	0.23	0.19	1.77	0.19	0.16	2.27
4	0.52	0.13	1.52	0.35	0.12	2.44	0.62	0.13	1.71	0.37	0.12	2.62
5	0.30	0.09	1.93	0.29	0.09	2.60	0.30	0.10	1.25	0.32	0.09	2.74
MW	0.32	0.13	1.66	0.27	0.12	2.34	0.34	0.14	1.53	0.28	0.12	2.35

Tabelle 5.2: Modellparameter für die beiden Versuchsbedingungen. Für beide Versuchsbedingungen (konsistente und inkonsistente Bedingungen) und beide Modelle (Subthreshold und Probability Summation) werden jeweils für alle fünf Beobachter die auditive Schwelle (α_{aud}), die visuelle Schwelle (α_{vis}), und der Exponent (β) gezeigt.

richtungsunabhängig funktioniert. Die Richtungsunabhängigkeit der Versuchsergebnisse ist wiederum im Einklang mit dem Ergebnis, dass die Integration erst nach der auditiven und visuellen frühen Verarbeitung stattfindet [AB01]. Dies bedeutet, dass sich die Integration nicht auf ökologisch sinnvolle Kombinationen beschränkt. Unter ökologisch sinnvollen Kombinationen sind dabei jene Stimuluskombinationen zu verstehen, bei denen sich beide Stimuli in die gleiche Richtung bewegen. Dies ist auch der vorherrschende Fall in einer natürlichen Umgebung, falls angenommen wird, dass beide Signale einem Objekt zugeordnet werden. Mit einer, bei früher neuronale Interaktion, erwarteten niedrigen Schwelle, könnte bei schwachen Umgebungsreizen eine bessere Wahrnehmungsleistung erzielt werden. Andererseits würde eine zu niedrige Schwelle unter Umständen zur Wahrnehmung von für den Menschen völlig irrelevanter Reize führen. Es stellt sich die Frage, warum sich im Gegensatz zum Experiment in Abschnitt 4.1 keine Verbesserung der Schwellen in den gleichsinnigen Stimuluskombinationen ergibt. Als mögliche Ursache könnten die artifiziellen Reize (bewegte Punkte, weißes Rauschen) in Frage kommen. Es wird angenommen, dass es durch diese abstrakten Stimuli beim Beobachter nicht zu dem Perzept eines Objekts, welches hör- und sichtbar ist, kommt. Deshalb wurde im nachfolgenden Experiment Wert darauf gelegt, eine möglichst objekthafte Stimuluskombination zu realisieren.

5.2 Objektbewegung

In diesem Experiment ([HWM⁺04], [HWMS03], [HWM⁺03]) wurde untersucht, ob und auf welche Weise Informationen aus dem auditiven und visuellen System miteinander verknüpft werden. Dabei wird sowohl uni- wie auch bimodal

eine seitliche Objektbewegung simuliert. Ausschlaggebend für dieses Experiment war die Verwendung möglichst objekthafter Stimuli. Darunter ist eine Stimuluskombination zu verstehen, bei der der auditive und visuelle Reiz keine globale Bewegung, wie im Experiment in Abschnitt 5.1, vermittelt. Vielmehr soll durch die Art der Stimuli ein bewegtes **Objekt** wahrgenommen werden. Der Beobachter sollte unter verschiedenen Bedingungen den Zeitpunkt des Erscheinens des bewegten Objekts an einem bestimmten Ort bestimmen. Gemessen wurde der Zeitpunkt, zu dem der Beobachter das Objekt erwartet. Verglichen wurden die Verteilungen des Antwortverhaltens der Beobachter in den verschiedenen Modalitäten. Überprüft wurde die Präzision der Vorhersage. Diese lässt sich durch mehrere Parameter charakterisieren. Berechnet wurde die Anzahl der korrekten Antwortzeiten, der Mittelwert der Antwortzeiten und die Standardabweichung der Antwortzeiten. Eine korrekte Antwortzeit wurde erreicht, falls die Antwort des Probanden innerhalb eines bestimmten Intervalls erfolgte.

5.2.1 Methode

Versuchsaufbau

Es wurden 31 LED's horizontal und halbkreisförmig im Abstand von 10 cm montiert. Direkt unterhalb der LED's befanden sich 31 Lautsprecher (siehe Abbildung 5.6). Der Beobachter saß 1.60 m vom Mittelpunkt der bogenförmigen Anordnung in einem Stuhl. Der Versuch fand in einem abgedunkelten, schalldichten Raum statt. Die LED's und Lautsprecher wurden von einem Tucker-Davis-realtime Signal Prozessor (Tucker-Davis Technologies) gesteuert. Der Signal Prozessor war mit einem PC verbunden, mit dem die Auswertung erfolgte.

Der Echtzeitprozessor wurde über das zugehörige RP-Visual-Design-Studio programmiert. Abbildung 5.7 zeigt das Modul „Sequence-Player“. Damit wurde die in Matlab generierte Stimulussequenz entsprechend getaktet an die digitalen Ausgänge des Signalprozessors geschickt und an der bogenförmigen Anordnung abgespielt.

In Abbildung 5.8 sind die Module „Button-Detection“ und „Timer“ dargestellt. Ein Druck auf die Buttonbox wurde an einem digitalen Eingang des Signalprozessors registriert. Blieb die Antwort des Probanden aus, wurde dies ebenfalls erkannt. Nach einem Warnsignal wurde dann die entsprechende Sequenz nochmals abgespielt.

Im Timer-Modul wurde berechnet, wie groß der zeitliche Abstand zwischen dem Knopfdruck des Probanden und einem zuvor berechneten erwarteten Zeitpunkt war. Dadurch konnte die Genauigkeit der Vorhersage des Probanden über das Auftauchen des Stimulus an einer bestimmten Stelle gemessen werden.

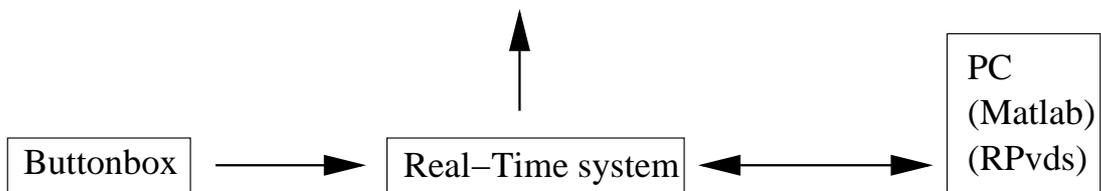


Abbildung 5.6: Durch einen Echtzeitprozessor wurde die Versuchsanordnung gesteuert. Aktionen der Probanden wurden durch die Buttonbox an den Echtzeitprozessor übertragen. Die Auswertung der Daten erfolgte mit einem PC.

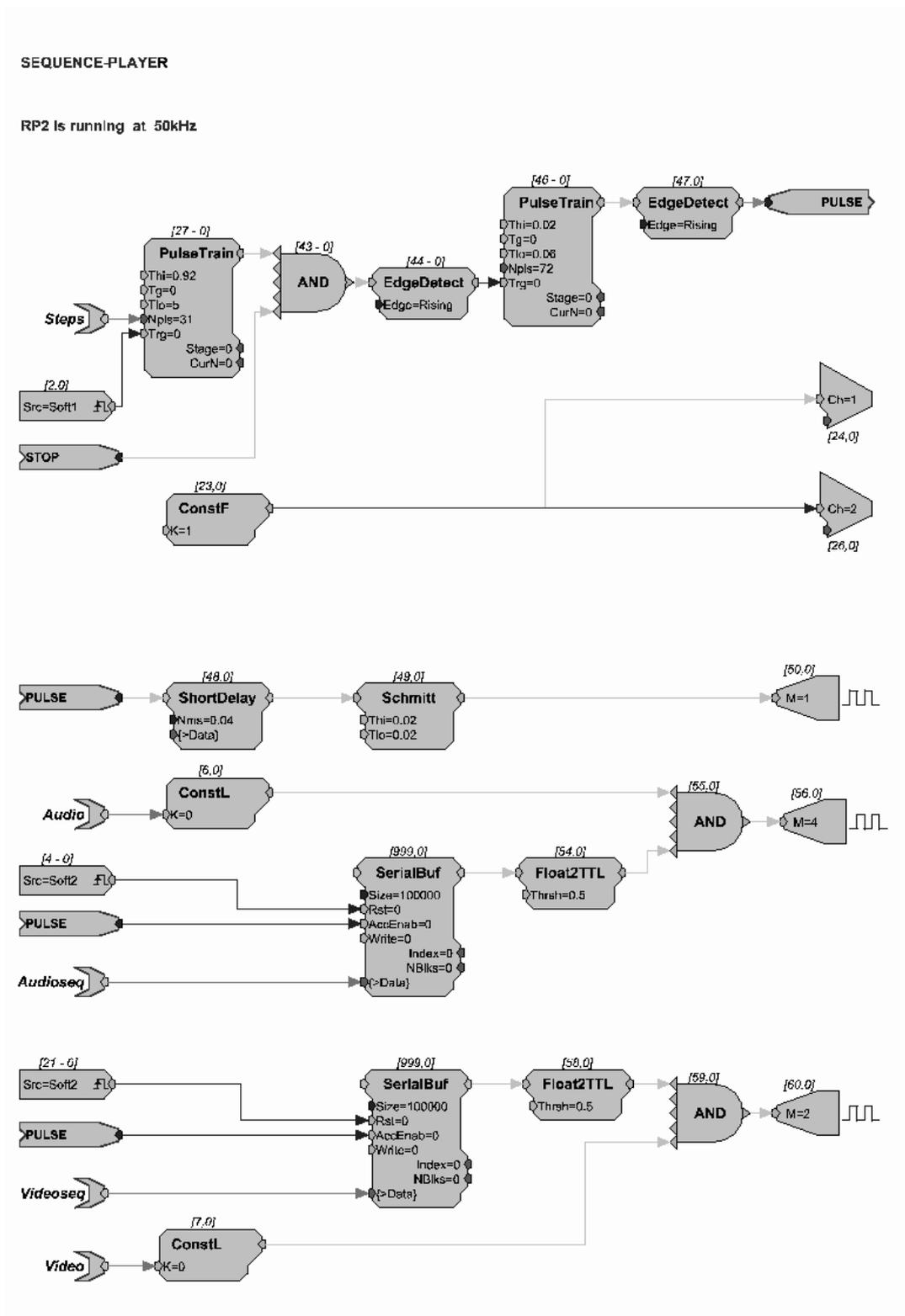


Abbildung 5.7: Mit dem Sequence-Player wurde die aktuelle Sequenz an die Lautsprecher und an die LED's übertragen.

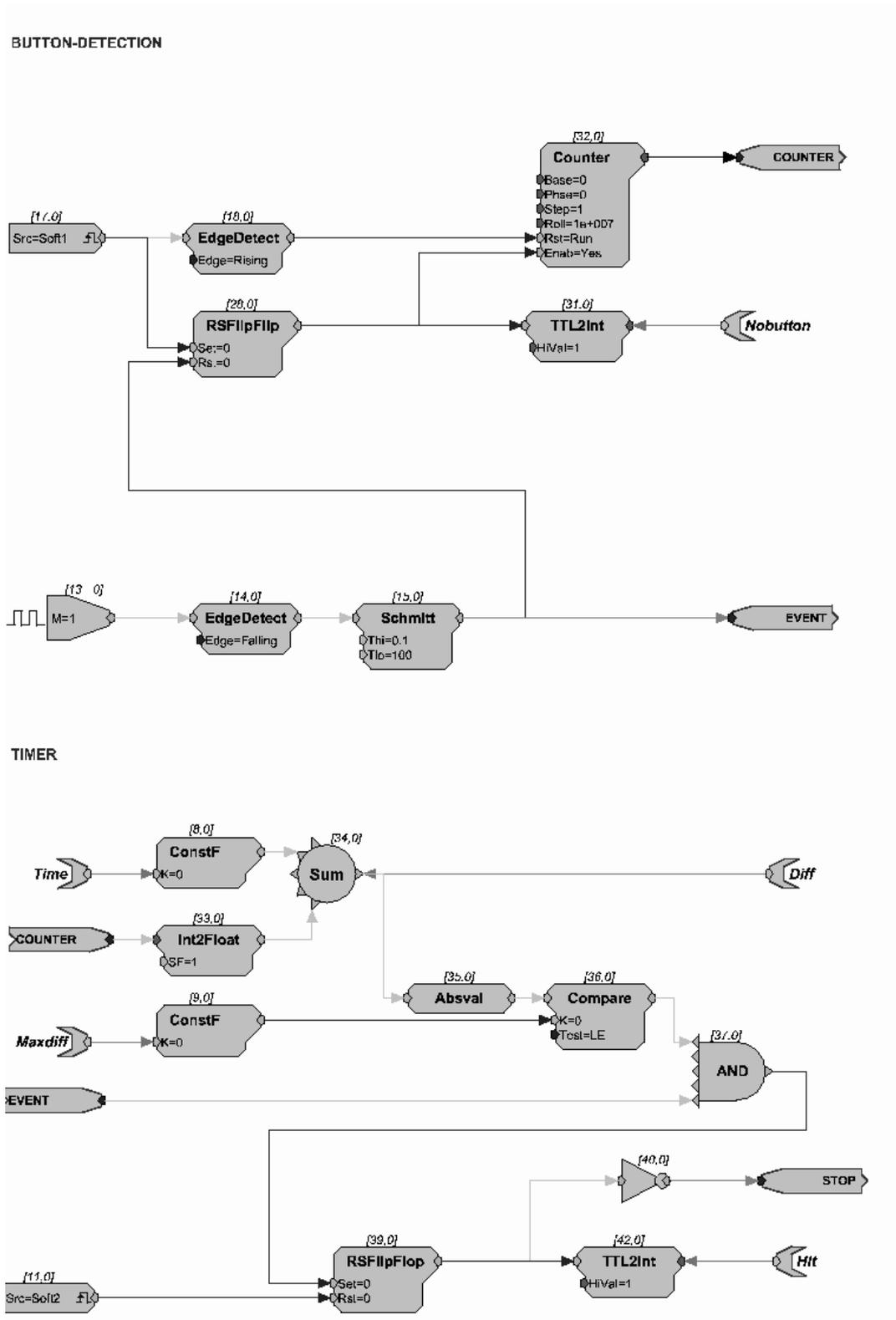


Abbildung 5.8: Mit dem Modul „Button-Detection“ wurde ein Druck auf die Buttonbox registriert. Mit dem „Timer“ wurde die Zeitdifferenz zwischen dem tatsächlichen und dem erwarteten Zeitpunkt des Drückens auf die Buttonbox ermittelt.

Visueller Stimulus

Durch zeitversetzte Ansteuerung der LED's konnte eine visuelle Objektbewegung simuliert werden.

Auditiver Stimulus

Mit den Lautsprechern konnten Klick-Geräusche erzeugt werden. Auch hier entstand durch zeitversetztes Ansteuern der Eindruck einer auditiven Bewegung.

Durchführung

Das Objekt bewegte sich von einem Ende des Bogens aus mit einer bestimmten Geschwindigkeit nach rechts oder links. Die Bewegung war entweder auditiv, visuell oder bimodal. Der Stimulus bewegte sich mit fünf verschiedenen Geschwindigkeiten (17, 23, 27, 34, 47 deg/s). Modalität, Richtung und Geschwindigkeit des Stimulus wurden zufällig ausgewählt und über eine Sitzung gleich verteilt. Der Beobachter musste per Knopfdruck den Zeitpunkt bestimmen, an dem der Stimulus genau in der Mitte der Anordnung war. Der Beobachter erhielt Rückmeldung für zu frühes oder zu spätes, als auch für korrektes (± 2 cm vom exakten Zeitpunkt) Bestimmen des Zeitpunktes. Gemessen wurde der Zeitpunkt, zu dem der Beobachter das Objekt in der Mitte schätzte, relativ zum exakten Zeitpunkt des Erscheinen des Objekts in der Mitte.

Jeder Beobachter absolvierte zuerst drei Trainingssitzungen. Danach folgten, an verschiedenen Tagen, sechs Testsitzungen mit insgesamt 720 Durchgängen. Daraus ergaben sich pro Beobachter 48 Durchgänge für jede der 15 Konditionen (fünf Geschwindigkeiten, drei Modalitäten).

Es gab zwei unterschiedliche Versuchsbedingungen:

- **Noise-Bedingung**
Bei dieser Bedingung wurde der Stimulus durch Rauschen überlagert. Dies geschah durch, über den ganzen Bogen verteiltes, zufälliges Aufblitzen von LED's und zufälliges Klicken der Lautsprecher. Durch das Rauschen wurde die Verfolgung des Stimulus erschwert. Es nahmen acht Personen (vier weibliche, vier männliche) zwischen 20 und 42 Jahren teil.
- **Tunnel-Bedingung**
Bei dieser Bedingung startete der Stimulus an einem Ende des Bogens und lief 30 cm vor der Zielposition verdeckt weiter. An der Zielposition war der Stimulus kurz hör- bzw. sichtbar und lief dann für weitere 30 cm verdeckt weiter, bis er wieder auftauchte. Dadurch war der Beobachter gezwungen,

die Geschwindigkeit des Stimulus im Tunnel zu schätzen, um dessen Erscheinen an der Zielposition zu bestimmen. An diesem Versuch nahmen zehn Personen (drei weibliche, sieben männliche) zwischen 18 und 48 Jahren teil.

5.2.2 Ergebnisse

Es sollte die Leistung der Probanden unter den verschiedenen Bedingungen gemessen werden. Dazu wurden drei Leistungsmerkmale ausgewertet:

1. Trefferquote
Aufgabe der Probanden war es, so viele Objekte wie möglich zu „fangen“. Ein Treffer wurde erzielt, wenn ein Proband den Knopf innerhalb des räumlichen Fensters von ± 2 cm von der Zielposition des Objekts drückte.
2. Mittlere geschätzte Ankunftszeit
Dieses Merkmal lieferte den mittleren Fehler zwischen dem tatsächlichen Auftauchen des Objekts an der Zielposition und dem Zeitpunkt des Knopfdrucks des Probanden.
3. Standardabweichung
Für die verschiedenen Bedingungen wurde jeweils die Standardabweichung der Antwortzeiten der Probanden ermittelt. Dies war ein Maß für die Variabilität der Antworten der Probanden.

Die folgenden sechs Grafiken zeigen die Ergebnisse aus den zwei unterschiedlichen Versuchsbedingungen. Jede Grafik zeigt die entsprechenden Mittelwerte über alle 8 bzw. 10 Beobachter. Die Abszisse ist in fünf Blöcke unterteilt. Diese gruppieren die fünf Geschwindigkeiten. Im linken Block stehen die Ergebnisse der langsamsten, im rechten Block stehen die Ergebnisse der schnellsten Stimuli. Innerhalb eines Blocks sind die Ergebnisse noch nach den jeweiligen Modalitäten unterteilt. Links innerhalb jeden Blocks steht das auditive (a) Ergebnis, rechts daneben das visuelle (v) Ergebnis und danach das audio-visuelle (av) Ergebnis.

Die Signifikanztests wurden jeweils mit dem Friedman-Test durchgeführt. Durch diese zwei-faktorielle Varianzanalyse konnten die Einflüsse der Geschwindigkeit und der Modalität auf das Ergebnis untersucht werden.

Noise-Bedingung

In Abbildung 5.9 ist die Trefferquote in Prozent für die Noise-Bedingung zu sehen. Für jede der fünf Geschwindigkeiten war eine signifikante Erhöhung (a-v: $p=0.023$; a-av: $p<0.001$; v-av: $p<0.001$) der Trefferquote im bimodalen (av) Fall, gegenüber den unimodalen Fällen (a, v) zu beobachten. Für langsamere Objekte

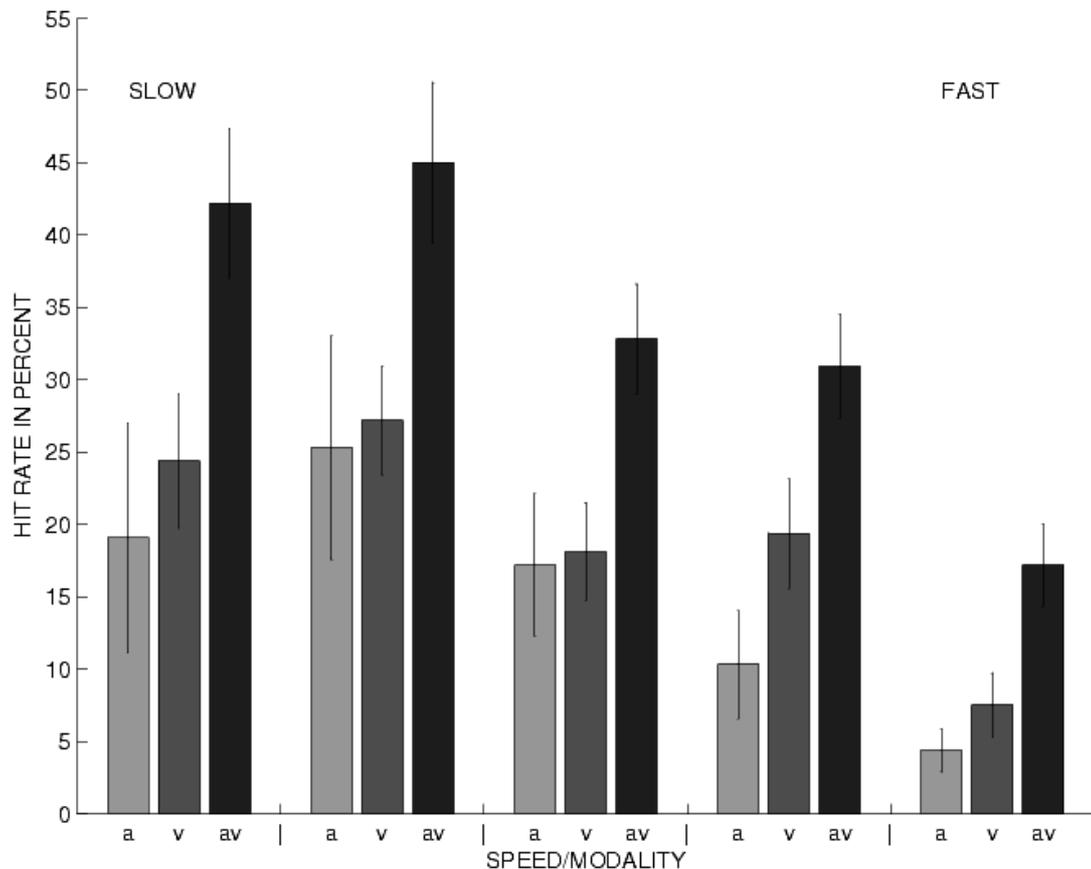


Abbildung 5.9: Trefferquote in der Noise-Bedingung in Prozent, gemittelt über alle 8 Beobachter. Es zeigte sich eine signifikante Zunahme der Trefferquote im bimodalen (av) Fall, gegenüber den unimodalen (a, v) Fällen.

zeigte sich eine Steigerung der Trefferquote für alle Modalitäten, gegenüber schnelleren Objekten ($p < 0.001$).

Abbildung 5.10 zeigt die Fehler für die geschätzte Ankunftszeit für alle 8 Beobachter und alle Geschwindigkeiten. Ein Fehler von 0 bedeutet, dass die Ankunftszeit des Objekts korrekt geschätzt wurde. Ein positiver Fehler bedeutet, dass die Ankunftszeit zu spät geschätzt wurde. Ein negativer Fehler bedeutet eine zu frühe Schätzung. Jeder Balken repräsentiert den Mittelwert des Fehlers über alle Beobachter für eine einzelne Kondition. Die Mittelwerte der drei Modalitäten (a, v, av) unterschieden sich nicht signifikant ($p = 0.10$) innerhalb einer Geschwindigkeit. Entlang der Geschwindigkeiten ergab sich eine Verschiebung ($p = 0.001$) der geschätzten Zeiten. Je langsamer der Stimulus, desto eher tendierten die Beobachter dazu, den Zeitpunkt zu früh zu schätzen. Der Bias in der bimodalen Kondition (av) war wesentlich geringer als in den unimodalen Konditionen.

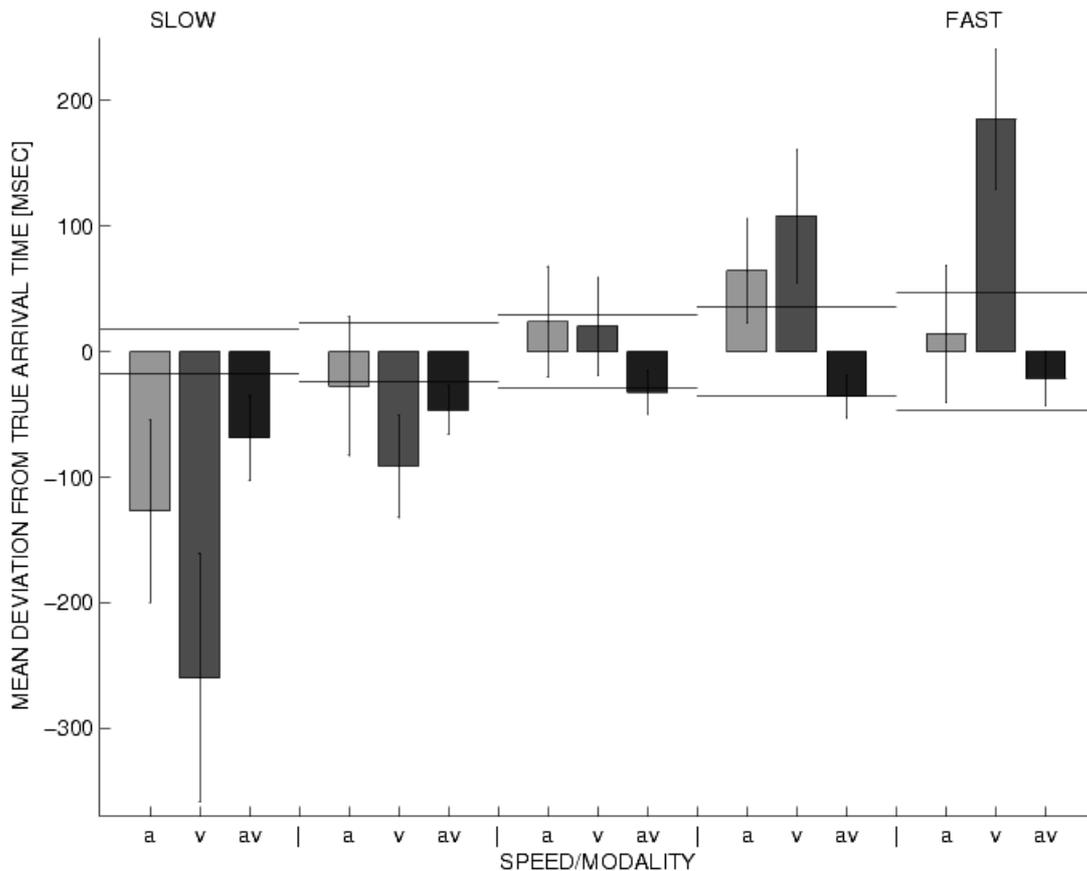


Abbildung 5.10: Mittlerer Fehler für alle acht Beobachter in der Noise-Bedingung. Die Unterschiede der drei Modalitäten innerhalb einer Geschwindigkeit waren nicht signifikant.

In Abbildung 5.11 werden die Standardabweichungen der geschätzten Ankerzeiten gezeigt. Die drei Balken in jedem Geschwindigkeitsfeld zeigen die gemittelten Standardabweichungen für die beobachteten Daten in den jeweiligen Modalitäten (a, v, av). Der Einfluss der Geschwindigkeit des Stimulus auf die Variabilität war nicht signifikant ($p=0.37$). Die Variabilität war, ebenso wie die Trefferquote, für den bimodalen Fall, stets signifikant kleiner (a-v: $p<0.001$; a-av: $p<0.001$; v-av: $p<0.001$) als für die unimodalen Konditionen.

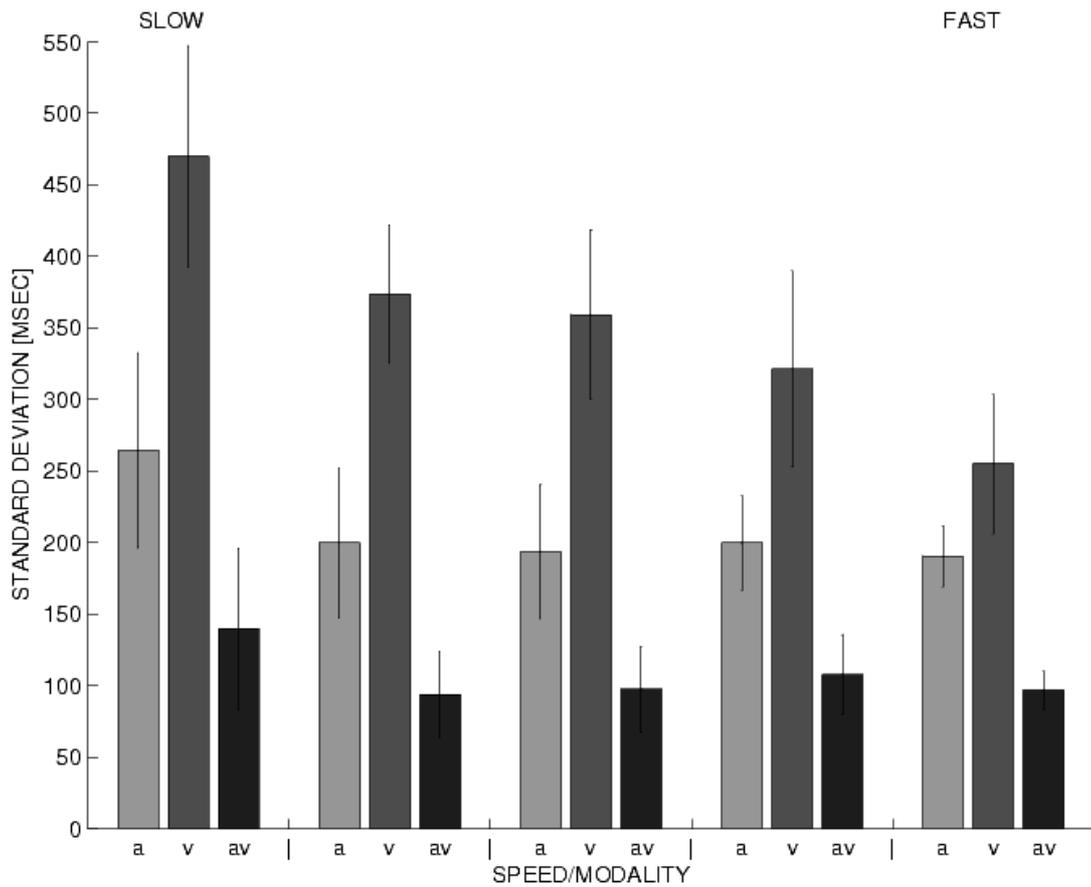


Abbildung 5.11: Mittlere Standardabweichungen für alle 8 Beobachter in der Noise-Bedingung. Die Variabilität war innerhalb einer Geschwindigkeit für den bimodalen Fall stets kleiner als für die unimodalen Fälle.

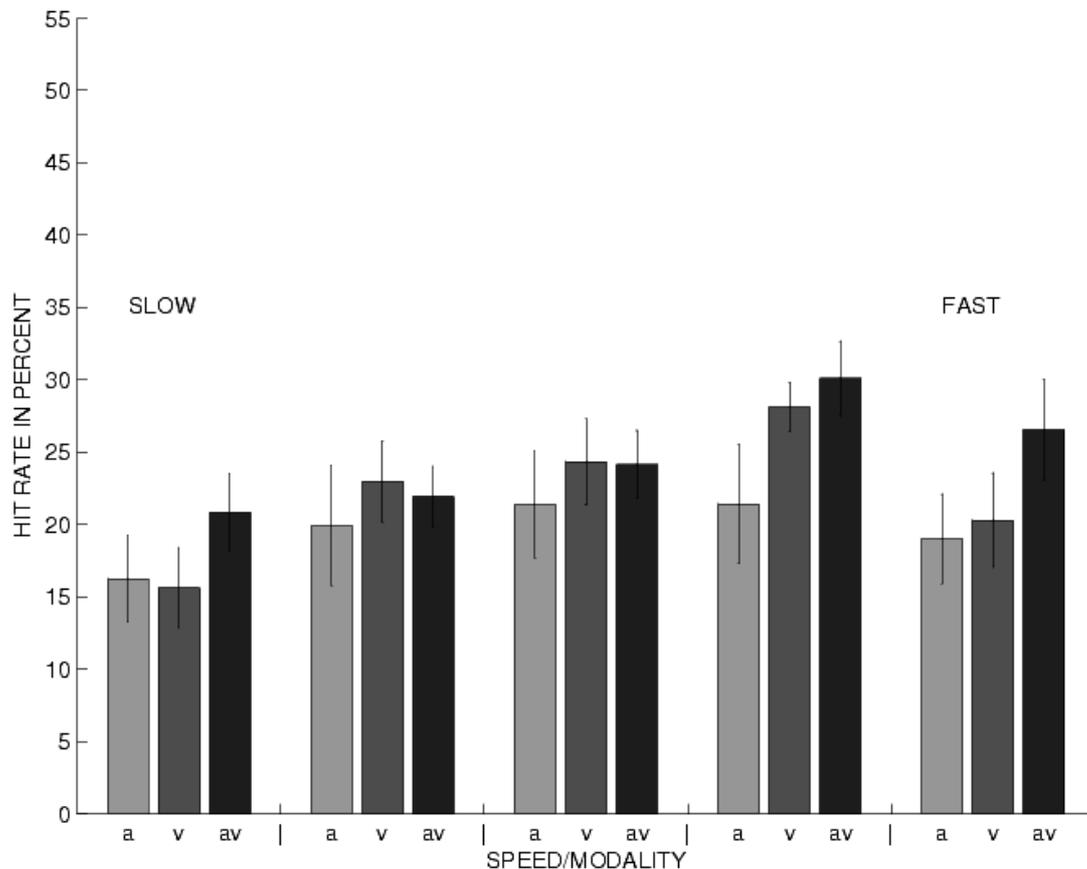


Abbildung 5.12: Mittlere Trefferquote für die Tunnel-Bedingung. Die bimodale Trefferquote war innerhalb einer Geschwindigkeit stets höher als die auditive Trefferquote.

Tunnel-Bedingung

In Abbildung 5.12 ist die durchschnittliche Trefferquote aller 10 Beobachter für die Tunnel-Bedingung zu sehen. Die Trefferquote war höher ($p=0.02$) für den bimodalen Fall (av), verglichen mit dem auditiven Fall (a). Zum visuellen Fall (v) konnte sich die bimodale Trefferquote nicht signifikant verbessern ($p=0.08$).

Abbildung 5.13 zeigt die Fehler für die geschätzte Ankunftszeit für die 10 Beobachter und alle Geschwindigkeiten. Die Mittelwerte für die verschiedenen Modalitäten (a, v, av) unterschieden sich nicht signifikant voneinander ($p=0.97$). Zwischen den verschiedenen Geschwindigkeitsstufen war ein signifikanter Bias festzustellen ($p<0.001$). Die Beobachter schätzen die Ankunftszeit früher, wenn sich das Objekt langsamer bewegt.

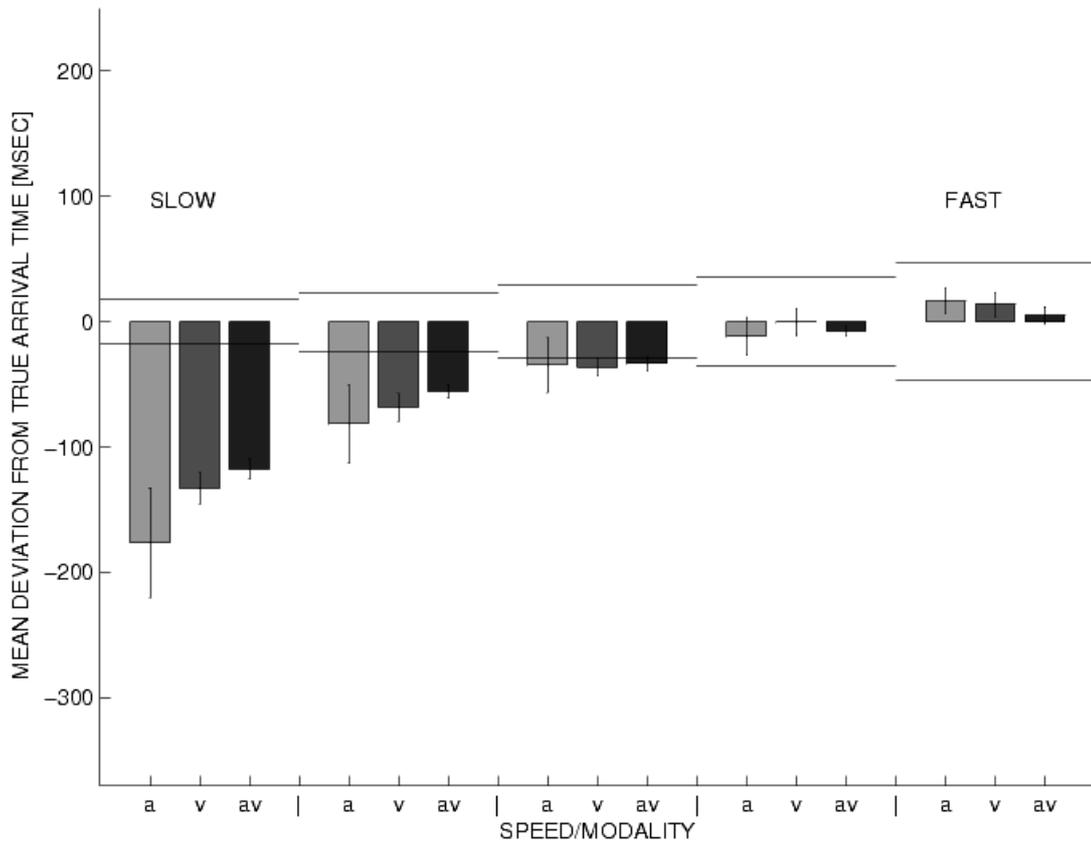


Abbildung 5.13: Mittlerer Fehler der geschätzten Ankunftszeit der Objekte für alle Beobachter in der Tunnel-Bedingung. Innerhalb einer Geschwindigkeit waren die Unterschiede in den drei Modalitäten nicht signifikant.

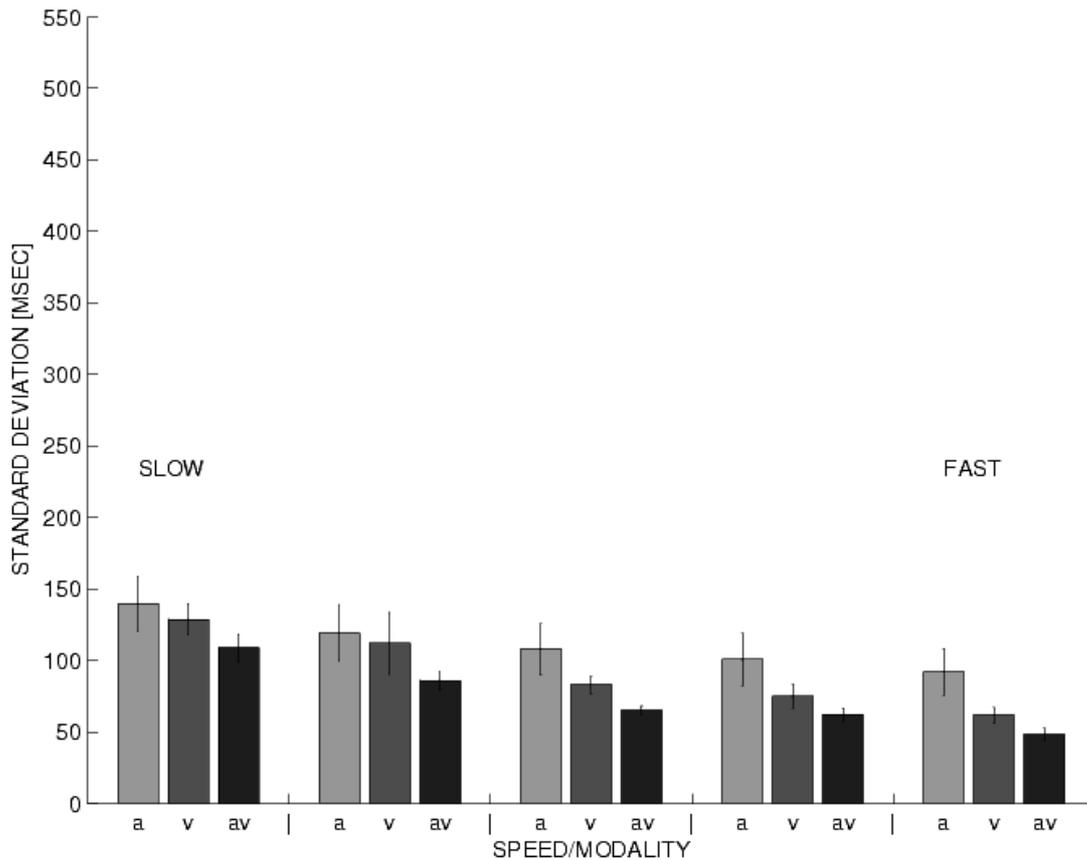


Abbildung 5.14: Mittlere Standardabweichung in der Tunnel-Bedingung. Die Variabilität war innerhalb einer Geschwindigkeit für den bimodalen Fall stets kleiner als für die unimodalen Fälle.

In Abbildung 5.14 sind die Standardabweichungen der geschätzten Ankunftszeiten zu sehen. Die Variabilität stieg, ebenso wie in der Noise-Bedingung, mit abnehmender Geschwindigkeit des Stimulus ($p < 0.001$). Innerhalb jedes Blocks war die Variabilität für den bimodalen Fall (av) stets signifikant kleiner ($p < 0.001$) als für die unimodalen Konditionen (a, v).

5.2.3 Modellierung

Die Modellierung der bimodalen Ergebnisse erfolgte mit zwei unterschiedlichen Modellen. Es wurden das Maximum-Likelihood-Modell (MLE-Modell) aus Kapitel 2.3.1 und das Minimum-Modell (MIN-Modell) aus Kapitel 2.3.5 verwendet. Mit jedem Modell sollten die bimodalen Ergebnisse jedes Experiments aus den jeweils unimodalen empirischen Daten vorhergesagt werden. Mit dem MLE-Modell

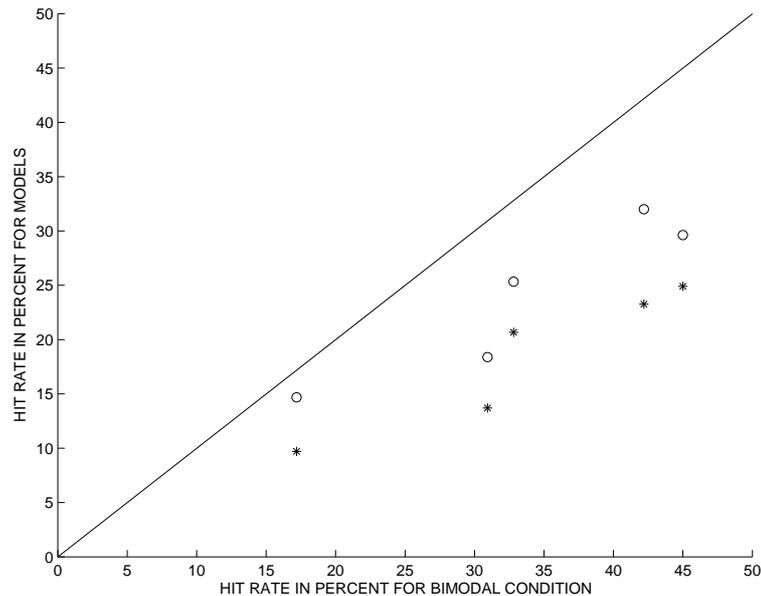


Abbildung 5.15: Modellvergleich für die Trefferquote in der Noise-Bedingung. Das MIN-Modell (o) zeigt für jede der fünf Geschwindigkeiten eine bessere Anpassung als das MLE-Modell (*).

wurde in [EB02] bereits erfolgreich die Interaktion zwischen dem visuellen und haptischen System modelliert. Nun soll überprüft werden, ob sich mit diesem Modell auch die bimodalen Daten aus den hier durchgeführten Experimenten vorhersagen lassen.

Die Abbildungen 5.15-5.20 zeigen jeweils die Modellvorhersagen in Korrelation zu den empirischen Ergebnissen. Auf den Abszissen befinden sich jeweils die bimodalen Daten. Auf der Ordinate sind die Modellergebnisse abgebildet. Die Modellvorhersagen für das MLE-Modell sind mit „*“ und die Vorhersagen für das MIN-Modell mit „o“ gekennzeichnet. Für jede der fünf Geschwindigkeitsstufen ergibt sich pro Modell ein Datenpunkt. Je näher sich die Punkte an der gezeichneten Winkelhalbierenden befinden, desto besser ist ihre Übereinstimmung mit den bimodalen Versuchsergebnissen. Die Signifikanztests zur Überprüfung der Güte der Modelle wurden mit dem Wilcoxon-Rangsummentest durchgeführt.

Noise-Bedingung

In Abbildung 5.15 werden die Vorhersagen zwischen den beiden Modellen und den Versuchsdaten für die Trefferquote in der Noise-Bedingung gezeigt. Das MIN-Modell (o) zeigt eine signifikant ($p=0.05$) bessere Anpassung an die bimodalen Trefferquoten als das MLE-Modell (*).

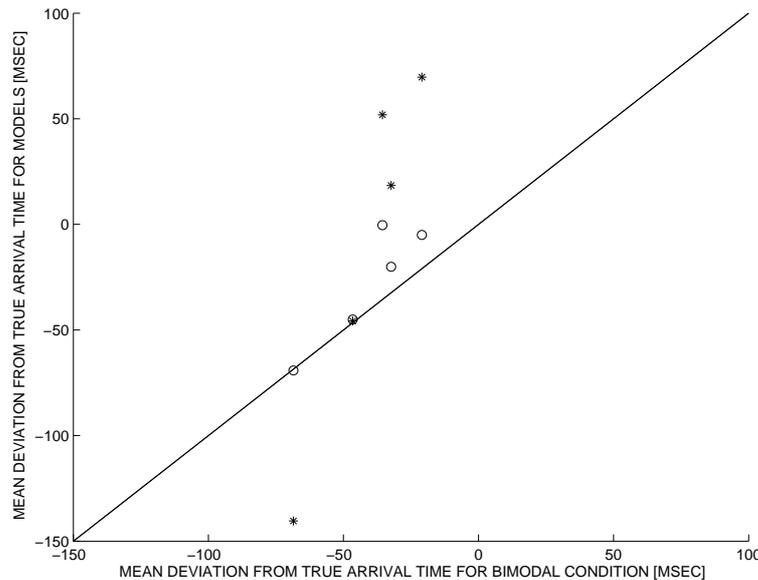


Abbildung 5.16: Modellvorhersagen für den Fehler der geschätzten Ankunftszeit in der Noise-Bedingung. Die Vorhersage des MIN-Modells (o) ist signifikant besser als die des MLE-Modells (*).

Abbildung 5.16 zeigt die Modellvorhersagen für den Fehler der geschätzten Ankunftszeit in der Noise-Bedingung. Falls die sensorische Integration durch einen Maximum-Likelihood-Integrator beschrieben werden kann, ergibt sich für die Vorhersage des Mittelwerts im bimodalen Fall das gewichtete arithmetische Mittel aus den unimodalen Fällen. Durch den starken Bias der unimodalen Konditionen (siehe Abbildung 5.10) kann das MLE-Modell (*) die beobachteten Daten nicht vorhersagen. Das MIN-Modell (o) ist gegenüber diesem Bias wesentlich unempfindlicher und liefert eine deutlich bessere Vorhersage ($p < 0.01$).

Die Modellvorhersagen für die bimodale Standardabweichung sind in Abbildung 5.17 zu sehen. Das MLE-Modell (*) liefert für alle Geschwindigkeitsstufen zu große Ergebnisse. Das MIN-Modell (o) hingegen liefert durchgängig signifikant ($p = 0.05$) bessere Schätzungen der Standardabweichung als das MLE-Modell (*).

Tunnel-Bedingung

Die Modellvorhersagen für die bimodale Trefferquote in der Tunnel-Bedingung sind in Abbildung 5.18 zu sehen. Das MIN-Modell (o) und das MLE-Modell (*) zeigen keine signifikanten ($p = 1$) Unterschiede in der Güte der Vorhersage.

In Abbildung 5.19 sind die Modellvorhersagen für den Fehler der geschätzten Ankunftszeit in der Tunnel-Bedingung zu sehen. Die Anpassung der beiden Modelle

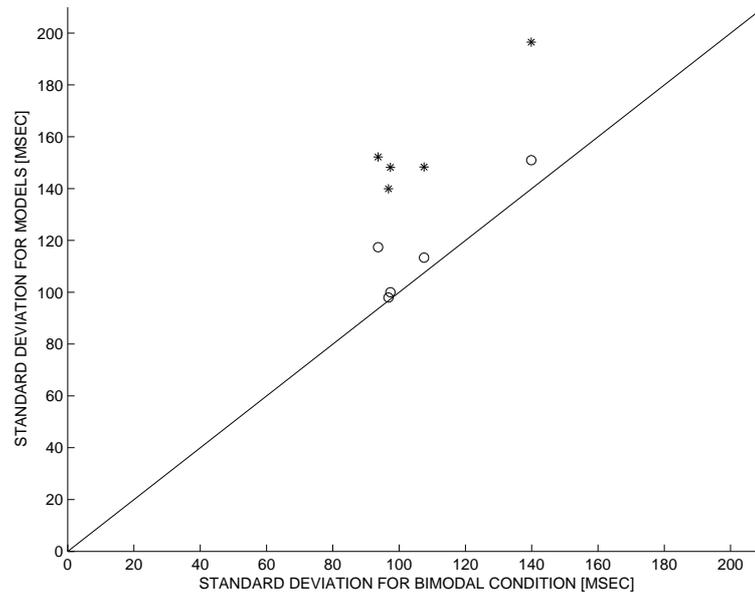


Abbildung 5.17: Modellvorhersagen für die Standardabweichungen in der Noise-Bedingung. Die Standardabweichungen werden vom MLE-Modell (*) wesentlich mehr überschätzt als vom MIN-Modell (o). Der Fehler des MIN-Modells (o) ist signifikant kleiner.

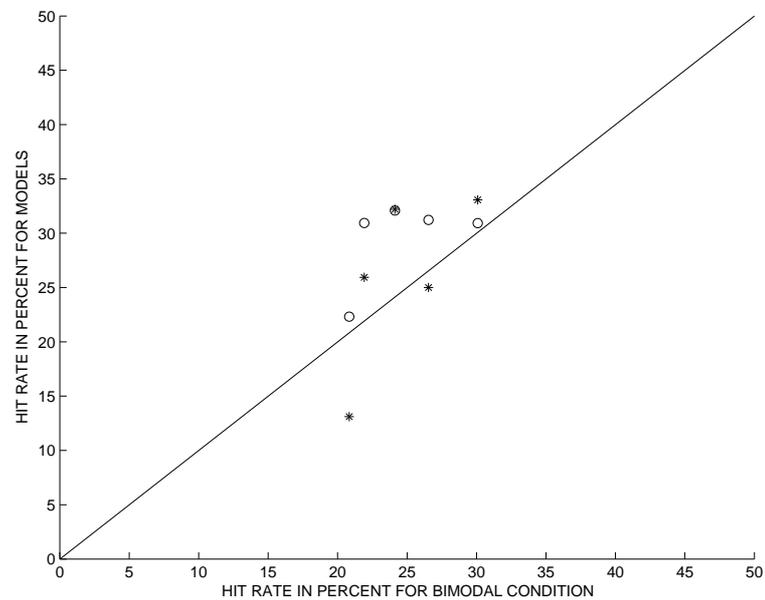


Abbildung 5.18: Modellvorhersagen für die Trefferquote in der Tunnel-Bedingung. Das MLE-Modell (*) hat keine Vorteile gegenüber dem MIN-Modell (o).

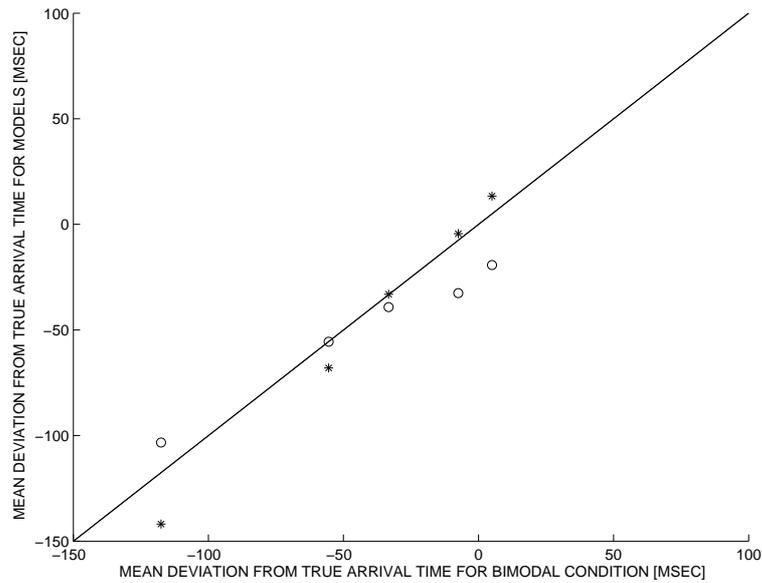


Abbildung 5.19: Modellvergleich für die Fehler der geschätzten Ankunftszeit in der Tunnel-Bedingung. Weder das MLE-Modell (*) noch das MIN-Modell (o) zeigt eine signifikant bessere Anpassung an die empirischen Daten.

(*) und (o) an die Daten zeigt keine signifikanten ($p=1$) Unterschiede.

Abbildung 5.20 zeigt die Modellvorhersagen für die Standardabweichungen in der Tunnel-Bedingung. Das MIN-Modell (o) und das MLE-Modell (*) zeigen ebenfalls keine signifikanten ($p=0.84$) Unterschiede.

5.2.4 Folgerungen

Die Ergebnisse werden zunächst innerhalb einer Bedingung betrachtet. Anschließend werden die zwei Bedingungen gegenübergestellt.

Noise-Bedingung

In Abbildung 5.9 ist bereits ein starker Interaktionseffekt in der Trefferquote zu beobachten.

Abbildung 5.10 zeigt, dass für die Mittelwerte der geschätzten Ankunftszeiten im bimodalen Fall der Bias wesentlich geringer war, als in den unimodalen Fällen. Für die audio-visuellen Ergebnisse war auch eine gewisse „Vorhaltezeit“ zu beobachten. Die Probanden drückten den Knopf bei jeder Geschwindigkeitsstufe zu früh. Diese Vorhaltezeit steigt bei langsameren Geschwindigkeiten. Dies könnte darauf hindeuten, dass der Proband die interne Annahme machte, dass der

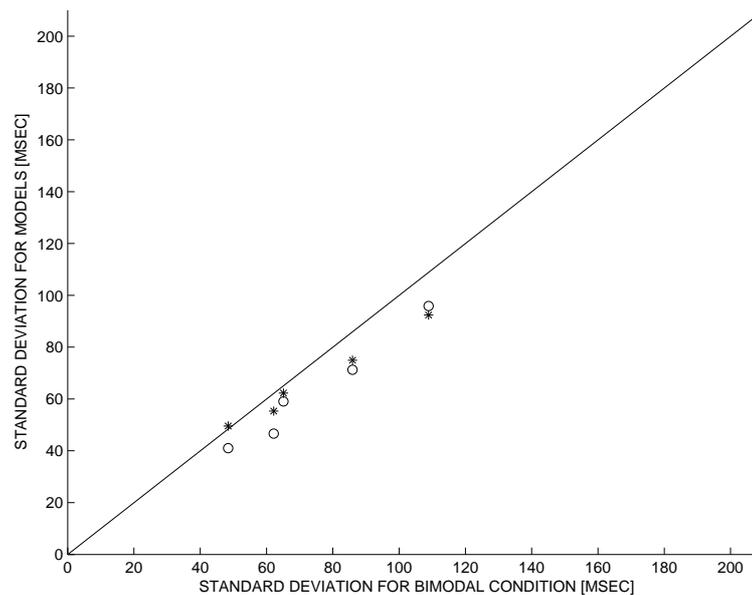


Abbildung 5.20: Modellvorhersagen für die Standardabweichungen in der Tunnel-Bedingung. Das MIN-Modell (o) zeigt eine sehr ähnliche Anpassung wie das MLE-Modell (*).

Knopfdruck eine gewisse Zeit benötigt, bis er an der Zielposition eine Wirkung zeigt.

Die bimodalen Standardabweichungen (siehe Abbildung 5.11) zeigten ebenfalls einen starken Interaktionseffekt.

Besonders bei den langsamen Geschwindigkeiten wird deutlich, dass zur sinnvollen Beurteilung der Leistung der Beobachter alle drei Faktoren (Trefferquote, Mittelwert der geschätzten Ankunftszeit und die zugehörige Standardabweichung) betrachtet werden müssen. Für den auditiven Stimulus waren bei den beiden langsamsten Geschwindigkeitsstufen sowohl beim Mittelwert (Abbildung 5.10) als auch bei der Standardabweichung (Abbildung 5.11) geringere Werte als für den visuellen Stimulus zu beobachten. Trotzdem wurde jeweils für den visuellen Stimulus eine höhere Trefferquote (Abbildung 5.9) registriert.

Die Modellvorhersagen (Abbildungen 5.15-5.17) gelingen für alle drei Faktoren (Trefferquote, Mittlerer Fehler und Standardabweichungen der geschätzten Ankunftszeit) mit dem MIN-Modell deutlich besser als mit dem MLE-Modell. Durch die großen unimodalen Standardabweichungen wird auch die mit dem MLE-Modell berechnete bimodale Standardabweichung überschätzt. Die beim MLE-Modell verwendete gewichtete Mittelwertschätzung aus den unimodalen Daten liefert durch den starken Bias der unimodalen Daten ebenfalls eine unzureichende Vorhersage. Das MIN-Modell wird durch diesen Bias wesentlich weniger be-

einflusst und erreicht dadurch eine signifikant bessere Schätzung der empirischen bimodalen Ergebnisse.

Tunnel-Bedingung

Für die bimodale Trefferquote bei der Tunnel-Bedingung (Abbildung 5.12) war gegenüber den unimodalen Bedingungen nur eine moderate Steigerung zu messen.

Bei den Mittelwerten der geschätzten Ankunftszeiten (Abbildung 5.13) war ein geschwindigkeitsabhängiger Bias für alle Modalitäten zu beobachten. Der Zeitpunkt des Austritts des Objekts aus dem Tunnel ist anscheinend schwer einzuschätzen. Sowohl die Trefferquote als auch die Mittelwerte deuten darauf hin, dass eine optimale Stimulusgeschwindigkeit existiert, bei der eine maximale Leistung erreicht wird (zweitschnellste Geschwindigkeitsstufe, 34 deg/s). Abbildung 5.21 zeigt die Mittelwerte der Antwortzeiten aller Probanden in allen Modalitäten (*auditiv* $\equiv o$, *visuell* $\equiv x$ und *bimodal* $\equiv +$) als Funktion der Geschwindigkeit. Die Abbildung zeigt auch drei mögliche Modelle für die Schätzung der Zeit, welche das Objekt im Tunnel verbringt. Würde die Zeit korrekt geschätzt werden, ergäbe sich die horizontale Nulllinie (MODEL PERFECT). Falls für die Tunnelzeit eine, von der tatsächlichen Geschwindigkeit unabhängige, konstante Zeit angenommen würde, ergäbe sich die gestrichelte Linie (MODEL CONSTANT TUNNELTIME). Im Beispiel wurde die sich aus dem Mittel der fünf verschiedenen Geschwindigkeiten ergebende Zeit verwendet. Für das dritte Modell (MODEL MIX) wird angenommen, dass die Probanden die Geschwindigkeit des Objekts nach dem Eintritt in das Tunnel noch für 20 cm korrekt schätzen und dann eine von der Geschwindigkeit unabhängige, konstante Zeit von 110 ms hinzuaddieren. Die gestrichelte Kurve zeigt das Resultat für dieses Modell. Die Kurve stimmt gut mit den visuell beobachteten Daten überein. Möglicherweise gelingt die Schätzung der Geschwindigkeit nach dem Tunneleintritt des Objekts nur noch für einen gewissen Zeitraum. Diese Hypothese kann in weiteren Experimenten mit z.B. unterschiedlichen Tunnellängen überprüft werden.

Die Auswertung der Standardabweichung (Abbildung 5.14) ergab für alle Geschwindigkeiten kleinere Werte für die bimodalen Ergebnisse, verglichen mit den unimodalen Leistungen. Hierbei kann wieder die Interaktion zwischen dem auditiven und visuellen System bestätigt werden.

Nach der Analyse der Modellvorhersagen (siehe Abbildungen 5.18-5.20) kann keinem der beiden Modelle der Vorzug gegeben werden. Die Fehler der beiden Modelle zeigen keine signifikanten Unterschiede.

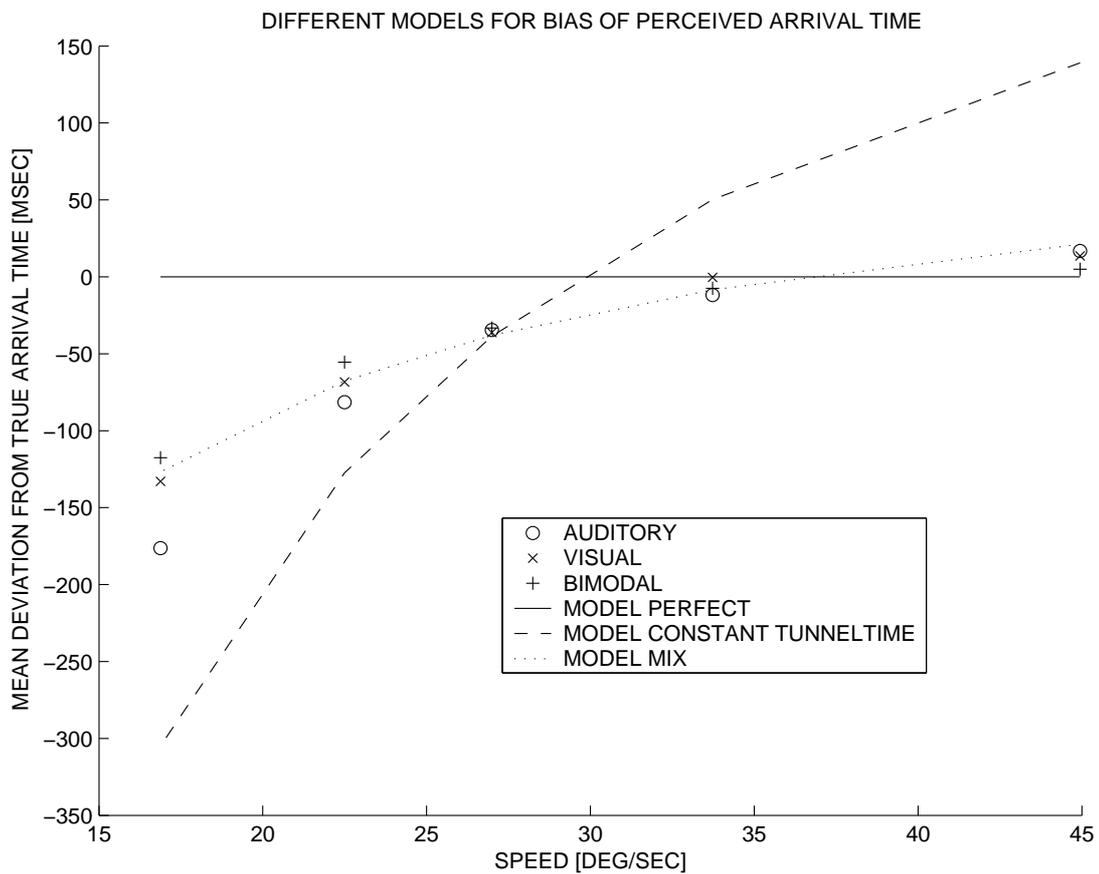


Abbildung 5.21: Die Mittelwerte der Antwortzeiten (*auditiv* \equiv *o*, *visuell* \equiv *x* und *bimodal* \equiv *+*) aller Probanden sind als Funktion der Geschwindigkeit aufgetragen. MODEL PERFECT nimmt eine korrekte Schätzung der Geschwindigkeit des Objekts im Tunnel an. MODEL CONSTANT TUNNELTIME nimmt eine konstante Zeit des Objekts im Tunnel an. Für MODEL MIX wird angenommen, dass die korrekte Geschwindigkeit des Objekts im Tunnel nur für eine bestimmte Strecke geschätzt werden kann.

Vergleich zwischen Tunnel- und Noise-Bedingung

Bei einem Vergleich der unimodalen (a und v) Trefferquoten zwischen den beiden Bedingungen (Abbildungen 5.9 und 5.12) zeigte sich für die Tunnelbedingung eine insgesamt höhere Leistung als für die Noise-Bedingung. Dies wurde aber durch den viel stärkeren Interaktionseffekt in der Noise-Bedingung mehr als ausgeglichen, wodurch sich bei einer Betrachtung aller Messwerte (a, v und av) insgesamt für die Noise-Bedingung eine höhere Trefferquote ergab.

Bei der Betrachtung der Mittelwerte der geschätzten Ankunftszeiten (Abbildungen 5.10 und 5.13) war in beiden Experimenten ein ausgeprägter geschwindigkeitsabhängiger Bias zu beobachten. Dieser Bias war bereits in Vorversuchen zu sehen. Daraufhin wurde den Probanden nach jeder Stimuluspräsentation mitgeteilt, ob sie die Ankunftszeit zu früh oder zu spät geschätzt haben. Trotzdem konnte der Bias nicht eliminiert werden. Die Wahrnehmungsprozesse, welche der geschwindigkeitsabhängigen Verschiebung der Antwortzeiten zu Grunde liegen, müssen durch weitere Experimente untersucht werden.

Vergleicht man die Standardabweichungen der beiden Versuchsbedingungen in Abbildung 5.11 und Abbildung 5.14, fallen die durchgängig für alle Geschwindigkeiten und alle Modalitäten wesentlich kleineren Werte für die Tunnel-Bedingung ins Auge. Dies lässt aber nicht uneingeschränkt auf eine bessere Leistung der Probanden schließen, da trotzdem bei der Noise-Bedingung insgesamt eine höhere Trefferquote erzielt wurde. Die bimodalen Standardabweichungen zeigen in beiden Versuchsbedingungen, dass die Informationsintegration dazu benutzt wurde, Ergebnisse mit kleinerer Variabilität zu produzieren. Es wurden also bei der bimodalen Stimulierung konstantere Leistungen erbracht.

Der Vergleich der Modelle (Abbildung 5.15-5.20) zeigt, dass vor allem in der Noise-Bedingung, das MIN-Modell gegenüber dem MLE-Modell große Vorteile hat. Insgesamt sind für beide Modelle die Vorhersagen in der Tunnel-Bedingung besser als in der Noise-Bedingung.

Kapitel 6

Diskussion

In dieser Arbeit wurden verschiedene Aspekte der audio-visuellen Interaktion im menschlichen Gehirn untersucht.

In einer Computersimulation zur Kovariation von Schallintensität und Größe bewegter Objekte konnte gezeigt werden, dass die von einem Sensorsystem empfangenen physikalischen Größen ein systematisches, statistisches Muster ergeben. Daraus ergab sich die Vorhersage für die nachfolgenden Wahrnehmungsexperimente, dass die statistische Redundanz bei gleichsinniger Veränderung von Schallintensität und scheinbarer Objektgröße Einfluss auf die Wahrnehmung audio-visueller Stimuli hat. Diese Vorhersage aus Abschnitt 3.3 konnte durch die Experimente bestätigt werden.

Die durchgeführten Experimente zeigten, dass die Informationen aus unterschiedlichen Sinnesorganen nicht notwendigerweise zu einer starken Interaktion führen (siehe Kapitel 5.1). Vielmehr zeigte sich, dass für einen starken Interaktionseffekt bestimmte Voraussetzungen erfüllt sein müssen. Die Modellierung der Versuchsergebnisse wurde mit unterschiedlichen Ansätzen durchgeführt. In Kapitel 4.1 war mit der Dempster-Shafer Theorie eine gute Beschreibung der experimentellen Daten möglich, während das Probability-Summation-Modell und Linear-Summation-Modell nur eine ungenügende Anpassung lieferten. Mit der Maximum-Likelihood-Methode konnten in Kapitel 5.2 nur bestimmte Teilaspekte der Versuchsergebnisse dargestellt werden. Das Minimum-Modell hingegen liefert eine deutlich bessere Vorhersage der empirischen Daten.

Ein Ziel dieser Arbeit war es, herauszufinden unter welchen Bedingungen eine Interaktion zwischen dem auditiven und dem visuellen System beobachtet werden kann. In Kapitel 4.1 ist für die gleichsinnigen Veränderungen von Schallintensität und Objektgröße ein starker Interaktionseffekt zu beobachten. Die wesentlich niedrigeren Schwellen in diesen Bedingungen lassen darauf schließen, dass das menschliche Wahrnehmungssystem auf diese Stimuluskombinationen optimiert ist. Die größeren Schwellen bei den räumlich inkonsistenten Bedingungen sind

durchaus sinnvoll. Dadurch kann erreicht werden, dass solche Stimuluskombinationen nicht fälschlicherweise ein und demselben Objekt zugeschrieben werden. Das gleiche Muster bei den Schwellen zeigt sich auch im Nachfolgeexperiment in Kapitel 4.2. Durch den Austausch der Stimuli konnte keine Verbesserung der Schwellen erreicht werden. Die höheren Inkrementalschwellen können durch die komplexeren Stimuli zustande kommen. Der Einfluss der Komplexität der Stimuli auf die Wahrnehmungsleistung sollte daher in weiteren Experimenten genauer untersucht werden.

Die Ergebnisse in Kapitel 5.1 zeigen nur einen sehr schwachen Interaktionseffekt und keine Unterschiede in den konsistenten und inkonsistenten Bedingungen. Dies ist ein weiterer Hinweis darauf, dass die Datenintegration aus verschiedenen sensorischen Kanälen nur auf ökologisch sinnvolle Stimuluskombinationen beschränkt ist. Bei diesem Experiment handelt es sich um sehr abstrakte Bewegungsmuster, die weitgehend getrennt verarbeitet werden. Die beiden Signale werden nicht auf einen gemeinsamen Ursprung zurückgeführt. Daher findet auch keine Interaktion der Signale statt.

Im Gegensatz dazu liefern die Ergebnisse aus dem Experiment in Kapitel 5.2 einen sehr deutlichen Interaktionseffekt. Vor allem in der Noise-Bedingung ergibt sich bei der bimodalen Stimulierung eine wesentlich höhere Wahrnehmungsleistung, als für die unimodalen Stimulierungen. Der Interaktionseffekt scheint gerade bei verrauschten Signalen sehr stark zu sein. Dies ist auch im Einklang mit der „inverse effectiveness rule“ aus Abschnitt 1.3. Entsprechend ist in der Tunnelbedingung der Interaktionseffekt geringer, da hier bereits unimodal sehr deutliche Signale vorhanden sind.

Sowohl bei den Experimenten mit der Stimulusbewegung in Blickrichtung als auch bei den Experimenten mit Stimulusbewegung in frontparalleler Ebene sind unter ökologisch relevanten Bedingungen starke Interaktionen zwischen dem auditiven und dem visuellen Signal zu beobachten.

Ein weiterer Aspekt dieser Arbeit war es, zu beschreiben, wo die gefundene Interaktion im menschlichen Gehirn stattfindet. Die in Abschnitt 1.3 durch neurophysiologische, psychophysische, klinische oder bildgebende sowie auch elektrophysiologische Methoden dargestellten Erkenntnisse bilden dabei die Grundlage für die in dieser Arbeit durchgeführten Wahrnehmungsexperimente. Unser Gehirn ist in der Regel bestrebt, die Signale aus unterschiedlichen Modalitäten einem einzigen Objekt zuzuordnen. Dabei kommen erstaunliche Kompromisse zustande. Das bekannteste Phänomen ist wohl der Bauchrednereffekt. Etliche Arbeiten (siehe [BA03] für eine Übersicht) untersuchen dabei, welche Wahrnehmung ein auditiver und ein dazu räumlich divergierender visueller Reiz erzeugen. Die Position des wahrgenommenen Signals befindet sich dabei in der Regel zwischen den tatsächlichen Positionen der beiden Stimuli. Abhängig von der Versuchsbedingung dominiert dabei entweder die Position des visuellen oder akustischen Stimulus.

Ein weiteres klassisches Phänomen wird als McGurk-Effekt bezeichnet [MM76]. Hier wurden die Einflüsse der Lippenbewegung auf die Sprachwahrnehmung untersucht. Die Wahrnehmung einer gesprochenen Silbe wird dabei durch die visuelle Präsentation einer anderen Silbe beeinflusst. Wenn z.B. akustisch die Silbe „ba-ba“ und gleichzeitig visuell eine Lippenbewegung mit der Silbe „ga-ga“ präsentiert wird, wird dies vom Beobachter als „da-da“ wahrgenommen.

Solche Ergebnisse aus psychophysischen Experimenten lassen den Schluss zu, dass die Verarbeitung eines Reizes in den verschiedenen Modalitäten zuerst getrennt erfolgt und erst später auf kognitiver Ebene eine Integration der wahrgenommenen Stimuli stattfindet.

Neurophysiologische Lokalisationsexperimente [SM93] zeigen aber bereits einen starken Interaktionseffekt im Mittelhirn. Dieser Effekt bei den multimodalen Neuronen im Superior Colliculus ist dann besonders stark, wenn die modalitätsspezifischen Reize eine große räumliche und zeitliche Nähe aufweisen. Dies lässt sich damit erklären, dass die multimodalen Neurone rezeptive Felder besitzen, die bei zeitlicher bzw. räumlicher Konsistenz der Stimuli eine hohe Aktivierung der Neurone erzeugen.

Experimente zur Bewegungsdetektion audio-visueller Stimuli ([RHM⁺03], [WRM⁺03]), die an der in Kapitel 5.2 verwendeten Apparatur durchgeführt wurden, machen ebenfalls deutlich, dass sowohl die räumliche als auch zeitliche Übereinstimmung der Stimuli in den beiden Modalitäten bei der Bewegungserkennung eine wichtige Rolle spielt. Bei diesen Schwellenmessungen befanden sich der auditive und der visuelle Stimulus entweder in der gleichen Hemisphäre oder in verschiedenen Hemisphären. Außerdem konnten sich die Stimuli gleich- oder gegensinnig bewegen. Bei konsistenten Stimuli (gleiche Hemisphäre, gleiche Bewegungsrichtung) sind die gemessenen Schwellen deutlich niedriger als in den anderen Fällen.

Die Ergebnisse aus Lokalisationsexperimenten finden ihre Entsprechung also auch bei Versuchen zur Bewegungsdetektion. Dies deutet darauf hin, dass auch die in dieser Arbeit gefundenen Interaktionseffekte nicht nur das Ergebnis einer kognitiven Verarbeitung sind. Vielmehr deuten die gemessenen Interaktionseffekte auf eine frühe Verarbeitung der Signale hin und lassen sich somit subkortikalen Prozessen zuordnen.

Desweiteren sollte in dieser Arbeit untersucht werden, mit welchen Modellen die gefundenen Interaktionen aus den unimodalen Daten vorhergesagt werden können. Die in dieser Arbeit verwendeten Methoden zur Modellierung der Interaktionseffekte zeigen teilweise eine gute Annäherung an die experimentellen Ergebnisse. Für den Einsatz dieser Methoden in technischen Sensorsystemen ergeben sich aber noch erhebliche Einschränkungen (siehe dazu Kapitel 1.2). Oft erfolgt die Realisierung der Datenfusion durch eine Kombination verschiedener Modelle. Neuronale Netze werden z.B. mit der Dempster-Shafer-Methode trainiert [GZ96]. Meist wird aber bei der Modellierung dem dynamischen Aspekt der

Umgebungsbedingungen zu wenig Beachtung geschenkt. Zur effektiven Anwendung, z.B. der Dempster-Shafer-Methode, zur Datenfusion in technischen Sensorsystemen ist eine explizite Modellierung des zeitlichen Verlaufs der Stimuli wünschenswert. Dies zieht aber eine erhöhte Komplexität des Fusionsmechanismus nach sich.

Zur Überprüfung der verwendeten Modelle ist eine fundierte mathematische Grundlage dieser Modelle notwendig. Desweiteren ist es notwendig, dass die vorgeschlagenen Algorithmen nicht zu komplex, in ihrer Funktionsweise nachvollziehbar, robust und schnell sind. Damit wird der Einsatz von günstigen Hardware-Komponenten erleichtert.

Die Maximum-Likelihood-Methode, das Probability-Summation-Modell sowie das Linear-Summation-Modell erfüllen alle die eben genannten Voraussetzungen. Das in Abschnitt 4.1 verwendete Dempster-Shafer-Modell verlangt zwar einen potentiell hohen Rechenaufwand, bietet aber für diesen Fall eine intuitive Hypothesenauswahl an. Die Berechnung der Evidenzwerte für die einzelnen Hypothesen entspricht physikalischen Grundsätzen. Die Anpassung an die experimentellen Daten ist deutlich besser als mit dem hybriden Modell. Die Dempster-Shafer-Methode ist daher zur Modellierung dieser audio-visuellen Integrationseffekte gut geeignet.

Das Minimum-Modell liefert ebenfalls eine gute Anpassung an die experimentellen Daten aus den Experimenten in Abschnitt 5.2. Bei diesem Modell wird immer derjenige Wert ausgewählt, der näher am tatsächlichen korrekten Wert liegt. Dieser Wert müsste vom menschlichen System zuerst geschätzt werden. Durch die in den Experimenten gegebene Rückmeldung für eine zu frühe, zu späte oder korrekte Antwort ist eine solche Vorhersage des Zeitpunkts durchaus denkbar. Aus den vom auditorischen und visuellen Kanal aktuell empfangenen Daten könnte dann derjenige Wert ausgewählt werden, welcher der Vorhersage am Nächsten ist.

Die in dieser Arbeit verwendeten Modelle mit einer teilweisen guten Anpassung an die experimentellen Ergebnisse haben gezeigt, dass die Komplexität der menschlichen Wahrnehmung und multisensorischen Informationsverarbeitung noch nicht vollständig beschrieben werden kann. Insbesondere ist der Einfluss unterschiedlicher Verarbeitungsebenen des Gehirns auf die Wahrnehmung noch nicht geklärt. Neben dem zeitlichen Verlauf der Wahrnehmung unserer Umgebung sollte daher durch weitere Experimente untersucht werden, auf welche Art und Weise die unterschiedlichen Top-Down und Bottom-Up Verarbeitungsprinzipien die menschliche Informationsverarbeitung beeinflussen.

Literaturverzeichnis

- [AB01] D. Alais and D. C. Burr. No low-level interaction between visual and auditory motion. In *ECVP*, 2001.
- [ABH00] J. E. Aspell, D. I. Bramwell, and A. C. Hurlbert. Interactions between visual and auditory movement perception in a direction discrimination task. In *ECVP*, 2000.
- [Alb97] J. S. Albus. The nist real-time control system (rcs): an approach to intelligent systems research. *Journal of Experimental and Theoretical Artificial Intelligence*, 9:175–194, 1997.
- [BA03] P. Bertelson and G. Aschersleben. Temporal ventriloquism: crossmodal interaction on the time dimension. *International Journal of Psychophysiology*, 50(1-2):147–155, 2003.
- [BGM00] E. Binaghi, I. Gallo, and P. Madella. A neural model for fuzzy dempster-shafer classifiers. *International Journal of Approximate Reasoning*, 25(2):89–121, 2000.
- [BHI⁺03] Khalafalla O. Bushara, Takashi Hanakawa, Ilka Immisch, Keiichiro Toma, Kenji Kansaku, and Mark Hallett. Neural correlates of cross-modal binding. *Nature Neuroscience*, 6(2):190–195, 2003.
- [BLM⁺98] E. Binaghi, L. Luzzi, P. Madella, F. Pergalani, and A. Rampini. Slope instability zonation: a comparison between certainty factor and fuzzy dempster-shafer approaches. *Natural Hazards*, 17(1):77–97, 1998.
- [Bro99] M. C. Brown. Audition. In M. J. Zigmond, F. E. Bloom, S. C. Landis, J. L. Roberts, and L. R. Squire, editors, *Fundamental Neuroscience*, chapter 27. Academic Press, 1999.
- [CHIB01] G. A. Calvert, P. C. Hansen, S. D. Iversen, and M. J. Brammer. Detection of audio-visual integration sites in humans by application of electrophysiological criteria to the bold effect. *NeuroImage*, 14(2):427–438, 2001.

- [CSS04] G. A. Calvert, C. Spence, and B. E. Stein, editors. *THE HANDBOOK OF MULTISENSORY PROCESSES*. THE MIT PRESS, Cambridge, Massachusetts, 2004.
- [DCG98] J.-R. Duhamel, C. L. Colby, and M. E. Goldberg. Ventral intraparietal area of the macaque: Congruent visual and somatic response properties. *Journal of Neurophysiology*, 79:126–136, 1998.
- [Den00] T. Denoeux. Modeling vague beliefs using fuzzy-valued belief structures. *Fuzzy Sets and Systems*, 116(2):167–199, 2000.
- [Eas83] R. D. Easton. The effect of head movements on visual and auditory dominance. *Perception*, 12(1):63–70, 1983.
- [EB02] M. O. Ernst and M. S. Banks. Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, 415:429–433, 2002.
- [Ebn92] C. Ebner. *Statistik Für Soziologen, Pädagogen, Psychologen und Mediziner*, volume 1. Verlag Harri Deutsch, Thun und Frankfurt am Main, 1992.
- [FMM⁺00] J. J. Foxe, I. A. Morocz, M. M. Murray, B. A. Higgins, D. C. Javitt, and C. E. Schroeder. Multisensory auditory-somatosensory interactions in early cortical processing revealed by high-density electrical mapping. *Cognitive Brain Research*, 10:77–83, 2000.
- [GA99] K. Goebel and A. Agogino. Fuzzy sensor-fusion for gas turbine power plants. In *Sensor Fusion: Architectures, Algorithms, and Applications III*, volume 3719 of *PROCEEDINGS OF SPIE*, pages 52–61. SPIE - The International Society for Optical Engineering, 1999.
- [Gar95] M. A. Garcia. A hierarchical world model representation supporting heterogeneous multisensory integration. In *Proceedings of International Conference on Advanced Robotics*, pages 461–471, Sant Feliu de Guixols, Spain, September 1995.
- [GG00] M. S. A. Graziano and S. Gandhi. Location of the polysensory zone in the precentral gyrus of anesthetized monkeys. *Experimental Brain Research*, 135:259–266, 2000.
- [GP99] M. H. Giard and F. Peronnet. Auditory-visual integration during multimodal object recognition in humans: A behavioral and electrophysiological study. *Journal of Cognitive Neuroscience*, 11(5):473–490, 1999.
- [Gra89] N. Graham. *Visual Pattern Analyzers*. Oxford University Press, 1989.

- [GZ96] J. H. Graham and J. M. Zurada. A neural network approach for safety and collision avoidance in robotic systems. *Reliability Engineering and System Safety*, 53:327–338, 1996.
- [Hau74] G. Hauske. *Entscheidungstheorie in Biologie und Psychologie*. R. Oldenburg Verlag, München, 1974.
- [HK92] S. A. Hutchinson and A. C. Kak. Multisensor strategies using Dempster-Shafer belief accumulation. In M.A. Abidi and R.C. Gonzalez, editors, *Data Fusion in Robotics and Machine Intelligence*, chapter 4, pages 165–209. Academic Press, 1992.
- [HM99] C. Hiransoog and C. A. Malcolm. Multi-sensor / knowledge fusion. In *Proceeding 1999 IEEE / SICE / RSJ International Conference on Multisensor Fusion and Integration for Intelligent Systems*, pages 117–122. IEEE, 1999.
- [HRSZ01] M. Hofbauer, F. Röhrbein, K. Schill, and C. Zetzsche. Visuell-auditive Kovariation und Interaktion: Die statistische Struktur von Relativbewegungen zwischen Objekt und Beobachter und die Kombination multi-sensorischer Evidenzen. In H. H. Bülhoff, K.R. Gegenfurtner, H.A Mallot, and R. Ulrich, editors, *Beiträge zur 4. Tübinger Wahrnehmungskonferenz*, page 162, Tübingen, Germany, 2001. Knirsch Verlag.
- [Hun92] T. L. Huntsberger. Data fusion: A neural networks implementation. In M.A. Abidi and R.C. Gonzalez, editors, *Data Fusion in Robotics and Machine Intelligence*, chapter 13, pages 507–535. Academic Press, 1992.
- [HWM⁺03] M. Hofbauer, S. M. Wuerger, G. Meyer, F. Röhrbein, K. Schill, and C. Zetzsche. Catching audio-visual mice: the extrapolation of audio-visual speed. In *26th European Conference on Visual Perception*, volume 32, page 96, Paris, 2003.
- [HWM⁺04] M. Hofbauer, S. M. Wuerger, G. F. Meyer, F. Röhrbein, K. Schill, and C. Zetzsche. Catching audiovisual mice: Predicting the arrival time of auditory visual motion signals. *Cognitive, Affective, & Behavioral Neuroscience*, 4(2):241–250, 2004.
- [HWMS03] M. Hofbauer, S. M. Wuerger, G. Meyer, and K. Schill. Audio-visual interaction in a speed estimation task. In *4th International Multisensory Research Forum*, Hamilton, Canada, 2003.
- [KD79] J. J. Koenderink and A. J. Van Doorn. The internal representation of solid shape with respect to vision. *Biological Cybernetics*, 32:211–216, 1979.

- [Kü68] K. Küpfmüller. *Die Systemtheorie der elektrischen Nachrichtenübertragung*. Hirzel Verlag Stuttgart, 1968.
- [MM76] H. McGurk and J. MacDonald. Hearing lips and seeing voices. *Nature*, 264,(5588):746–748, 1976.
- [MNS87] M. A Meredith, J. W. Nemitz, and B. E. Stein. Determinants of multisensory integration in superior colliculus neurons: I. temporal factors. *Journal of Neuroscience*, 10:3215–3229, 1987.
- [MPS⁺99] F. Manes, S. Paradiso, J. A. Springer, G. Lamberty, and R. G. Robinson. Neglect after right insular cortex infarction. *Stroke*, 30:946–948, 1999.
- [MRM⁺02] S. Molholm, W. Ritter, M. M. Murray, D. C. Javitt, C. E. Schroeder, and J. J. Foxe. Multisensory auditory-visual interactions during early sensory processing in humans: a high-density electrical mapping study. *Cognitive Brain Research*, 14:115–128, 2002.
- [MS80] R. M. Mulligan and M. L. Shaw. Multimodal signal detection: Independent decisions vs. integration. *Perception & Psychophysics*, 28(5):471–478, 1980.
- [MS86] M. A Meredith and B. E. Stein. Spatial factors determine the activity of multisensory neurons in cat superior colliculus. *Brain Research*, 365:350–354, 1986.
- [Mur98] R. R. Murphy. Dempster-shafer theory for sensor fusion in autonomous mobile robots. *IEEE Transactions on Robotics and Automation*, 14(2):197–206, 1998.
- [MW01] G. Meyer and S. Wuerger. Cross-modal interaction of auditory and visual motion signals. *NeuroReport*, 12:2557–2600, 2001.
- [NH81] E. A. Newman and P. H. Hartline. Integration of visual and infrared information in bimodal neurons of the rattlesnake tectum. *Science*, 213:789–791, 1981.
- [NSSW97] J. Neumann, S. Schlüter, A. Steiff, and P.-M. Weinspach. Früherkennung sicherheitsrelevanter Betriebszustände in Chemieanlagen mit neuronalen Netzen. In C. Freksa, editor, *Proceedings in Artificial Intelligence*, volume 5, pages 486–492. infix, 1997.
- [PLD03] F. Pavani, E. Ladavas, and J. Driver. Auditory and multisensory aspects of visuospatial neglect. *TRENDS in Cognitive Sciences*, 7(9):407–414, 2003.

- [PRD98] S. Petit-Renaud and T. Deneux. A fuzzy-neuro system for reconstruction of multi-sensor information. In W. Brauer, editor, *Proceedings in Artificial Intelligence*, volume 7, pages 322–329. infix, 1998.
- [PS99] R. Pfeifer and C. Scheier. *Understanding Intelligence*. MIT Press, 1999.
- [PY02] L. C. Populin and T. C. T. Yin. Bimodal interactions in the superior colliculus of the behaving cat. *The Journal of Neuroscience*, 22(7):2826–2834, 2002.
- [RHM⁺03] F. Röhrbein, M. Hofbauer, G. Meyer, S. M. Wuerger, K. Schill, and C. Zetzsche. The role of co-localisation for auditory-visual motion signals. In H. H. Bülhoff, K.R. Gegenfurtner, H.A Mallot, R. Ulrich, and F. A. Wichmann, editors, *Beiträge zur 6. Tübinger Wahrnehmungskonferenz*, page 105, Tübingen, Germany, 2003. Knirsch Verlag.
- [RZ00] F. Röhrbein and C. Zetzsche. Auditory-visual interactions and the covariance structure generated by relative movements in natural environments. In *Seventh international congress on sound and vibration*, pages 2427–2434, 2000.
- [Sac84] L. Sachs. *Angewandte Statistik*. Springer-Verlag, Berlin, 1984.
- [Sha76] Glenn Shafer. *a mathematical theory of evidence*. Princeton University Press, 1976.
- [SM93] B. E. Stein and M. A. Meredith. *The merging of the senses*. MIT Press, Cambridge, MA, 1993.
- [Sme88] P. Smets. Belief functions. In P. Smets, E. H. Mamdani, D. Dubois, and H. Prade, editors, *Non-Standard Logics for Automated Reasoning*. Academic Press, London, 1988.
- [SMW94] B. E. Stein, M. A. Meredith, and M. T. Wallace. Neural mechanisms mediating attention and orientation to multisensory cues. In M. Gazzaniga, editor, *The Cognitive Neurosciences*, pages 683–702. MIT Press, 1994.
- [SSL97] R. Sekuler, A. B. Sekuler, and R. Lau. Sound alters visual motion perception. *Nature*, 385:308, 1997.
- [SSVW01] B. A. Stein, T. R. Stanford, J. W. Vaughan, and M. T. Wallace. Multisensory integration. In R. A. Wilson and F. C. Keil, editors, *The MIT Encyclopedia of the Cognitive Sciences*, pages 574–575. MIT Press, 2001.

- [Str87] N. A. Streitz. Die Rolle der Psychologie. In K.-P. Fährnich, editor, *Software-Ergonomie*, pages 43–53. R. Oldenburg Verlag GmbH, 1987.
- [Tre97] Bernhard Treutwein. Yaap : Yet another adaptive procedure. *Spatial Vision*, 11:129–134, 1997.
- [Umk00] E. Umkehrer. *Formalisierung von unsicherem Wissen*. Akademische Verlagsgesellschaft Aka GmbH, Berlin, 2000. Dissertation an der Technischen Universität München, Fakultät für Informatik, Univ.-Prof. Dr. Dr. h.c. W. Brauer.
- [War79] D. H. Warren. Spatial localization und conflict conditions: is there a single explanation. *Perception*, 8(3):323–337, 1979.
- [WHM⁺02] S. M. Wuerger, M. Hofbauer, G. Meyer, F. Röhrbein, K. Schill, and C. Zetsche. The integration of auditory and visual motion signals. In *25th European Conference on Visual Perception*, volume 31, page 148, Glasgow, 2002.
- [WHM03] S. M. Wuerger, M. Hofbauer, and G. F. Meyer. The integration of auditory and visual motion signals at threshold. *Perception & Psychophysics*, 65(8):1188–1196, 2003.
- [WMS98] M. T. Wallace, M. A. Meredith, and B. E. Stein. Multisensory integration in the superior colliculus of the alert cat. *Journal of Neurophysiology*, 80:1006–1010, 1998.
- [WRM⁺03] S. M. Wuerger, F. Röhrbein, G. Meyer, M. Hofbauer, K. Schill, and C. Zetsche. Auditory and visual motion signals have to be colocalized to be effectively integrated. In *Journal of Vision*, volume 3(9), page 771a, 2003.
- [WW80] R. B. Welch and D. H. Warren. Immeditate perceptual response to intersensory discrepancy. *Psychological Bulletin*, 88(3):638–667, 1980.
- [ZE99] E. Zrenner and U. Eysel. Visuelles System. In P. Deetjen and E.-J. Speckmann, editors, *Physiologie*, chapter 3.3. URBAN & FISCHER, 1999.
- [ZRHS02] C. Zetsche, F. Röhrbein, M. Hofbauer, and K. Schill. Audio-visual sensory interactions and the statistical covariance of the natural environment. In *Revista de Acustica*, volume XXXIII(3-4), Sevilla, 2002.