

# Multimedia & Multimodalität

Interaktion zwischen akustischer  
und visueller Wahrnehmung bei  
der Rezeption audiovisueller  
Medien

**Diplomarbeit von  
Christoph Eggener**

Matrikelnummer: 410223

Fachhochschule Düsseldorf  
Fachbereich Medien  
Studiengang Medientechnik

Prof. Dr. Karin Welkert-Schmitt  
Fachbereich Medien  
FH Düsseldorf

Prof. Dr. Heike Sperling  
Institut für Musik und Medien  
Robert Schumann Hochschule Düsseldorf

Düsseldorf, März 2009

# INHALTSVERZEICHNIS

<b>1</b>	<b>EINLEITUNG</b>	4
1.1	Forschungsinteresse der Diplomarbeit	7
1.2	Ziel der Diplomarbeit	7
1.3	Aufbau und Gliederung	8
<b>2</b>	<b>MULTIMEDIA</b>	11
<b>3</b>	<b>DIE MENSCHLICHE WAHRNEHMUNG</b>	22
3.1	Definitionen	22
3.2	Physikalische und Physische Grundlagen	23
3.2.1	Optik	23
3.2.2	Akustik	30
3.3	Neuronale Grundlagen	34
<b>4</b>	<b>MULTISENSORISCHE PROZESSE</b>	48
4.1	Forschungsgebiete und Untersuchungsmethoden	48
4.1.1	Kreuzmodale Prozesse bei der exogenen Lenkung der Aufmerksamkeit	49
4.1.2	Modulation der visuellen Wahrnehmung durch Klang	51
4.1.3	Modulation der akustischen Wahrnehmung durch das Sehen	56
4.1.4	Multisensorische Integration von Bewegungsinformation	60
4.1.5	Kreuzmodale Interaktionen in schnellen Klassifizierungen	65
4.2	Zusammenfassung Multisensorische Prozesse	76
<b>5</b>	<b>SYNÄSTHESIE</b>	77
5.1	Allgemeine Grundlagen	79
5.2	Forschung	80
5.2.1	Sind Synästhesien perzeptuell oder kognitiv?	81
5.2.3	Top-Down-Verarbeitung bei Synästheten	86
5.2.4	Higher und lower synesthets	88
5.2.5	Synästhesie und Metaphern	89
5.2.6	Künstler, Poeten und Synästhesie	90
5.2.7	Der Gyrus angularis und synästhetische Metaphern	91
5.3	Zusammenfassung	93

---

<b>6</b>	<b>MICHEL CHIONS AUDIO-VISION</b> .....	94
6.1	The audiovisual contract.....	95
6.2	The three listening modes .....	98
6.3	Analyse und Vergleich mit wissenschaftlichen Studien .....	101
<b>7</b>	<b>PRAKTISCHE ARBEIT</b> .....	108
7.1	Theoretische Hintergründe .....	108
7.2	Die Umsetzung.....	110
7.2.1	Die Musik.....	110
7.2.2	Die Visualisierung .....	115
7.3	Zusammenfassung .....	127
<b>8</b>	<b>FAZIT UND AUSBLICK</b> .....	129
<b>9</b>	<b>LITERATURVERZEICHNIS</b> .....	139
<b>10</b>	<b>ABBILDUNGSVERZEICHNIS</b> .....	143
<b>11</b>	<b>DANKSAGUNG</b> .....	145
<b>12</b>	<b>EIDESSTATTLICHE ERKLÄRUNG</b> .....	146

**ANHANG: VIDEO-CLIP „CONGRUENCE“.**

Abrufbar unter [http://www.youtube.com/watch?v=oH\\_-DdezpYg](http://www.youtube.com/watch?v=oH_-DdezpYg)

# 1 EINLEITUNG

Audiovisuelle Medien sind heutzutage fester Bestandteil des Alltags vieler Menschen. Ob Fernsehen, Kino, Internet, Computerspiele, Lernsoftware oder Navigationssysteme; Multimedia ist allgegenwärtig und der Umgang damit zu einer Selbstverständlichkeit geworden. Dementsprechend groß ist ihre kulturelle Bedeutung, denn audiovisuelle Medien sind in fast alle Lebensbereiche vorgedrungen: Sie dienen der Unterhaltung, der Information, der Werbung, der Erleichterung in täglichen Aufgaben und vielem mehr. Besonders seit das Fernsehen zum Leitmedium aufgestiegen ist, dominieren sie die Massenkommunikation und spielen nicht zuletzt deshalb auch eine entscheidende Rolle in gesellschaftlichen und politischen Entwicklungen und Prozessen. Begriffe wie *Mediendemokratie* oder *Informationsgesellschaft* stehen beispielhaft für die Rolle der Medien als gesellschaftliche Instanz. Audiovisuelle Medien sind somit in unserer Kultur fest verankert.

Medien dienen in einer technisierten Gesellschaft der Kommunikation zwischen Menschen. Kommunikation ist geprägt vom Zusammenspiel aus Inhalt und Form, Information und Interpretation. Unabhängig vom Medium ist die Frage „*Was sage ich?*“ immer eng verknüpft mit der Frage „*Wie sage ich es?*“. Beide Axiome sind demnach bei der Analyse von Kommunikation stets gemeinsam zu betrachten. Für die audiovisuelle Kommunikation bedeutet dies, dass Gestaltung und Ästhetik eine zentrale Rolle einnehmen, denn das bewegte Bild und der Ton als Medium sind in ihrem Informationsgehalt besonders abhängig von der Interpretation des Rezipienten. Dies verdeutlicht beispielhaft ein Blick auf die täglichen Fernsehnachrichten: Die Auswahl der Bilder, Lautstärke, Art und Beschaffenheit des Umgebungstons der gezeigten Szene, Wortlaut und Stimmfall des Berichterstatters bzw. Erzählers, Schnitt, Montage und nicht zuletzt Helligkeit, Kontrast und Farbe des Bildes beeinflussen, meist unbewusst, die innere Einstellung des Rezipienten und damit indirekt die vermittelten Informationen. Dies ist ein besonderes Beispiel für die Bedeutung von Gestaltung und Ästhetik im audiovisuellen Kommunikationsprozess, da Nachrichten in erster Linie der Information dienen sollen und somit von ihnen ein hohes Maß an Objektivität erwartet wird, anders als von einem

unterhaltendem Spielfilm. Daran wird deutlich, dass Gestaltung und Ästhetik essentiell für die Vermittlung von Inhalten, Botschaften und Gefühlen ist.

Die Ästhetik der Medien befindet sich stets im Wandel und ist ein Spiegel der Gesellschaft. Auf die ästhetische Entwicklung des Fernsehens, des Kinos und anderer audiovisuelle Medien, hatte das Musikvideo als Kunstform in den letzten 20 Jahren großen Einfluss. Es eröffnete ein weites Experimentierfeld an gestalterischen Freiräumen, was auch daran lag, dass seine Zielgruppe hauptsächlich musikinteressierte Jugendliche und junge Erwachsene waren. Das Ziel von Musikvideos war und ist es, Künstler und Bands zu vermarkten, aus ihnen auch visuell ‚eine Marke‘ zu machen und somit die Identifikation zwischen Fan und Künstler zu verstärken (Keazor/Wübbena, 2007, S.18, S.65). Es entstanden neue Sender und Formate, die sich hauptsächlich der Ausstrahlung dieser Videoclips widmeten: Das Musikfernsehen war geboren. Es galt als jung, bunt, laut, innovativ und war Ausdruck und Teil der Jugend- und Popkultur. Scheinbar alles war erlaubt, die Bild-Schnitte wurden immer schneller, das Spiel mit Farben, Formen, Metaphern und Bildern immer ausgefeilter, Konventionen wurden aufgelöst und die technischen Möglichkeiten ausgelotet. Regisseure wie Michel Gondry, Chris Cunningham, Anton Corbijn u.a. schufen vor allem Mitte der 1990er Jahre eine Vielzahl von Videos, die bis heute Kultstatus haben und in ihrer Expressions- und Innovationskraft einzigartig sind (Keazor/Wübbena, 2007, 14; vgl. auch DVD-Serie „Directors Label“, Palm Pictures, 2003 & 2005). Wie fast alle sub- oder popkulturellen Phänomene entwickelte sich das Musikvideo nach der Jahrtausendwende immer deutlicher aus der Avantgarde heraus, hinein in eine zunehmende Kommerzialisierung und in den Mainstream. Musikfernsehen steht heutzutage kaum mehr für Videokunst als vielmehr für Handyklingelton-Werbung und Reality-Soaps. Ästhetisch und technisch hochwertige Videos sind immer seltener anzutreffen (Keazor/Wübbena, 2007, S.12). Sein Einfluss auf die Fernsehästhetik ist im Hinblick auf Schnitttechnik, den Einsatz von Effekten, das Arbeiten mit Bildern und Metaphern und die Verknüpfung von Musik mit bewegten Bildern bis heute deutlich erkennbar.

Auch wenn es in seiner Form völlig neuartig war, modernisierende Einflüsse und in vielerlei Hinsicht Pionier-Charakter hatte, so ist das Musikvideo dennoch nicht der erste Versuch des Menschen, Visuelles und Akustisches in einem Kunstwerk

zu verbinden: Opern, das antike Theater oder sogar frühzeitliche Riten aus Gesang und Malerei sind Beispiele für frühe, audiovisuelle Kunst und Kultur. Die Verbindung von Musik als „Sprache der Gefühle“ (Altenmüller, 2005, S.139) und bewegten visuellen Szenen ist also tief in der Kulturgeschichte des Menschen verankert (Daniels, 2004). Dies verdeutlicht, wie groß die Faszination dieser Verbindung ist und welche besondere Ausdruckskraft ihr innewohnt. Michel Chion erkannte: „Ton zeigt das Bild anders, als es sich allein zeigt, gleichsam lassen Bilder Töne anders erscheinen, als wenn sie alleine auftreten würden“ (Chion, 1994, S.21).

Akustisches und Visuelles kann sich also zu einem neuen, in sich geschlossenen Medium verbinden, dessen Wirkung auf den Rezipienten größer ist, als die Summe seiner Einzelteile Ton und Bild. Doch was genau steckt hinter dieser Verbindung? Wodurch entsteht sie und was beeinflusst sie? Was sind die relevanten Mechanismen? Unter welchen Bedingungen verschmelzen Ton und Bild zu einer Einheit, bzw. welche Bedingungen sind eher hinderlich? Die Aktualität dieser Fragen manifestiert sich an der rasanten Entwicklung audiovisueller Medien und damit nicht zuletzt auch am Musikvideo als ein innovatives Element dieser Kommunikationsform.

Die Frage nach dem Zusammenspiel von Akustischem und Visuellem führt unmittelbar zur Auseinandersetzung mit der menschlichen Wahrnehmung: Ton und Bild als mediale Repräsentationen der Sinne Hören und Sehen werden beim Rezipienten (zusammen mit allen anderen Sinneseindrücken) zu einem Gesamterleben zusammengeführt. Durch diese Integration von Gesehenem und Gehörtem kommt es zu gegenseitigen Beeinflussungen zwischen den Sinnen. Daher ist es sinnvoll bei der Betrachtung audiovisueller Medien die akustische und visuelle Wahrnehmung als interagierend anzusehen und gemeinsam zu betrachten. Nach aktuellen Untersuchungen im Bereich der Neurokognition kann davon ausgegangen werden, dass diese Integration der Sinnesmodalitäten nicht bloß „semantisch und kulturell erlernt“ (Floto/Lepa, 2003, S.2) ist, sondern Verknüpfungen bereits auf sehr frühen Wahrnehmungsebenen stattfinden, teilweise noch vor der Objekterkennung.

## 1.1 Forschungsinteresse der Diplomarbeit

Das Forschungsinteresse dieser Arbeit konzentriert sich vor allem auf Interaktionen der frühen, perzeptuellen Wahrnehmungs-Ebenen. Untersucht werden sollen vor allem Verbindungen zwischen Ton und bewegtem Bild, die auf den originären Eigenschaften von akustischen und visuellen Reizen basieren. Gibt es wissenschaftlich belegbare Zusammenhänge zwischen Klangeigenschaften des Schalls, wie Tonhöhe, Lautstärke, Definition, Reinheit, Volumen und Klangfarbe bzw. den musikalischen Parametern wie Melodik, Harmonik, Rhythmik und Dynamik, sowie den optischen Eigenschaften des bewegten Bildes, wie Helligkeit, Kontrast, Farbe, geometrische Form, Textur und Animation? Wenn ja, wie sehen diese aus und sind sie auf perzeptuelle Prozesse zurückzuführen? Können aus den Ergebnissen kognitionswissenschaftlicher Forschung Erkenntnisse für die Gestaltung audiovisueller Medien gewonnen werden?

## 1.2 Ziel der Diplomarbeit

Ziel dieser Diplomarbeit ist es zum einen, Forschungsergebnisse der Kognitionswissenschaften und der Wahrnehmungspsychologie auf dem Gebiet der multisensorischen Wahrnehmung zu finden, die bei der Rezeption audiovisueller Medien eine Rolle spielen, und diese zu analysieren. Es soll darüber hinaus gezeigt werden, dass Ästhetik ein Bestandteil von Wahrnehmung ist, der an der Konstruktion von Erleben beteiligt ist und demzufolge audiovisuell Ästhetik eine wichtige Rolle bei der Konstruktion medialer Wirklichkeiten spielt.

Im praktischen Teil wird zum anderen eine Musikvisualisierung (Visual Music) erarbeitet, die einige der zuvor analysierten Untersuchungsergebnisse aufgreift und die dargestellten Interaktionen zwischen akustischer und visueller Wahrnehmung als Relation zwischen Musik und bewegtem Bild umsetzt. Hierdurch werden wissenschaftliche Ergebnisse in audiovisuelle Gestaltung transformiert. Dieser Video-Clip stellt damit ein künstlerisches Experiment dar, welches der Frage nachgeht, ob eine derartige Umsetzung wissenschaftlicher Ergebnisse in mediales Erleben möglich ist. Ziel dieser Umsetzung ist dabei auch zu erörtern, ob es eine Art perzeptuelle Ästhetik gibt, also eine, die nicht, oder

nur sehr begrenzt semantische, narrative und linguistische Konzepte anspricht, sondern frühe, perzeptuelle Stufen der Wahrnehmung rekurriert. Eine solche Ästhetik wäre weitgehend unabhängig von Interpretationen des Rezipienten und von Bedeutungen, die er den Reizen zuschreibt.

Der geschilderten Sicht auf dieses Thema liegt auch die Idee zugrunde, dass Ästhetik, Kunst und Gestaltung stets Aufschluss über den Menschen geben, da sie Teil seiner Natur und Ausdruck seiner Existenz sind. Hier liegt eine Schnittstelle zur Wissenschaft, da deren Ziel der Erkenntnisgewinn ist. Ebenso wie ein/e KünstlerIn auf dem Weg der Kunst zu einer Erkenntnis, beispielsweise über den Menschen, gesellschaftliche Prozesse oder sich selbst, kommen will, sucht die Wissenschaft nach Lösungen für Probleme, obgleich ihre Methoden häufig grundlegend verschieden zu sein scheinen. In dieser Diplomarbeit werden beide Methoden angewandt, indem ein theoretischer Teil wissenschaftliche Ergebnisse darlegt und analysiert und ein praktischer Teil -auf Grundlage dieser Ergebnisse- den Versuch einer künstlerischen Umsetzung unternimmt.

### 1.3 Aufbau und Gliederung

Entsprechend der o.g. Zielsetzung gliedert sich diese Diplomarbeit in folgende Kapitel:

Nach der Einleitung in Kapitel 1 wird in Kapitel 2 der Begriff Multimedia definiert und erläutert. Es wird geklärt, was Medien sind, wozu sie dienen, und welche gesellschaftliche Funktion sie haben, sowie ein Bezug zu Wahrnehmung und Ästhetik hergestellt. Im Hinblick auf Ton-Bild-Relationen in audiovisuellen Medien werden das fernsehspezifische Genre Musikvideo sowie Visual Music als eine Form der Medienkunst erläutert und ihr Einfluss auf die ästhetische Entwicklung der Medien insgesamt diskutiert.

Kapitel 3 erläutert die Grundlagen der menschlichen Wahrnehmung. Es wird zunächst ein Überblick über die verschiedenen menschlichen Sinne gegeben und anschließend ausführlich auf das Sehen und Hören eingegangen. Dazu werden die physikalischen Eigenschaften von Licht und Schall beschrieben, sowie die

Funktionsweisen, Fähigkeiten und Einschränkungen der akustischen und visuellen Wahrnehmungsapparate erläutert.

Es folgen neuronale Grundlagen, in denen der Aufbau und grundlegende Funktionsprinzipien des menschlichen Gehirns erläutert werden. Begriffe wie Aktionspotential, Neuronen und Synapsen werden erklärt, um kognitive Prozesse und neuronale Verarbeitung verständlich zu machen. Dabei steht der cerebrale Cortex als Sitz all jener kognitiven Leistungen, die den Menschen auszeichnen, im Mittelpunkt der Betrachtung.

Aufbauend auf diesen Grundlagen werden in Kapitel 4 multisensorische Prozesse erläutert. Hier werden kognitionswissenschaftliche Forschungen und deren Ergebnisse dargestellt, die sich mit der Interaktion zwischen akustischer und visueller Wahrnehmung befassen. Dies sind in erster Linie empirische Studien, in denen einer Anzahl Testpersonen verschiedenste Reize dargeboten werden und durch Reaktionsmessungen und/oder anschließende Befragung statistische Daten erhoben werden. Daraus lassen sich dann allgemeine Schlüsse über die Funktionsweisen der menschlichen Wahrnehmung ziehen.

Kapitel 5 befasst sich mit dem Thema Synästhesie. Dieses Phänomen beschreibt Menschen, die zu bestimmten Reizen zusätzlich eine Wahrnehmung in einer weiteren Modalität haben, also beispielsweise zu Tönen ein visuelles Erleben haben. Dies ist vor allem deshalb interessant, da von Synästhesien gegebenenfalls Erkenntnisse über die menschliche Wahrnehmung im Allgemeinen abgeleitet werden können. Dazu wird erläutert, welche neuronalen Mechanismen Synästhesien zu Grunde liegen könnten und dass es sich nicht um Einbildungen oder übermäßige Phantasien handelt.

Kapitel 6 geht auf einzelne Aussagen des Medientheoretikers Michel Chion ein. Chion beschäftigt sich bereits seit den 1980er Jahren mit dem Wirkungszusammenhang von Ton und Bild im Bezug auf das Kino. Seine Erkenntnisse leitet er als Komponist und Filmkritiker vor allem aus persönlichen Erfahrungen und Beobachtungen ab. Der Vergleich mit den in Kapitel 4 und 5 erläuterten kognitionswissenschaftlichen Ergebnissen erweist sich dabei als äußerst spannend.

Kapitel 7 behandelt die praktische Arbeit. Es wird erläutert, welche wissenschaftlichen Ergebnisse und theoretischen Thesen darin aufgegriffen wurden. Ausführlich wird beschrieben, wie diese in praktische Gestaltung umgesetzt wurden. Dies umfasst die Komposition der Musik, die Wahl von Tönen und Klängen sowie die gesamte Bildkomposition aus Formen, Farben und Animation. Jedes Gestaltungselement steht hierbei in Verbindung zu wissenschaftlichen Erkenntnissen und Theorien.

Die Arbeit schließt in Kapitel 8 mit einem Fazit und einem Ausblick.

## 2 MULTIMEDIA

### *Medien und Kommunikation*

„ [...] Die Begriffe Medium und Kommunikation [sind], zumindest in ihrer heute weit verbreiteten Bedeutung, wesentlich jünger als die Sachverhalte oder Vorgänge, die mit ihnen im Allgemeinen bezeichnet werden.“ (Wilke, 2000, S.1)

In der Publizistikwissenschaft werden unter „Medien“ im engeren Sinn jene technischen Mittel verstanden, die zur Verbreitung von Aussagen an ein potentiell unbegrenztes Publikum geeignet sind, also Presse, Hörfunk, Film und Fernsehen. Damit hat der Medien-Begriff einen technischen Aspekt. Neben der Technik selbst werden darüber hinaus auch die Produkte dieser Technik bezeichnet.

Medien dienen somit der Kommunikation zwischen Menschen (Humankommunikation), durch ihre potentiell unbegrenzte Anzahl an Teilnehmern können sie auch als Massenmedien bezeichnet werden, die der Massenkommunikation dienen (Wilke, 2000, S.1ff.).

Neben dieser technischen Dimension haben Medien eine kulturelle, soziologische und psychologische Bedeutung. Die menschliche Wahrnehmung ist abhängig vom Körper und dessen Fähigkeiten, die Dinge quantitativ und qualitativ zu verarbeiten oder Ereignisse in Raum und Zeit zu verorten. Sie ist dadurch auf natürlichem Wege begrenzt, denn der Mensch kann zunächst nur erfahren und erleben, was in seiner Reichweite ist. Medien sind Werkzeuge zur Ausdehnung dieser Reichweite, ebenso wie es prähistorische Kulturtechniken wie der Werkzeug- und Symbolgebrauch sind. Die menschliche Sinneserfahrung ist zwar unersetzbar, aber sie kann durchaus ‚virtualisiert‘ werden. Das heißt, dass Simulationen des Erfahrbaren möglich sind. Das ist Grundlage jeder Medienerfahrung, denn ein Bildschirm zeigt beispielsweise kein Bild, sondern bloß ein Pixelmuster, das der Mensch interpretiert und ihm einen Sinn sowie Bedeutung gibt. „Medien dienen grundsätzlich der Erweiterung der menschlichen Wahrnehmung in Raum und Zeit. Indem sie Wahrnehmungswelten ermöglichen, sind sie viel mehr als bloß Mittel und Verstärker der menschlichen Kommunikation.“ (Hartmann, 2008, S.15)

In einer „Medienkultur“ (Hartmann, 2008, S.9) werden durch Bilder, Töne und Texte Informationen von anderen Orten und aus anderen Zeiten vermittelt, wodurch Wahrnehmungen simuliert und Erfahrungen substituiert werden.

### *Multimedia: Viele Medien?*

Multimedia ist seit den 1990er Jahren ein universell eingesetzter Begriff für die Integration multipler Medienformate wie Text, Bild, Animation, Video und Audio und steht für „Multiple Content Media“, was bedeutet, dass sich Information aus verschiedenen Informationskanälen zusammensetzt. „Multimedialität liegt vor, wenn unterschiedliche Sinne des Menschen gleichzeitig durch integrierte Medienanwendungen angesprochen werden.“ (Hartmann, 2008, S.19)

Die Massenkommunikation des 20. Jahrhundert war geprägt durch audiovisuelle Medien. Begonnen hatte dies mit der Entwicklung des Kinos gegen Ende des 19. Jahrhunderts. Dadurch wurde neben den gedruckten Text, dem „privilegierte[n] Medium westlicher Rationalität“ (Hartmann, 2008, S.15), eine neue Form der Kommunikation gestellt. Dies erzeugte neue Wahrnehmungsmöglichkeiten in einer Medienwirklichkeit, die zuvor seit Erfindung des Buchdrucks durch Literatur und Presse dominiert wurde. Wissen und Information reproduzierte sich nun nicht mehr exklusiv über Bücher und andere Texte, sondern auch über technische Bilder und audiovisuelle Verfahren. „Aus dem Film wurde [...] ein Instrument, das die Wahrnehmung des Menschen tiefgreifend veränderte.“ (Wilke, 2000, S.312).

Das Kino wurde zwar nicht erst mit Einführung des Tonfilms 1927 multimedial, denn vorher wurden Filme durch direkt gespielte Musik eines Orchesters, diverse Geräusche durch Kinoorgeln auditiv sowie durch Texttafeln literarisch ergänzt. Dennoch stellte diese technische Weiterentwicklung einen wichtigen Schritt dar, da durch sie ein synchrones Abspielen von aufgenommenem Ton möglich wurde und damit erstmals das Sehen und gleichzeitige Hören von gesprochenen Dialogen. Das Veränderte das Medium Film grundlegend und schuf die Grundlage für alle ihm folgenden audiovisuellen Medien (Wilke, 2000, S.310f.).

Der Film fand zunächst nur im Kino statt, also im öffentlichen Raum. Zwischen Ende des 19. Jahrhunderts und dem beginnenden 20. Jahrhundert entwickelte sich die Übertragung zunächst nur akustischer Signale über elektromagnetische Wellen und daraus der Hörfunk. Die Erfindung der Nipkow-Scheibe und der

Braunschen Röhre legten Grundsteine für die Übertragung von Bildern auf elektronischem Wege. In den 1950er Jahren begannen die massenhafte Verbreitung des Fernsehens und sein Aufstieg zum Leitmedium. Damit erreichten audiovisuelle Medien auch den privaten Raum. Film und Fernsehen haben dem bewegten Bild zu einer medialen Dominanz verholfen. „Man kann das Medienrevolution nennen: Neben das Lesen trat die Rezeption optisch-ikonischer Zeichensysteme“ (Wilke, 2000, S.303).

Um die Jahrtausendwende geschah ein weiterer Umbruch der Medienkultur durch das Zusammenwachsen von Telekommunikationstechnologien und Computer. Die rasch fortschreitende Entwicklung der Digitaltechnik führte zu einer Konvergenz der Kanäle und Integration der Formate durch multimodale Codierung. In einer digitalen Medienkultur werden Informationen nicht mehr isoliert angeboten, sondern bereits auf technischer Ebene verknüpft. Dadurch entstand ein neuer Multimedia-Begriff: „Die Frage nach der technischen Codierung führt zu einer Definition von Multimedia. [...] Das Medium der Integration von unterschiedlichen Medien ist der Computer.“ (Hartmann, 2008, S.19)

Multimedia bezeichnet einerseits eine neue Technologie, die Produktion auf Basis des Computers, als auch eine neue kulturelle Form, die Distribution auf Basis des Internets. Durch die massenhafte Verbreitung des Computers entsteht ein neuer Umgang mit Medien, der zunehmend bestimmt ist durch Interaktivität und Online-Kommunikation. Für Bilder und Videos gibt es zahlreiche kostenlose Portale, das Schreiben von *emails* gehört zur alltäglichen Kommunikation ebenso, wie *social networking* Portale, auf denen *user* (multimediale) Profile erstellen. In *weblogs* diskutieren Wissenschaftler, Journalisten und Autoren. Online Enzyklopädien wie Wikipedia, an denen jeder mitschreiben kann, ersetzen oft gedruckte Bände wie den des Brockhaus Verlag und in der Freizeit werden *multiplayer online games* gespielt (Hartmann, 2008, S.8ff.).

Die audiovisuelle Prägung der medialen Wirklichkeiten durch Kino und Fernsehen setzt sich in diesem veränderten Multimedia-Begriff fort. Nach anfänglichen technischen Einschränkungen durch Rechenleistung und Netzbandbreiten, hielten in der jüngsten Vergangenheit in zunehmendem Maße audiovisuelle Medien Einzug in die neue Netz- und Computer-Kultur. Dem tragen

auch die Rundfunkanstalten Rechnung, indem sie ihre Inhalte in Form von *Streams* und *Podcasts* online stellen. Demnach vollzieht sich im Internet ein ähnlicher Wandel von der Dominanz der Schrift und des Textes, über die (Digital-) Fotografie hin zum bewegten Bild und zu audiovisuellen Medien, für den ebenfalls technische Entwicklungen die Grundvoraussetzungen liefern. Der Wunsch nach audiovisueller Kommunikation scheint dem Menschen innezuwohnen. Dies kann in einer Langzeitbetrachtung als „Normalisierung“ (Wilke, 2000, S.303) angesehen werden, nämlich als die Rückkehr zur audiovisuellen Prädominanz, die 400 Jahre durch die Erfindung des Buchdrucks unterbrochen war.

### *Multimediale Ästhetik*

Menschliche Sinnestätigkeit lässt sich beschreiben „als Wahrnehmung auf verschiedenen Ebenen, mithin als Produktion einer bestimmten Ästhetik, basierend auf der biologischen Ausstattung des Menschen.“ (Hartmann, 2008, S.13) Ästhetik ist abgeleitet vom griechischen *aisthesis*, eine Bezeichnung, die nicht für „das Schöne“ steht, sondern ganz allgemein für das, was auf die Sinne wirkt. Somit bestimmen die menschlichen Sinne und die Ästhetik die Vorstellungen von Wirklichkeit (Hartmann, 2008, S.13).

Durch die Dominanz der audiovisuellen Medien hat deren Ästhetik den Umgang mit Medien im Allgemeinen und damit auch die Wahrnehmung geprägt. Seit Anfang der 1930er liegt das Basisinstrumentarium der audiovisuellen Ästhetik in Form der Gestaltung und Montage des Bildes, sowie der Montage von Bild und Ton, bereit. Dieses wurde in vielfachen Variationen wiederholt, in geringem Maße erweitert, jedoch in seinen Grundzügen nicht verändert. Durch diese Technik ist im Film fast alles möglich: Zusammenhänge zerreißen, Wahrnehmungen begründen, Widersprüche versöhnen, Konsistenz zerschlagen, Harmonie aufsprengen sowie Gegensätze aufheben (Schnell, 2000, S.98). Die Montage ist das filmische Organisationsprinzip, mit der die Produktionsseite Entscheidungen über die Qualität der diskontinuierlichen Bildfolge trifft. Dies sind künstlerische Entscheidungen von politischer Qualität, denn sie bestimmen die Ästhetik des Films, erzeugen Spannungsmomente oder stellen Ruhe, Besinnung und Kontemplation her. Kraft seiner einprägsamen Sprache

übersteigen die Wirkungsmöglichkeiten des Films die aller anderen Medien um ein Vielfaches (Schnell, 2000, S.98 & S.192f.).

Das Fernsehen bedient sich im Wesentlichen derselben Gestaltungsprinzipien, hat allerdings darüber hinaus eine eigene Ästhetik entwickelt. Dies ist bedingt durch die Diversifikation der Formate (Spielfilm, Dokumentation, Serie, Nachrichten, Sport, Live-Sendung etc.) und das andere Konsumverhalten gegenüber dem Kino.

Eines dieser TV-spezifischen Formate ist das Musikfernsehen. Im Hinblick auf die Fragestellung dieser Diplomarbeit, die die Interaktion zwischen akustischer und visueller Wahrnehmung als Untersuchungsgegenstand hat, werden im Folgenden die Ästhetik und deren Einfluss auf andere Medien genauer untersucht.

### *Das Musikvideo und seine ästhetische Bedeutung*

Video-Clips stellen das erfolgreichste der wenigen TV-spezifischen Genres dar. Seit Beginn der 1960er repräsentiert diese Gattung den Versuch, eine genuin fernsehästhetische Form der visuellen Umsetzung von Musik zu bieten. In den Anfängen gelang das nur mühsam; Musikvideos bestanden meist daraus, die Künstler beim Spielen ihres Musikstücks zu zeigen (Schnell, 2000, S.229f.). Die 1975 von Lasse Hallström gedrehten Video-Clips zu den ABBA-Songs „Bang-a-boomerang“, „SOS“ und Mamma mia“ zeigen erste Ansätze einer eigenen Ästhetik. Hallström fand einen Ansatz, für jedes der Lieder ein eigenes Vokabular zu entwickeln, das die Struktur des Stücks zugleich auf der Ebene der Bilder interpretiert, ohne in starre Schemata der Wiederholung zu verfallen. Zwei der genannten Videos spielen im Freien, die Blickwinkel sind variabel und der Regisseur experimentiert mit ungewöhnlichen Kamera-Einstellungen, es gibt Sequenzen mit schnellen Schnitten in Übereinstimmung mit der Musik, Zooms und optische Effekte, wie Spiegelungen und Kaleidoskop (Keazor/Wübbena, 2007, S.62). Diese und ähnliche Clips dienten zunächst dem Zweck, den Kosten- und Zeitaufwand von Fernsehauftritten, zu denen Künstler anreisen mussten, zu minimieren und sie dennoch im Programm zu platzieren, obgleich die genannten Videos von ABBA bereits darüber hinausgehen, indem sie die Künstler inszenieren und somit Identifikationsmöglichkeiten für den Fan schaffen.

Das Phänomen erfuhr einen deutlichen Entwicklungsschub durch die Gründung von MTV in New York 1981. Der Sender nahm am 1. August um 12:01 mit dem

Video zu dem Guggles-Song „Video killed the radio star“ seinen Betrieb auf, was der Startpunkt einer Entwicklung der Verquickung von Jugend- und Pop-Kultur mit dem Fernsehen war. Die Symbiose aus Musikindustrie, Werbung, Fernsehen und Unterhaltung wurde in der Folge durch MTV noch enger. Waren Musikvideos zuvor Pausenfüller im regulären Programm gewesen, so wurden sie nun in einer scheinbar endlosen Kette aneinandergereiht präsentiert. Dadurch wurden sie in einen völlig neuen Kontext gestellt, der entscheidenden Einfluss auf ihre Produktion und ihr Erscheinungsbild hatte. Zum einen wurde die Clip-Produktion routinierter, zum andern wuchs die Konkurrenz untereinander. Um im Strom der Clips wahrgenommen zu werden, musste sich ein Video nun von den anderen abheben, wodurch die ästhetische und technische Qualität der Videos stetig gesteigert wurde (Keazor/Wübbena, 2007, S.66f.).

Bis in die späten 1990er Jahre entwickelte sich das Genre zu einer eigenen, „z.T. durchaus avantgardistischen Kunstform“ (Keazor/Wübbena, 2007, S.9). Es wurde auf immer vielfältigere Arten die Musik, der darin enthalten Text und das Bild miteinander verwoben und in Beziehung gesetzt. Dabei wurden Bezüge zu Film, Fernsehen und bildender Kunst hergestellt, aber auch zu Cartoons, Werbung, Computerspielen und anderen Musikvideos. Regisseure wie Michel Gondry, Chris Cunningham, Spike Jones, Anton Corbijn u.a. entwickelten jeweils eigene Bildsprachen. Die von der Firma Palm Pictures veröffentlichte DVD-Serie „Directors Label“ zeigt Zusammenstellungen der Werke der genannten und anderer Regisseure. Ferner widmeten sich Ausstellungen (z.B. im Düsseldorfer NRW-Forum „25 Jahre Videoästhetik“, 2004) dem Thema. Dies zeigt einerseits die kulturelle und ästhetische Bedeutung, andererseits wird auch erkennbar, dass das Genre seinen künstlerischen und wirtschaftlichen Zenit offenbar überschritten haben könnte. In der Tat wurden die Budgets für Videoproduktionen drastisch gekürzt, die Programme der Musiksender MTV und VIVA werden in zunehmendem Maß von Werbung, vor allem für Handyklingeltöne, bestimmt. Prominente Regisseure betätigen sich vermehrt in Werbung oder dem Kinofilm. Das heißt nicht, „dass der Clip in Kürze aussterben wird [...]“ sondern eher „[...] in der bekannten Form vielleicht nicht mehr lange weiter bestehen und sich in verschiedene Varianten und Spielformen ausdifferenzieren wird, die dann auch nicht mehr -wie bisher- an einzelnen

monopolisierten Quellen (den Musikfernsehsendern) bereitgestellt sind.“ (Keazor/Wübbena, 2007, S.12ff.).

Die fernsehästhetischen Qualitäten des Musikvideos wurden trotz der skizzierten Entwicklung lange Zeit verkannt. Dies hängt mit seinem Mischcharakter zusammen: Ein Videoclip ist keine Bild-Reproduktion der Alltagswelt, obwohl er diese zitiert. Ebenso ist er nicht an künstlerischen Traditionen orientiert, obwohl er diese aufnimmt. Alltag und Kunst werden so miteinander verschmolzen, dass aus dem Zusammenspiel ein spannungsreiches Mit- und Gegeneinander unterschiedlicher kultureller Paradigmen entsteht. Damit setzt das Musikvideo gängige Orientierungsmuster außer Kraft, es stellt Alltag gegen Kunst. Das Verbindungselement ist dabei die Musik. Sie setzt die unterschiedlichen Bildmaterialien in Verbindung, rhythmisiert sie, akzentuiert und pointiert, so dass ein ästhetisch eigenständiges Produkt entsteht. Dadurch bekommt es ein medienästhetisch anregendes Moment. „Die Geschichten, die erzählt werden, die Pointen, die sie setzen, die Wirkungen, die das Zusammenspiel mit der Musik hervorrufen, die simultanen Bild-Ton-Effekte, die Kontrapunktik der Bildarrangements - all das bietet, konzentriert auf weniger mehr als 3 Minuten Laufzeit, innovative Impulse von einer Dichte und Dynamik, wie sie die Geschichte des Films schon lange nicht mehr entwickelt hat.“ (Schnell, 2000, S.230)

Da das Musikvideo allein historisch auf den Film und das Fernsehen folgt und sich folglich aus diesen heraus entwickelte, gibt es Verbindungen zwischen den Formaten auf verschiedenen Ebenen. Zum einen wird auf etablierte Szenographien und Erzähltechniken von Film und Fernsehen zurückgegriffen, zum anderen kann im kürzeren und daher schneller zu produzierenden Musikvideo mit technischen Errungenschaften und Dramaturgien experimentiert werden kann, die dann erst Eingang in den vergleichsweise schwerfälligen Film finden. Bei der gegenseitigen Beeinflussung kann es „zuweilen [...] dabei dann auch zu jener Art Spiraleffekt kommen, bei dem das Element eines Films im Rahmen eines Videos aufgegriffen, dort modernisiert und dynamisiert wird, um schließlich in eben dieser mobilisierten Form wieder Eingang in den Film zu finden.“ (Keazor/Wübbena, 2007, S.167).

Wim Wenders sagte in einem Interview, das Handwerk des Regisseurs lerne man heute durch Werbespots und Videoclips. Die erzählökonomische Verknappung des Geschichtenerzählens, Konzentration auf Spannungsmomente, Computeranimation als Realitätsersatz tragen nach Wenders zur Ausbildung einer neuen Handschrift, der individuellen ‚Bildersprache‘ des Regisseurs bei, die der Entwicklung des Films zu Gute kommt (Schnell, 2000, S.232). Vorbehalte gegen das Musikvideo gab es dennoch, trotz einer reichen Bilderästhetik, wegen der klar kommerziellen Ausrichtung. Dabei sind eben jene kommerziellen Effekte ein Teil der Clip-Ästhetik. In Europa wurden die produktiven Impulse auch vom jungen deutschen Film genutzt, der im Streben nach einer eigenen Ästhetik gerade auch die Abgrenzung zu Hollywood suchte. Dort zu erkennende Einflüsse, die sich auf die Clip-Kunst zurückführen lassen, sind variable Zeitformate, die Emanzipation des Tons, offene dramaturgische Formen, die Miniaturtechnik, serielle Formen sowie die Kombination aus dokumentarischen und fiktiven Ausdrucksformen (Schnell, 2000, S.234f.).

### *Visual Music*

Visuelle Musik bezeichnet allgemein die Idee, Klänge und Musik in visuelles Erleben umzusetzen. Dabei geht es darum, Entsprechungen musikalischer Eigenschaften wie Melodie, Harmonie und Rhythmus auf visuellen Ebenen zu finden. Durch simultane Präsentation von Ton und Bild wird daraus ein multimodales Gesamtkunstwerk erschaffen. Es können folgende Methoden der Umsetzung unterschieden werden: (1) Mechanische Apparaturen, die Klänge oder Musik mittels vordefinierter Relationen in eine zugehörige visuelle Präsentation verwandeln (z.B. Farborgeln). (2) Computer-Algorithmen, die Audio-Daten gemäß vordefinierter Relationen in Video-Daten umwandeln, die einer visuellen Präsentation zugeführt werden und (3) die freie Interpretation und Assoziation des Künstlers.

Die Idee der visuellen Musik reicht kunsthistorisch weit zurück; Moritz vermutet sogar bis in die Antike: „Der Traum, eine Farbmusik für das Auge zu schaffen, die der tonalen Musik für das Ohr entspricht, datiert mindestens bis zu Pythagoras und Aristoteles in die Antike zurück [...]“ (Moritz, 1987, S.17)

In der frühen Neuzeit machten Künstler wie Leonardo da Vinci, Giuseppe Arcimboldo, Athanasius Kircher, oder Louis-Bertrand Castel mit ihren

Lichtprojektoren (Leonardo) und optischen Musikinstrumenten erste Versuche, durch die Simultanität von Bewegung und Farbe, Formen und Musik einen auf Klang und Optik zurückgreifenden Gesamteindruck zu erzielen (Keazor/Wübbena, 2007, S.56). Wagner formulierte um 1850 Theorien zu einem „Kunstwerk der Zukunft“, das die Einzelkünste Malerei, Tanz, Musik und Dichtung in einem „Gesamtkunstwerk“ vereinen sollte. Ende des 19. Jahrhunderts wurde mit Farborgeln und Lichtklavieren experimentiert, Apparaturen, die gemäß einer angeschlagenen Tonhöhe Licht einer den Tonhöhen zugeordneten Farbe in den Raum warfen (Daniels, 2004).

In den 1920er Jahren waren es Künstler der Avantgarde wie Walther Ruttmann, Viking Eggeling und Oskar Fischinger, die inspiriert von der modernen Technik und Geschwindigkeit im Zeitalter der Industrialisierung Musikvisualisierungen mit Hilfe der Filmtechnik schufen. Fischinger versuchte dabei von Beginn an, mit vorwiegend geometrischen, abstrakten Formen die Musik zu interpretieren (Fischer/Hubert, 2005). Die Werke der Künstler dieser Zeit können als Beginn dessen angesehen werden, was heute als Medienkunst verstanden wird.

In den 1950/1960er Jahren nutzen Videokünstler die neuen Möglichkeiten der elektronischen audiovisuellen Medien. Ein bekannter Vertreter ist Nam June Paik, der verschiedenste mediale Installationen und Skulpturen schuf, für die er u.a. Plattenspieler, Tonbandgeräte, Fernseher und Projektionen nutzte (Schnell, 2000, S.298).

Heute bieten die modernen Multimedia-Techniken vielfältige Möglichkeiten zur Realisierung von *visual music*. Häufig sind es Video-Clips die vermehrt elektronische Musik, vereinzelt auch klassische oder neue Musik visualisieren. Dabei kommen die digitale Videotechnik, mit ihren vielfältigen Methoden der Nachbearbeitung, sowie 2D- und 3D-Animationssoftware zum Einsatz. Möglich sind auch durch die Künstler selbstprogrammierte Software-Applikationen und Codes, die algorithmisch Video- in Audiosignale transferieren, als auch alte analoge Techniken, wie Overhead- und Filmprojektoren und vieles mehr.

Eine weitere Ausprägung von *visual music* ist das VJing. Unter diesem Begriff sind Video-Performance-Künstler zu verstehen, die in Kombination mit Musik, oder zu einer vorhandenen Musik (Live-Band, DJ, Orchester etc.) verschiedene Arten Bilder erzeugen. Dies geschieht in Echtzeit z.B. durch VJ-Software, die das mixen und filtern von verschiedenen Videoquellen, meist als Dateien

vorhandener kurzer Sequenzen, erlaubt. Das VJing ist weit verbreitet in der Club-Kultur und dient dort der visuellen Ergänzung der Musik. Es gibt darüber hinaus Künstler und Kollektive, die audiovisuelle Performances durchführen, bei denen Musik und Bild hierarchisch auf einer Ebene angesiedelt sind. Die Produktion erfolgt aufeinander abgestimmt, dabei rekuriert die visuelle Gestaltung auf die Musik ebenso, wie umgekehrt die Musik auf visuelle Elemente abgestimmt wird (Fischer/Hubert, 2005; Daniels, 2004).

*Visual Music* ist eine Kunstform, die heute die audiovisuellen Möglichkeiten moderner Multimedia-Technologien ausschöpft. Sie ist Medienkunst, welche die ästhetischen Paradigmen der Medien-Gesellschaft spiegelt und beeinflusst. Ihr Sinn und Ziel ist es, Relationen und Beziehungen zwischen der Domäne Klang, meist in Form von Musik und Visuellem zu schaffen. Daher ist sie relevant im Bezug auf die Fragestellung dieser Diplomarbeit und wird im praktischen Teil beispielhaft umgesetzt.

### *Zusammenfassung*

Medien sind (in der Publizistikwissenschaft) technische Mittel zur Verbreitung von Aussagen an ein potentiell unbegrenztes Publikum und dienen somit der Massenkommunikation. Neben dieser technischen Dimension sind sie von kultureller, soziologischer, politischer und psychologischer Bedeutung, da sie Information an andere Orte und Zeiten bringen und damit die Wahrnehmung des Menschen erweitern.

Multimedialität liegt vor, wenn unterschiedliche Sinne durch integrierte Medienanwendungen gleichzeitig angesprochen werden. Im 20. Jahrhundert wurde dies durch die Entwicklung des Kinos und des Fernsehens möglich. Audiovisuelle Medien gewannen dadurch an Bedeutung und wurden zu Leitmedien. Um die Jahrtausendwende brachte die beschleunigte Entwicklung der Digital- und Computertechnik einen Wandel der Medienkultur. Das Zusammenwachsen der Telekommunikationstechnologien mit dem Computer machte die technische Konvergenz der Kanäle möglich. Durch das Internet entwickelte sich ein neuer Umgang mit Medien hin zu Interaktivität und Online-Kommunikation.

Medienästhetik spielt eine wichtige Rolle innerhalb des Kommunikationsprozesses, denn die Sinne des Menschen und die Ästhetik bestimmen seine Vorstellungen von Wirklichkeit. Grundlage audiovisueller Ästhetik ist die Filmmontage, die in ihren Grundformen seit den 1930er Jahren existiert. Eine besondere Rolle bei der Entwicklung der Medienästhetik kommt dem Musikvideo zu. Durch die Verknüpfung von Musik, Pop- und Jugendkultur, Kunst, Werbung und Unterhaltung entwickelte dieses Genre ein hohes kreatives Potential, das Auswirkungen auf die Medienästhetik im Allgemeinen hatte.

Visuelle Musik als Kunstform reflektiert und moduliert medienästhetische und damit kulturelle Prozesse. Sie ist gespeist aus der kulturhistorisch bis in die Antike zurückzufolgenden Idee einer Umsetzung von Musik in visuelles Erleben. Teile von *visual music* sind ebenfalls in der Pop-Kultur angesiedelt, wo sie als VJing elektronische Musik in Clubs unterstützen. Die Essenz visueller Musik -das Erzeugen von Verknüpfungen von akustischer und visueller Wahrnehmung- trifft die Fragestellung dieser Diplomarbeit, weshalb sie im praktischen Teil exemplarisch umgesetzt wurde.

### 3 DIE MENSCHLICHE WAHRNEHMUNG

Wahrnehmung ist so alltäglich und selbstverständlich, wie essen, trinken und schlafen. Sie ist eine Grundvoraussetzung für das Überleben und hat sich im Laufe der Evolution nahezu perfekt an die Lebensbedingungen angepasst. Ihre Funktion ist die des Vermittlers zwischen der äußeren, realen und der inneren Welt, dem Ich. Sie liefert dem Individuum Informationen über die Beschaffenheit und das Verhalten seiner Umwelt und bildet somit die Basis seines Bewusstseins und Verhaltens. Als Vermittler oder Botschafter ist die Wahrnehmung indes weder objektiv, noch vollständig. Allein durch die Beschaffenheit der Sinnesorgane und ihrer Rezeptoren erfolgt eine Selektion und Bewertung der einströmenden Informationen, darüber hinaus prägen Vorannahmen und Wissen das Erleben auf kognitiven Stufen der Verarbeitung. Wahrnehmungsprozesse spielen daher auch bei der Rezeption von (audiovisuellen) Medien eine entscheidende Rolle. So werden die Eigenschaften und Einschränkungen des menschlichen sensorischen Apparats sowohl technisch für Kompressionsalgorithmen in der Übertragung und Speicherung von Medien-Daten genutzt, als auch im Kommunikationsprozess gestalterisch, ästhetisch und inhaltlich mit ihnen, der menschlich-subjektiven ‚Sicht auf die Welt‘, umgegangen.

Dieses Kapitel erörtert die physiologischen, psychologischen und neuronalen Grundlagen insbesondere der akustischen und visuellen Wahrnehmung. Dies umfasst die physikalischen und psychophysikalischen Eigenschaften von Licht und Schall, Aufbau und Funktion der Sinnesorgane Auge und Ohr sowie ihre Rezeptoren Retina und Cochlea, neuronale Strukturen des Zentralnervensystems, sowie die kognitive Verarbeitung.

#### 3.1 Definitionen

Die Psychologie definiert Wahrnehmung als den Prozess der Aufnahme, Selektion, Identifikation, Integration und Klassifikation von sensorischen Erfahrungen. Dieser Prozess lässt sich in drei Stufen unterteilen: Sensorische Prozesse, perzeptuelle Organisation und Identifikation/Wiedererkennung von

Objekten (Zimbardo, 2008, S.108). Der Mensch verfügt über folgende Sinne (Modalitäten):

- Visuelle Wahrnehmung: Sehen
- Akustische Wahrnehmung: Hören
- Olfaktorische Wahrnehmung: Riechen, Geruchssinn
- Gustatorische Wahrnehmung: Schmecken, Geschmackssinn
- Hautsinne: Sensibilität für Berührung, Druck und Temperatur
- Vestibuläre Wahrnehmung: Gleichgewichtssinn
- Kinästhetischer Sinn: Wahrnehmung über Bewegung und Stellung von Körperteilen
- Schmerzsinne

Im Folgenden werden die akustische und visuelle Wahrnehmung genauer betrachtet.

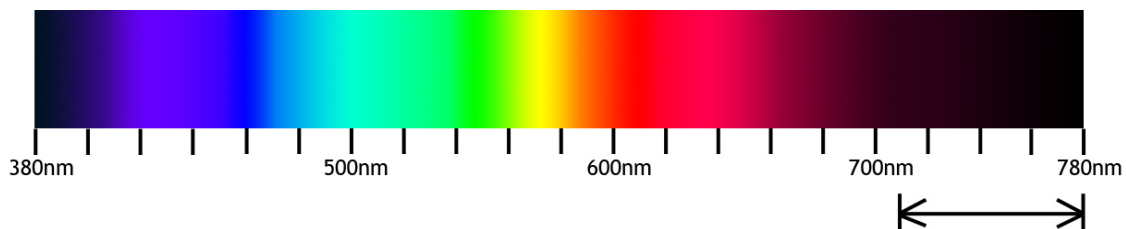
### 3.2 Physikalische und Physische Grundlagen

Es werden zunächst die wesentlichen Grundbegriffe der Optik und Akustik geklärt. Die physikalischen Eigenschaften des Licht und des Schalls, als auch Funktionsweise und Fähigkeiten der Sinnesorgane des Sehens und des Hörens sowie ihrer Rezeptoren werden dazu erläutert.

#### 3.2.1 Optik

Der physikalische Reiz der visuellen Wahrnehmung ist das Licht. Im allgemeinen Sprachgebrauch bezeichnet Licht den für den Menschen sichtbaren Teil des elektromagnetischen Spektrums. Damit folgt es denselben physikalischen Gesetzen wie Röntgenstrahlung und Radiowellen. Laut Quantenphysik kann es jedoch auch Eigenschaften von Teilchen annehmen, sogenannten Lichtquanten oder Photonen. Diese besitzen eine endliche Masse und geben ihre Energie beim Auftreffen auf Oberflächen quantisiert, also nicht kontinuierlich ab, was den Grundsätzen der Wellenphysik widerspricht. Ein Photon ist definitionsgemäß die kleinste absorbierbare Energiemenge. In der Optik ist jedoch das Wellenmodell für die meisten Betrachtungen ausreichend (Meschede, 2006, S.576).

Als elektromagnetische Welle lässt sich Licht durch zwei physikalische Eigenschaften klassifizieren: Die Wellenlänge, welche den subjektiven Farbeindruck vermittelt, sowie die Amplitude, welche Ausdruck der Intensität der Strahlung ist und somit den subjektiven Eindruck der Helligkeit mitbestimmt. Die Lichtgeschwindigkeit  $c$  ist eine der wichtigsten Naturkonstanten und beträgt  $299.792.458 \text{ m/s}$ , also rund  $300.000 \text{ km/s}$  (Meschede, 2006, S.505). Für den Menschen sichtbares Licht hat Wellenlängen zwischen  $380\text{nm}$  und  $780\text{nm}$ , das entspricht Frequenzen von  $790\text{THz}$  bis  $385\text{THz}$ . Monochromatisches Licht, das einer einzelnen Wellenlänge, wird als Spektralfarbe bezeichnet. *Abb. 3.1* zeigt den kontinuierlichen Verlauf der Spektralfarben mit den dazugehörigen Wellenlängen in Nanometern.



*Abb.3.1: approximierter Spektralfarbverlauf, durch Pfeile markierter Bereich= Farbtöne außerhalb des darstellbaren Farbraums (Anm.: Die Abbildung dient lediglich der Orientierung, da voll gesättigte Spektralfarben durch Druck- bzw. Displaytechnik nicht darstellbar sind.)*

Am linken Rand schließt sich die ultraviolette (UV) Strahlung an, am rechten Teil die infrarote (IR), oder Wärmestrahlung (Zimbardo, 2008, S.126).

Jede Farbe lässt sich durch die psychischen Dimensionen Farbwert, Helligkeit, Sättigung beschreiben. Die Dimension Farbwert erfasst den subjektiven Farbeindruck, das heißt, ob eine Farbe rot, blau, grün, gelb etc. ist. Wie oben erläutert bestimmt die Wellenlänge des Lichts den Farbwert. Die Dimension Sättigung erfasst die Reinheit und Klarheit der Farbempfindung. Voll gesättigte Farben sind reine Farben, die weder mit weiß noch mit anderen Farbtönen vermischt sind. Nicht vollgesättigte sowie unbunte Farben (schwarz, weiß, grau) setzen sich zu unterschiedlichen Mischverhältnissen und Anteilen aus einer Vielzahl von Wellenlängen zusammen. Die Dimension Helligkeit erfasst die Lichtintensität. Weiß besitzt die größte Helligkeit, schwarz die geringste (Zimbardo, 2008, S.126). Physikalische Messgrößen für die Helligkeit sind

- der Strahlungsfluss  $\Phi_e$  (von einer punktförmigen Lichtquelle ausgehende Strahlungsleistung)
- die Strahlstärke  $I_e$  (Strahlungsfluss im Raumwinkel)
- die Strahldichte  $L_e$  (der von dem Flächenelement abgestrahlte Strahlungsfluss pro Raum- und Flächeneinheit)
- die Bestrahlungsstärke  $E_e$  (empfangener Strahlungsfluss pro Flächeneinheit)

Durch die Strahlungsbewertung des menschlichen Auges werden diese physikalischen Grundgrößen in bewertete photometrische Größen transformiert:

- Lichtstrom  $\Phi_v$  (bewerteter Strahlungsfluss)
- Lichtstärke  $I_v$  (bewertete Strahlstärke)
- Leuchtdichte  $L_v$  (bewertete Strahldichte)
- Beleuchtungsstärke  $E_v$  (bewertete Bestrahlungsstärke)

(Meschede, 2006, S.569).

Die Strahlungsbewertungskurven des menschlichen Auges  $V(\lambda)$  für das Tag- und das Nachtsehen bezogen auf die Wellenlänge und auf 1 normiert zeigt folgende Abbildung 3.2:

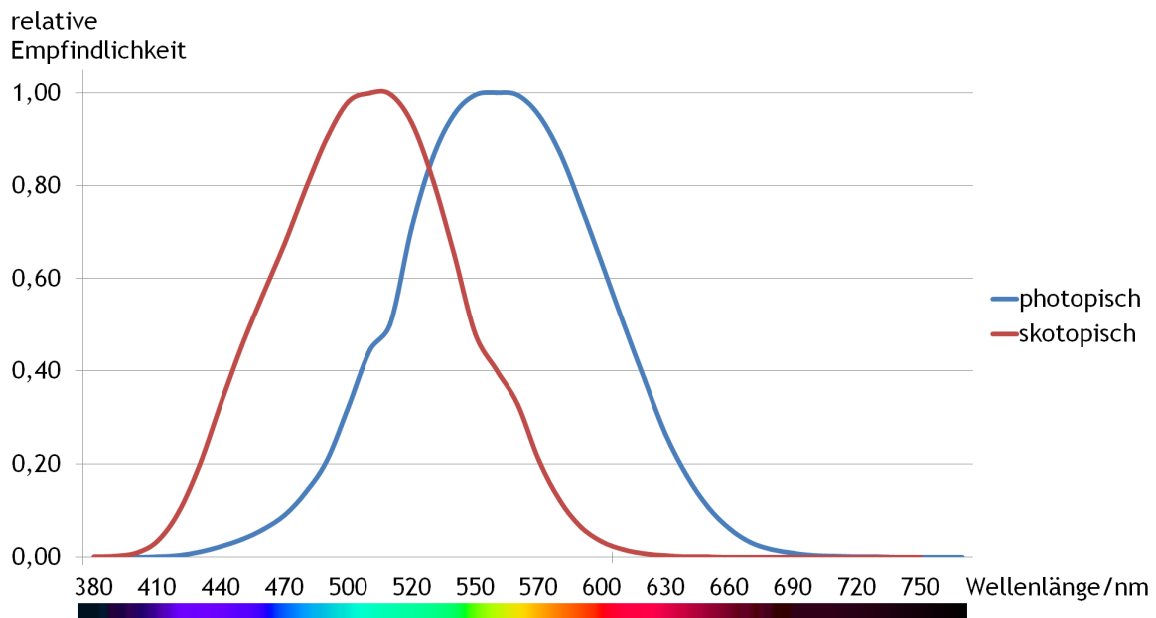


Abb. 3.2: Strahlungsbewertung des menschlichen Auges, auf 1 normiert. photopisch/Tagsehen (blaue Kurve) und skotopisch/Nachtsehen (rote Kurve)

Das menschliche Auge ist demnach nicht über dem gesamten Bereich des sichtbaren Lichts gleichmäßig empfindlich. Beim photopischen Sehen liegt die maximale Empfindlichkeit bei 560nm, was einer gelb-grünen Farbe entspricht. Für das skotopische Sehen liegt diese bei 500nm, was einer blau-grünen Farbe entspricht. Generell ist die Empfindlichkeit des Auges für mittlere Wellenlängen des sichtbaren Lichts (Grün) am höchsten und nimmt zu lang- und kurzwelligerem Licht ab. Beide Kurven sind auf 1 normiert, das heißt es wird die relative Empfindlichkeit abgebildet. Die absolute Empfindlichkeit des Nachtsehens ist indes größer, als die des Tagsehens. Photometrische Größen beziehen sich indes auf die photopische Kurve. (Pinel, 2007, S.174f.).

### Das Auge

Das äußere Sinnesorgan für das Sehen ist das Auge. Zwei Augäpfel sitzen beweglich und durch die Augenmuskeln steuerbar in den Augenhöhlen. Sie bestehen jeweils aus Hornhaut, vorderer Augenkammer, Pupille, Linse, Iris, Glaskörper und Retina.

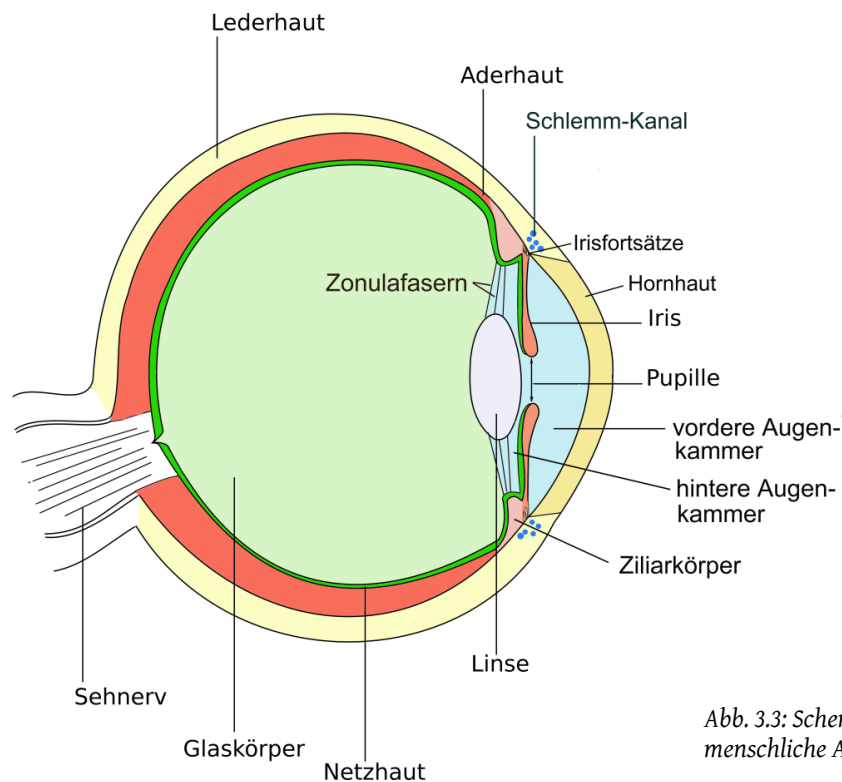
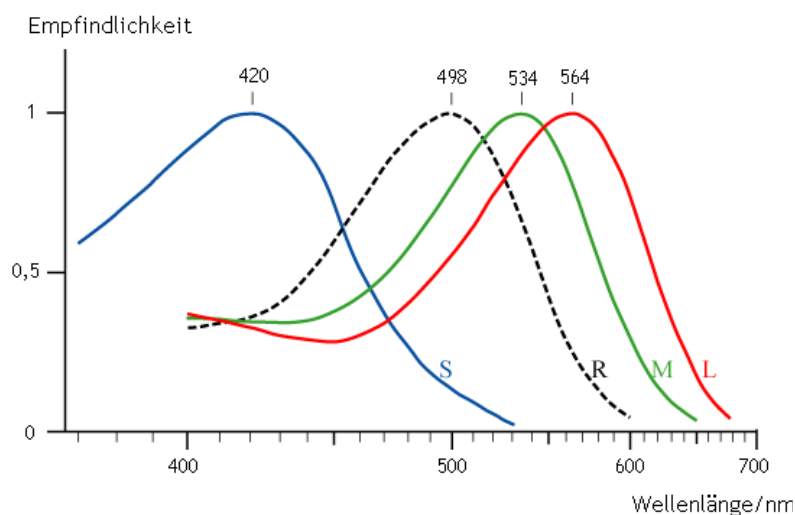


Abb. 3.3: Schematischer Querschnitt durch das menschliche Auge

Die Hornhaut ist eine transparente Ausstülpung nach vorn, durch die das Licht einfällt und die vordere Augenkammer durchquert. Dort trifft es anschließend

auf die lichtundurchlässige Pupille, deren Öffnung (Iris) der einfallenden Lichtintensität angepasst werden kann. Die dahinter liegende Linse ist elastisch und wird durch die Ziliarmuskeln unterschiedlich stark gekrümmt. So ist eine Fokussierung auf unterschiedliche Entfernungen möglich. Dieser Prozess der Akkommodation befähigt unbeeinträchtigte Menschen von ca. 7,6 cm vor ihrer Nasenspitze bis zur größtmöglichen Entfernung scharf zu sehen. Hinter der Linse durchquert das Licht den mit Flüssigkeit gefüllten Glaskörper und trifft schließlich auf die gegenüberliegende Netzhaut, die Retina. Diese setzt die Lichtreize in Nervenimpulse um, die vom Sehnerv weiter ans Gehirn geleitet werden (Zimbardo, 2008, S.119f.).

Die Retina ist mit zwei unterschiedlichen Typen von lichtempfindlichen Fotorezeptoren belegt, den Stäbchen und Zapfen. Die rund 120 Mio. Stäbchen dienen vor allem dem skotopischen Sehen bei geringen Lichtverhältnissen und sind wesentlich lichtempfindlicher als die ca. 7 Mio. Wellenlängen-selektiven Zapfen, die das Farbsehen ermöglichen. Das ist der Grund dafür, warum der Mensch bei schwachem Licht (z.B. in der Dämmerung) keine Farben unterscheiden kann. Ihre Lichtempfindlichkeit ist bei einer Wellenlänge von ca. 500nm am höchsten (siehe *Abb.3.4*).

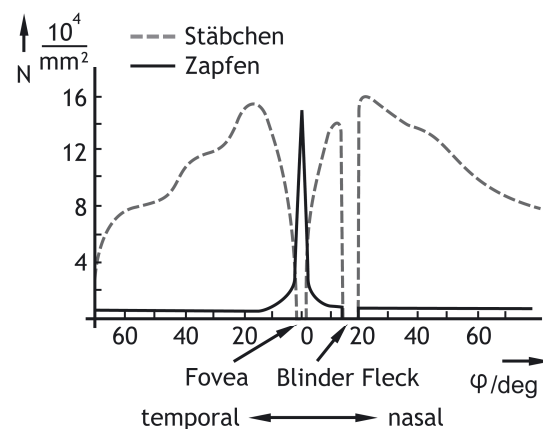


*Abb.3.4: Empfindlichkeitskurven der Zapfen (S-, M-, L-Rezeptoren) und der Stäbchen (R), normiert auf die maximale Empfindlichkeit des jeweiligen Rezeptors*

Die Zapfen unterteilen sich wiederum in drei Typen, dem S-Typ (Blaurezeptor), dem M-Typ (Grünrezeptor) und dem L-Typ (Rotrezeptor). Sie sind jeweils für bestimmte Wellenlängenbereiche empfindlich und ermöglichen somit das Farbsehen (*Abb.3.4*) (Bruce, 2003, S.22f.). Je nach spektraler Zusammensetzung des auftreffenden Lichts werden die drei Zapfen-Typen unterschiedlich stark angeregt. Aus der so entstehenden Intensitätsverteilung wird der Farbeindruck durch Interpretation im Gehirn konstruiert (Roth, 2005, S.17).

Durch die Empfindlichkeitskurven aller Rezeptoren zusammen ergibt sich eine maximale Empfindlichkeit des Auges im grünen Farbbereich. Dies hat zur Folge, dass grüne Farben bei gleicher Bestrahlungsstärke heller erscheinen, als rote oder blaue.

Die Dichte-Verteilung von Stäbchen und Zapfen auf der Netzhaut ist sehr inhomogen, wie in *Abbildung 3.5* zu sehen ist. Es existiert ein kleines Areal, die Fovea (Sehgrube), in dem nur Zapfen und keine Stäbchen vorhanden sind. Dies ist der Bereich des Scharfsehens. Im gesamten Bereich außerhalb der Fovea, der Peripherie, sind beide Rezeptor-Typen vorhanden, allerdings deutlich weniger Zapfen als Stäbchen. Die Netzhaut besitzt insgesamt ca. 6 Mio. Zapfen, von denen ein Prozent in der Fovea liegen



*Abb.3.5: Dichte der Stäbchen und Zapfen auf der Netzhaut*

sowie etwa 120 Mio. Stäbchen. Der blinde Fleck ist der Punkt auf der Netzhaut, an dem der Sehnerv anliegt. Er ist komplett licht-unempfindlich, da er keine Rezeptoren enthält. Die Fovea liegt  $5^\circ$  temporal der optischen Achse und macht nur etwa  $2^\circ$  des Sehfeldes aus, weshalb die Augen sich immer zu Objekten ausrichten, um sie scharf sehen zu können. Durch das dortige Fehlen von Stäbchen, die wesentlich lichtempfindlicher sind, fällt es schwer, bei geringen Beleuchtungsstärken feinere Strukturen zu erkennen und beispielsweise einen Text zu lesen (Goldstein, 2008, S.34f.).

Stäbchen und Zapfen sind näherungsweise kreisförmige Zylinder. Jeweils sieben Zapfen werden zu einem rezeptiven Feld in einer hexagonalen Anordnung zusammengefasst, einem in der Mitte und jeweils sechs umgebende. Dieses rezeptive Feld hat eine etwa 3,2-mal so große Grundfläche wie ein einzelner Rezeptor. Dadurch kommt es einerseits zur Verringerung der tatsächlichen Auflösung, allerdings liegt in den rezeptiven Feldern die Fähigkeit begründet, Konturen erkennen zu können und sie von Helligkeitsschwankungen zu unterscheiden. Es gibt allerdings nicht siebenmal so viele Zapfen wie rezeptive Felder, denn ein Zapfen gehört mehreren rezeptiven Feldern an. Nicht jeder Zapfentyp ist in gleicher Anzahl vorhanden. Es gibt im Verhältnis weniger Kurzwellenzapfen (S-Typ), was zu einer geringen Auflösung in Blautönen führt, als in anderen Wellenlängenbereichen. Hinzu kommt die chromatische Aberration der Linse: Kurzwelliges Licht liegt ca. 1-2 Dioptrien außerhalb des Schärfbereichs. Insgesamt ist dieser Bereich des Spektrums durch das Auge deutlich geringer aufgelöst als andere. Die höchste Auflösung erreicht das Auge im mittleren Wellenlängenbereich (Grün) (Reeves, 1994, S.17ff.).

Der kleinste auflösbare Bereich des visuellen Systems (Ricciòsche Bereich) liegt bei einer Bogenminute. Durch stereoskopische Disparitäten oder andere Effekte kann dieser jedoch bis 10 Bogensekunden oder weniger reichen. Das zeitliche Auflösungsvermögen lässt sich mit der kritische Dauer beschreiben. Laut Definition ist dies die Zeitdauer, unterhalb derer optische Reize gleicher Energie nicht unterscheidbar sind. Die Energie ist dabei das Produkt aus Zeit und Beleuchtungsstärke. Die kleinste, wahrnehmbare Beleuchtungsstärke beträgt ein Troland (td), nach der Quantentheorie trifft bei dieser Lichtmenge alle 240ms ein Photon pro Zapfen auf die Retina. Das entspricht einer kritischen Dauer von 34ms pro rezeptives Feld. In Versuchen wurde eine kritische Dauer von 40ms gemessen. Experimente zur Separierbarkeit zweier Blitze ergaben 18ms bei verschiedenen Orten, 35ms bei Blitzen am selben retinalen Ort (Reeves, 1994, S.20ff.).

### 3.2.2 Akustik

Akustische Reizungen werden als Schall bezeichnet. Physikalisch betrachtet ist Schall die Ausbreitung von kleinsten Druck- und Dichteschwankungen in einem elastischen Medium (Gase, Flüssigkeiten, Festkörper). Diese gehen mit Schwingungen der Teilchen (Moleküle) des jeweiligen Trägermediums einher. Durch die entstehenden Kräfte zwischen den schwingenden Molekülen (Abstoßung und Anziehung) kommt es zur Ausbreitung des Schalls als Welle. Die Ausbreitungsrichtung der Schallwelle ist gleich der Schwingungsrichtung der Moleküle seines Mediums (Longitudinalwelle), im Gegensatz zu Wasserwellen, deren Trägermedium orthogonal zur Ausbreitungsrichtung der Welle schwingt (Transversalwelle). Periodische Schwingungen haben eine Frequenz  $f$ ; Diese ergibt sich aus der Dauer einer Schwingungsperiode  $T$  in sec., die Zeit zwischen zwei Druckminima bzw. Druckmaxima. Es gilt  $f = \frac{1}{T} [Hz]$ . Die Frequenz  $f$  wird in Hertz (Hz) angegeben, 1Hz entspricht einer Schwingung pro Sekunde ( $Hz = \frac{1}{s}$ ). Die Intensität einer Schallwelle ist von ihrer Amplitude abhängig. Als Messgröße für die Intensität wird der von der Welle erzeugte Schalldruck  $p$  in Pascal [pa] herangezogen. Die Ausbreitungsgeschwindigkeit des Schalls in Luft bei 20°C ist  $c = 343 \text{ m/s}$  (Webers, 1989, S.54; Meschede; 2006, S.189ff.).

In der Psychoakustik bezeichnet Schall allgemein Geräusche, Klänge und Töne, die vom Menschen auditiv wahrgenommen werden können. Töne sind dabei reine Sinusschwingungen, die nur aus einer einzigen Frequenz bestehen, ihrer Grundfrequenz. Töne lassen sich eindeutig durch ihre Frequenz sowie ihre Amplitude beschreiben. Dabei bestimmen die Frequenz die empfundene Tonhöhe und der Schalldruck die empfundene Lautstärke. Reine Töne treten in der Natur allerdings nicht auf: Natürliche Schallquellen produzieren Klänge oder Geräusche. Diese bestehen aus vielen verschiedenen Tönen, die sich zu einer komplexen Schwingung überlagern. Klänge bestehen aus einem Grundton und vielen verschiedenen Obertönen, Vielfache der Grundschwingung. Sie haben eine identifizierbare Tonhöhe, die vom Grundton bestimmt wird. Die Zusammensetzung der Obertöne und deren Amplituden bestimmt die Klangfarbe eines Klangs. Geräusche hingegen haben keine identifizierbare Tonhöhe; Sie sind eine Superposition unendlich vieler Töne und besitzen in der Summe keine

harmonischen Schwingungen und deshalb auch keine herausragende Frequenz (Zenner, 1994, S.3; Meschede, 2006, S.192).

Die akustische Messgröße für die Lautstärke eines Schallereignisses ist der Schalldruckpegel  $L_p$ . Dieser ist das logarithmierte Verhältnis des quadrierten Effektivwertes des Schalldrucks  $p$  eines Schallereignisses zum Quadrat eines Bezugswerts  $p_0$  und wird in Dezibel (dB) angegeben. Es gilt:  $L_p = 20 \lg \frac{p}{p_0} [dB]$ .

Laut Definition gilt  $p_0=20\mu\text{Pa}$  als der vom Menschen kleinste wahrnehmbare Schalldruck (Hörschwelle) (Webers, 1989, S.100f.). Die Tonhöhe lässt sich stets mit der Frequenz ihrer Grundschwingung beschreiben, in der Musik wird jedoch das Notensystem verwendet, welches eine Oktave (Frequenzverhältnis 2:1) in 12 Halbtöne unterteilt (Webers, 1989, S.96). Oberhalb des vom Menschen wahrnehmbaren Bereichs schließt sich der Ultraschall, unterhalb der Infraschall an.

### Das Ohr

Das Sinnesorgan für das Hören sind die Ohren. Sie bestehen jeweils aus Außen-, Mittel- und Innenohr.

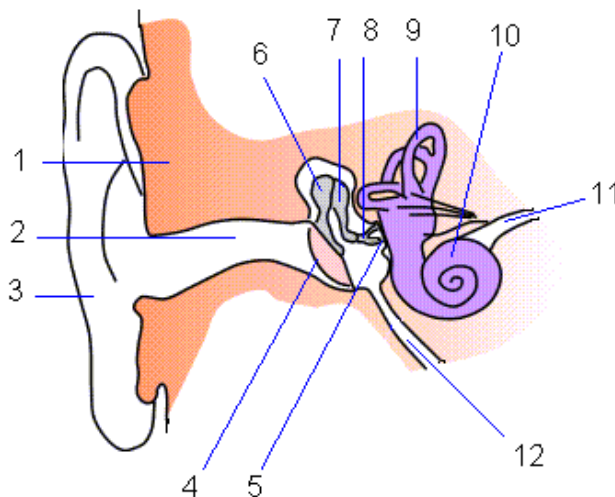


Abb. 3.6; Schematische Darstellung des Ohrs

Das Außenohr besteht aus Ohrmuschel (3), Gehörgang (2) und Ohrknorpel (1) und dient als verstärkender Schalltrichter sowie der besseren Richtungsbewertung. Diese erfolgt auf drei Arten: Erstens hat die Ohrmuschel eine Richtwirkung, zweitens durchläuft der eintreffende Schall verschieden lange Wege, die innere und äußere Rille, für die er unterschiedlich lange

Laufzeiten braucht. Die resultierenden Verzögerungszeiten führen zu drei Schallereignissen, die bei der Ortung helfen. Drittens kommt es zu richtungsabhängigen Frequenzanhebungen und -Absenkungen, den sogenannten richtungsbestimmenden Bändern, die vor allem für die Lokalisation in der Medianebene (vorn, oben, hinten, unten) entscheidend sind. Zum Mittelohr gehören das Trommelfell (4) sowie die Gehörknöchelchen Hammer (6), Amboss (7) und Steigbügel (8). Die Eustachische Röhre (12), auch Tube genannt, verbindet Mittelohr mit dem Nasenrachenraum, was den Druckausgleich zum Umgebungs-Luftdruck ermöglicht. Im Mittelohr findet eine mechanische Impedanz-Wandlung des Schalls statt. Die drei winzigen Gehörknöchelchen Hammer, Amboss und Steigbügel verstärken das Schallsignal durch die Flächenreduktion zwischen Trommelfell und Ovale Fenster (5) der Schnecke (10) um etwa 30dB. Eine weitere Verstärkung um ca. 10dB wird durch Hebelwirkungen erzielt. Das Innenohr besteht aus den Bogengängen (9), die dem Gleichgewichtssinn dienen, und der flüssigkeitsgefüllten Cochlea (Hörschnecke) (10). Der Steigbügel schwingt gegen die Membran im ovalen Fenster und versetzt so die Flüssigkeit im Inneren der Hörschnecke zum Schwingen. Diese ist durch die cochleare Trennwand längs in einen oberen und unteren Abschnitt geteilt, welche das Corti'sche Organ enthält. Die Basilarmembran stützt das Corti'sche Organ und vibriert als Antwort auf Schallereignisse. Die etwa 30.000 Haarzellen setzen diese mechanischen Schwingungen dann in Nervenimpulse um und leiten sie über den Hörnerv (11) an den auditiven Cortex weiter (Goldstein, 2008, S.267ff.; Haase, 1973, S.143).

Der Mensch ist in der Lage, Frequenzen von 20Hz bis 20kHz wahrzunehmen (die obere Grenze verschiebt sich mit zunehmendem Alter Richtung niedrigere Frequenzen), sowie Schalldrucke von ca. 10 $\mu$ Pa bis 100Pa zu verarbeiten. Das entspricht einem Frequenzumfang von etwa 10 Oktaven und einem Dynamikumfang (Verhältnisschalldruck) von ca. 140dB. Im Vergleich dazu liegt der Dynamikumfang (Signal/Rausch-Abstand) einer Audio-CD bei 96dB.

Wie das Auge, so ist auch das Ohr über das Frequenzspektrum nicht gleichmäßig empfindlich was zur Folge hat, dass das subjektive Lautstärkeempfinden stark frequenzabhängig ist.

Abbildung 3.7 zeigt die Kurven gleicher Lautstärke (Isophone): Diese Bewertungskurven zeigen an, welchen Schalldruckpegel ein Sinus-Ton haben muss, um gleich laut wie ein 1kHz Ton empfunden zu werden. Mit den Isophonen wird der Schalldruckpegel in die psychoakustische Messgröße Lautstärkepegel  $L$  [Phon] transformiert. Das Gehör ist am empfindlichsten bei ca. 3kHz, bedingt durch die Resonanzen des Gehörgangs, die individuell leicht unterschiedlich sind. Dieser Frequenzbereich ist zugleich ausschlaggebend für die Sprachverständlichkeit. Zu höheren und tieferen Frequenzen nimmt die Empfindlichkeit ab (Goldstein, 2008, S.263f.; Webers, 1989, S.100f.).

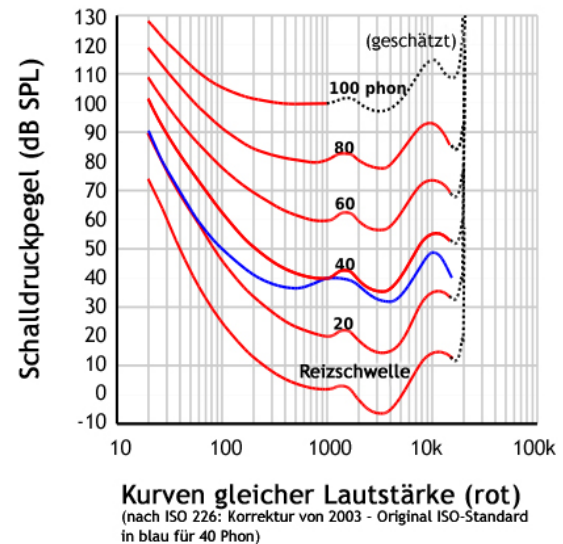


Abb. 3.7: Isophone

Das räumliche Auflösungsvermögen des Gehörs hängt stark von der Beschaffenheit des Schalls und der Richtung der Schallquelle im Bezug auf den Kopf ab, denn das Richtungshören basiert auf verschiedenen Phänomenen; Zum einen verfügt das Außenohr, wie oben beschrieben, über eine Richtwirkung sowie eine richtungsabhängige Frequenzbewertung. In der horizontalen Ebene spielt daneben vor allem das binaurale Hören eine zentrale Rolle. So werden die Pegel- und Zeitdifferenzen des eintreffenden Schalls zwischen beiden Ohren ausgewertet. Des Weiteren sind Beugungseffekte des Schalls am Kopf sowie Laufzeitunterschiede durch die Rillen der Ohrmuschel relevant. Insgesamt ist der Mensch dadurch in der Lage in der horizontalen Ebene Richtungsunterschiede von unter  $2^\circ$  auszuwerten (bei eintreffendem Schall von vorn). In der medialen Ebene wird vor allem der Frequenzgang des Schallereignisses ausgewertet. Je nach Frequenzspektrum des zu ortenden Signals beträgt die Auflösung hier zwischen 10 und 45 Grad (Eska, 1997, S.87f.; Webers, 1989, S.116ff.).

### 3.3 Neuronale Grundlagen

Um etwas wahrnehmen zu können bedarf es nicht nur der Sinnesorgane, denn diese sind letztlich bloß Sensoren, welche äußere Reize aufnehmen, in Nervenimpulse umsetzen und an das Gehirn weiterleiten. Die eigentliche Wahrnehmung, die Identifikation, die Interpretation, die Integration und Klassifikation aller eintreffenden Informationen, findet im Gehirn statt. Im Folgenden wird erläutert, wie die Informationsverarbeitung im Gehirn funktioniert, wie es aufgebaut und strukturiert ist und welche Funktionsprinzipien sich vor allem im Bezug auf Wahrnehmungsprozesse feststellen lassen.

#### *Neuronen, Synapsen, neuronale Netze*

Grundlage aller Informationsübermittlung und -verarbeitung bilden die Nervenzellen, die sogenannten Neuronen. Der ausgereifte menschliche cerebrale Cortex (Großhirnrinde) enthält etwa 100 Milliarden Neurone (Carlson, 2004, S.89), das gesamte Gehirn ca. 180 Milliarden (Goldstein, 2008, S.47). Neuronen sind auf Erregungsleitung spezialisierte Zellen, die elektrische Impulse empfangen und weiterleiten können. Dies geschieht durch sogenannte Membranpotentiale, elektrische Ladungen an den Zellwänden, die durch Trennung von elektrisch geladenen Teilchen (Ionen) entstehen. Es gibt drei Arten von Neuronen: Sensorische senden Informationen von Sinnesrezeptoren an das Zentralnervensystem. Sie sind hochspezialisiert, wie z.B. die Photorezeptoren der Netzhaut. Motorneurone leiten Informationen vom Zentralnervensystem weg, hin zu Muskeln und Drüsen. Das Gehirn enthält hauptsächlich Interneurone, welche Informationen von sensorischen Neuronen an Motorneurone oder andere Interneurone weiterleiten (Zimbardo, 2008, S. 75ff.).

Neuronen haben üblicherweise eine Vielzahl Eingangsverbindungen sowie eine Ausgangsverbindung. An den Eingängen erhält jedes Neuron zwei Arten von Reizungen, exzitatorische (erregende) und inhibitorische (hemmende), welche von ihm bilanziert werden. Das richtige Muster von exzitatorischen Inputs zur rechten Zeit oder am rechten Ort führt bei einem Neuron zur Erzeugung eines Aktionspotentials: Es „feuert“. Dies geschieht in Form einer kleinen elektrischen

Spannung von rund 40mV. Die Erregungsleitung ist grundsätzlich in beide Richtungen möglich, geschieht aber bedingt durch die chemischen Prozesse zur Polarisation bevorzugt in eine Richtung (Golenhofen, 2006, 37f.). Aktionspotentiale funktionieren nach dem „Alles-oder-nichts-Gesetz“. Das heißt, es gibt keine Abstufungen des Ausgangssignals in Abhängigkeit der Eingangssignale. Wenn ein Neuron feuert, erfolgt dies immer gleich, das Aktionspotential als solches ist nicht variant. Nach dem feuern benötigt das Neuron eine gewisse Zeit, bis es erneut ein Aktionspotential generieren kann. Dies ist die sogenannte Refraktärphase, während der selbst das richtige Eingangsmuster nicht zu einem Feuern führt. Diese beträgt ca. 1ms bis das Ruhepotential wieder endgültig erreicht ist, vergehen noch weitere 4ms. Die Stärke eines Reizes spiegelt sich in der Feuerungsrate wieder: Je öfter ein Neuron feuert, desto stärker ist der Reiz. Als Vermittler der erzeugten Impulse dienen die Synapsen an den äußeren Rändern der Neuronen. Sie sind die erregungsleitenden Verbindungen zwischen Nervenzellen und anderen Zellen des Körpers, wie Sinnes-, Muskel- oder Drüsenzellen, sowie zwischen Neuronen untereinander (Zimbardo, 2008, S.78ff.).

Die Informationsverarbeitung im Gehirn erfolgt in Form neuronaler Netze; Sensorische Reize werden zunächst von den Sinnesorganen in elektrische Nervenimpulse gewandelt (Transduktion). Über Nervenbahnen gelangen sie ins Gehirn, wo die neuronale Verarbeitung der eingehenden Signale stattfindet. Das heißt, gewisse Neuronen werden zum feuern angeregt, die über synaptische Verbindungen wiederum andere Neuronen anregen bzw. hemmen. Es handelt sich um komplexe Vorgänge, die das Antwortverhalten von Neuronen auf vielfältige Weise verändern. (Goldstein, 2008, S.5). Information hat in neuronalen Netzen somit die Form von Aktivierungsmustern, sogenannte sensorische Codes, die sich meist über viele verschiedene Hirnareale ausdehnen (Goldstein, 2008, S.94ff.). Dies ist die grundlegende Struktur und Informationsarchitektur des Gehirns und des Nervensystems bei Tieren und Menschen. Es gibt verschiedenste Modelle dazu, wie Information innerhalb dieser Netze repräsentiert wird. Als ein Beispiel soll hier das Modell von Finkel und Edelman skizziert werden. Sie beschreiben Wahrnehmung als informationsverarbeitendes System, dessen kleinste Einheiten sogenannte Units (oder *nodes*) bilden. Diese Units können Aktivierung empfangen, sie

transformieren und ggf. weiterleiten. Dabei bestimmt „das Ausmaß an Aktivierung einer Unit, zu welchem Grad diese Unit den informationellen Zustand, den sie repräsentiert, widerspiegelt.“ (Schneider, 1994, S.145). Funktional gleiche Units sind in Netzwerken zu sogenannten Schichten zusammengefasst. Einfache Verarbeitungsstufen bestehen meist aus zwei Schichten: In- und Output. Die Weiterleitung von Aktivierung entlang von ‚Verbindungen‘ (*connections*) erfolgt auf zwei Arten: ‚Exzitatorische Verbindungen‘ (erregende) sind solche, die Aktivität beim Empfänger erhöhen und ‚inhibitorische Verbindungen‘ (hemmende), die deren Aktivität senken. Je größer das Gewicht einer Verbindung, desto mehr Aktivierung fließt über sie. Drei Faktoren bestimmen die Aktivierung:

1. die Aktivierung zum vorherigem Zeitpunkt
2. die Aktivierung durch den Input von anderen Units und
3. die Verfalls-Funktion (*decay term*)

Units besitzen einen Schwellenwert, unter dem keine Aktivierung stattfindet. Lernen beruht demnach darauf, die Gewichtungen der Verbindungen zwischen Units und damit den Verlauf des Aktivierungsflusses zu modifizieren. Wiederholte Aktivierung eines Musters führt zu einer Verstärkung und damit zur Festigung des Musters (Lernen durch Repetition) (Werner X. Schneider, 1994, S.145ff.).

### *Gehirnstrukturen und ihre Funktionen*

Das Gehirn ist keine homogene Masse, sondern sowohl anatomisch, als auch funktional strukturiert. Unterschieden werden:

Hirnstamm aus

- Medulla oblongata  
(verlängertes Rückenmark)
- Pons (Brücke)
- Formatio reticularis
- Thalamus
- Cerebellum (Kleinhirn)

Limbisches System aus

- Hippocampus
- Amygdala (Mandelkern)
- Hypothalamus

Großhirn mit

- Corpus callosum (Balken)
- Cerebraler Cortex  
(Großhirnrinde)

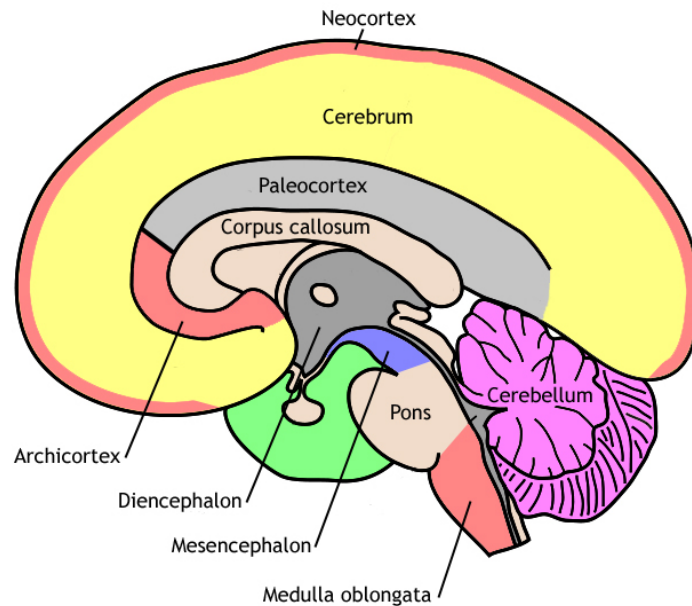


Abb. 3.8: Schematischer medialer Schnitt durch das menschliche Gehirn

Die Strukturen des **Hirnstamms** regeln in erster Linie interne Prozesse des Körpers. Die **Medulla oblongata** ist am oberen Ende des Rückenmarks angesiedelt und stellt das Zentrum für Atmung, Blutdruck und Herzschlag dar. Nervenfasern, die vom Körper hinauf ins Gehirn führen und umgekehrt, kreuzen sich in der Medulla oblongata. Somit ist die linke Hirnhälfte mit der rechten Körperseite verbunden, und die rechte Hirnhälfte mit der linken Körperseite. Über der Medulla oblongata liegt die **Brücke (Pons)**. Sie leitet alle ankommenden Informationen in andere Strukturen des Hirnstamms und Kleinhirns. Die **Formatio reticularis** ist ein dichtes Netzwerk aus Nervenzellen welches den Cerebralen Cortex anregt, die Aufmerksamkeit auf neue Situationen zu richten. Die Formatio reticularis besitzt lange Nervenfasern, die zum **Thalamus** führen, welcher eintreffende sensorische Information kanalisiert und zu den entsprechenden Bereichen des Cortex weiterleitet. Das **Cerebellum** koordiniert Körperbewegungen und -haltung sowie das Gleichgewicht (Zimbardo, 2008, S.92).

Das **limbische System** vermittelt zwischen motiviertem Verhalten, emotionalen Zuständen und Gedächtnisprozessen. Des Weiteren regelt es die Körpertemperatur, den Blutdruck, den Blutzuckerspiegel sowie andere Prozesse des Körperhaushalts. Es besteht aus drei Teilen; Der **Hippocampus** ist seine größte Struktur. Er spielt beim Erwerb expliziter Gedächtnisinhalte eine

wichtige Rolle. Die **Amygdala** ist maßgeblich an der Kontrolle und Formung emotionaler Gedächtnisinhalte beteiligt. Der **Hypothalamus** stellt eines der kleinsten Strukturen des Gehirns dar und ist aus verschiedenen Kernen und kleinen Neuronenbündeln zusammengesetzt. Er ist an der Regulation physischer Prozesse des motivationalen Handelns (Essen, Trinken, Temperatur, Sexualverhalten) beteiligt. Außerdem hält er das innere Gleichgewicht (Homöostase) und ist damit z.B. für das Zittern und die Verengung der Blutgefäße bei Kälte verantwortlich (Zimbardo, 2008, S.92f.).

Das **Großhirn** beim Menschen nimmt zwei Drittel der Masse des gesamten Gehirns ein und überragt somit alle anderen Teile. Es ist der phylogenetisch jüngste Teil und reguliert die höheren kognitiven und emotionalen Funktionen. Seine äußere Oberfläche ist der **cerebrale Cortex**, eine etwa 3mm starke Schicht aus Milliarden Nervenzellen. Sie besteht aus vielen Falten bzw. Hirnwindungen (Gyri, *sing.* Gyrus), die durch Sulci (Furchen) voneinander getrennt sind. Dadurch wird die Oberfläche des Cortexes vergrößert. Eine tiefe Längsfurche (Fissura longitudinalis) teilt das Großhirn in zwei fast symmetrische Hemisphären (linke und rechte), die über einen gewaltigen Strang Nervenfasern, dem Corpus callosum (Balken), miteinander verbunden sind. Dieser verbindet größtenteils identische Hirnareale der Hemisphären miteinander und macht eine effiziente Zusammenarbeit beider möglich (Carlson, 2004, S.93, S.98).

Beide Hemisphären besitzen zwei lange Furchen. Die **Sulcus centralis** (Zentralfurche) verläuft vertikal und die **Fissura lateralis** (Sylvische Furche) horizontal. Durch sie lassen sich auf jeder Seite vier Bereiche definieren (**Abb. 3.9**); Der **Frontallappen** (oder Stirnlappen) liegt vor der Sulcus centralis und oberhalb der Fissura lateralis. Er ist beteiligt an motorischer Kontrolle und kognitiven Aktivitäten wie Planen, Entscheiden und dem Setzen von Zielen. Hinter ihm liegt der **Parietallappen** (Scheitellappen) oberhalb der

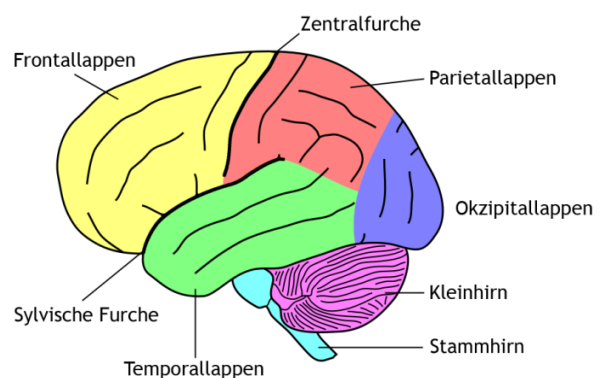


Abb. 3.9: Seitliche Ansicht des menschlichen Gehirns. Einteilung in vier Lappen

Fissura lateralis. In ihm werden Empfindungen wie Berührung, Schmerz und Temperatur verarbeitet. Am Hinterkopf liegt der **Okzipitallappen** (Hinterhauptslappen), in dem die Verarbeitung visueller Information stattfindet. Unterhalb der Fissura lateralis liegt der **Temporallappen** (Schläfenlappen). Dort werden akustische Informationen verarbeitet und es befindet sich das Wernicke-Areal, welches für das Sprachverständnis entscheidend ist (Zimbardo, 2008, S.93f.)

### *Visueller und auditiver Cortex*

Der visuelle Cortex ist der Bereich der Großhirnrinde, der sensorische Information von der Retina erhält und verarbeitet. Er wird in die primäre Sehrinde (V1) und sekundäre bzw. tertiäre (assoziative, V2-V5) Bereiche unterteilt und befindet sich am Okzipitallappen, an den inneren Oberflächen der cerebralen Hemisphären, genauer gesagt oberhalb und unterhalb des Sulcus calcarinus. Der auditive Cortex, der akustische Information empfängt, befindet sich an der inneren Oberfläche des Sulcus lateralis am Temporallappen (Carlson, 2004, S.211f., 220f., 248ff.).

### *Funktionsprinzipien des Gehirns: Konvergenz*

Die Neuronen im Gehirn arbeiten wie oben beschrieben nicht unabhängig voneinander, sondern im Verbund. Ein Neuron erhält von etwa 1.000 anderen Neuronen Signale, das heißt sein Antwortverhalten hängt von einer Vielzahl von Eingangssignalen ab (Goldstein, 2008, S.47).

### *Funktionsprinzipien des Gehirns: Spezialisierung (Modularität)*

Das Prinzip der Spezialisierung ist eng verbunden mit dem Aufbau des Gehirns, nämlich einer starken Unterteilung und Gliederung in Regionen. Diese Unterteilung ist auch eine funktionale. So sind bestimmte Areale (Module) spezialisiert auf das Lösen einer bestimmten Aufgabe. Im 19. Jahrhundert wurde die ‚Unspezifität‘ oder ‚Neutralität‘ der neuronalen Erregung gegenüber ihrem Inhalt entdeckt, das heißt die Art der Informationsverarbeitung im Gehirn durch neuronale Aktivitätsmuster ist grundsätzlich in allen Bereichen des Cortex gleich. Entscheidend dafür, ob es sich z.B. um einen Hör-, Seh-, Riech-, Schmeck- oder anderen Sinneseindruck handelt ist einzig der Ort, an dem diese

Aktivität stattfindet. So können elektrische Reize, die künstlich durch Elektroden in bestimmte sensorische Hirnregionen eingebracht werden, entsprechende Sinneseindrücke evozieren, z.B. eine Farbwahrnehmung durch einen Strom im visuellen Cortex.

Das Gehirn kann somit anhand seiner funktionalen Areale kartographiert werden; Der hintere (okzipitale) Teil sowie Anteile der Scheitel- und Schläfenlappen vermitteln den Seheindruck, wobei der Schläfenlappen für die Repräsentation des Raumeindruck zuständig ist. Der auditive Cortex befindet sich ebenfalls im Temporallappen. In den Scheitellappen befindet sich die Körperfühlsphäre, die Frontallappen gelten als Ort der Motorik und der Persönlichkeit. Körperfühlsphäre und Motorik liegen unmittelbar nebeneinander und sind nach Körperregionen geordnet. Es gibt dabei starke Entsprechungen aber auch Unterschiede zwischen der sensorischen und der motorischen „Landkarte“. Die unterschiedliche Ausdehnung der Hirnbereiche für bestimmte Körperregionen zeigt die Wichtigkeit dieser Körperregionen an, so sind beim Mensch große Areale z.B. für Finger und Zunge zuständig. Die Spezialisierung geht jedoch noch weiter in die Tiefe. So ist der okzipitale Cortex für die Analyse visueller Information weiter unterteilt: Beim Rhesusaffen lassen sich mindestens 40 Subregionen nachweisen, beim Menschen wird von noch mehr ausgegangen. Das Sehen ist demzufolge modular aus vielen Teilfunktionen zusammengesetzt. Welche das sind ist noch nicht abschließend geklärt, bekannt sind Bereiche für das Farbsehen, das Bewegungs- und das Tiefensehen. Beim Mensch ist etwa ein Drittel der Cortex mit dem Sehen beschäftigt, ein viel größerer Teil, als beispielsweise für das Lösen mathematischer Operationen. Das menschliche Gehirn ist folglich auf die visuelle Analyse der Umwelt spezialisiert (Fahle, 2005, S.27ff.; Goldstein, 2008, S.24, S.88).

### *Funktionsprinzipien des Gehirns: Ordnung*

Sensorische und motorische Signale werden in der jeweils entgegengesetzten Gehirnhälfte verarbeitet, d.h. Informationen der linken Körperhälfte werden in der rechten Hemisphäre verarbeitet und umgekehrt. Die genauen Gründe hierfür sind indes unklar, Manfred Fahle mutmaßt, es könnte eine Drehung des Kopfes um 180° während der Evolution stattgefunden haben (Fahle, 2005, S.68). Prinzipiell ist es so, dass benachbarte sensorische Areale auch an benachbarten

kortikalen Orten verarbeitet werden. Bei der visuellen Wahrnehmung lässt sich in Versuchen zeigen, dass auf der Hirnrinde des Okzipitallappens eine Art Projektion der Netzhautbilder entsteht. Diese ist allerdings räumlich verzerrt, was daran liegt, dass Fovea und Peripherie unterschiedlich gewichtet sind. Während die Sehgrube auf der Retina etwa 0,01% der Fläche einnimmt, sind es kortikal etwa 8-10%.

Diese Gewichtung ist Ausdruck der Dichteverteilung der Zapfen auf der Netzhaut, welche im Sehzentrum eine vielfach höhere Dichte haben, als in der Peripherie. Da die Zapfen für das Hell-Dunkel-Sehen zuständig sind, sind sie von zentraler Bedeutung für die Identifikation von Kontrasten und Konturen, was wiederum entscheidend für die Objektidentifikation ist (Fahle, 2005, S.68f.; Goldstein, 2008, S.76f.).

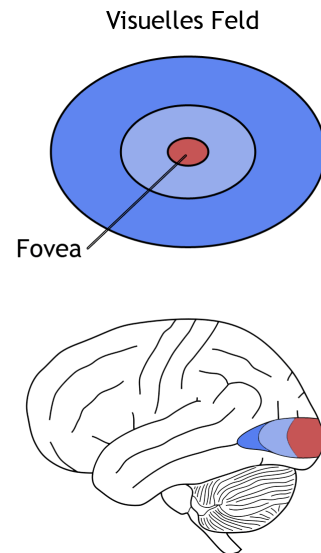


Abb. 3.10: Fläche der Fovea im visuellen Feld und im visuellen Cortex (nicht maßstabsgetreu!)

### ***Funktionsprinzipien des Gehirns: Parallelverarbeitung***

Das Prinzip der Parallelverarbeitung ist eng verbunden mit dem der Spezialisierung, denn die modulare Organisation erlaubt das zeitgleiche Verarbeiten unterschiedlicher Informationen. Dadurch erreicht das Gehirn eine enorme Beschleunigung der Prozesse, denn seine serielle Verarbeitungsgeschwindigkeit ist mit 10-100 Rechenoperationen pro Sekunde, gemessen an den Aufgaben, die es zu erfüllen hat, verhältnismäßig gering. Die Parallelverarbeitung beschleunigt die Verarbeitung einerseits, birgt andererseits das Problem, dass die modularen Einzelinformationen wieder zu einem Gesamterleben zusammengesetzt werden müssen (Bindungsproblem). Versuche zeigen, dass es sich hierbei in der Tat um ein reales Problem handelt und die Zusammenführung der Information nicht immer perfekt klappt. Werden Versuchspersonen Buchstaben in unterschiedlichen Farben präsentiert und deren Aufmerksamkeit durch eine andere kognitive Aufgabe abgelenkt, so werden anschließend die Buchstaben und Farben größtenteils richtig genannt, es kommt

allerdings häufig zu falschen Zuordnungen von Buchstaben und Farben (welcher Buchstabe hat welche Farbe?) (Fahle, 2005, S.69ff.). Eine mögliche Hypothese, wie dieses Problem im Gehirn gelöst wird, ist die Merkmalsintegrationstheorie, bei der die Aufmerksamkeit eine zentrale Rolle spielt. Sie besagt, dass zuerst die Merkmale eines Objekts (Farbe, Orientierung, Position etc.) getrennt verarbeitet werden. Die Konzentration von Aufmerksamkeit auf das Objekt setzt die Stufe der aufmerksamkeitsgerichteten Verarbeitung in Gang, welche die Objektmerkmale zu einer kohärenten Wahrnehmung kombiniert. Örtliche Korrelation spielt dabei eine große Rolle (Goldstein, 2008, S.145).

### *Funktionsprinzipien des Gehirns: Laterale und vertikale Integration*

Wie beschrieben arbeiten die Module und Areale des Gehirns nicht isoliert voneinander, sondern stehen im gegenseitigen Kontakt und Einfluss. Der Prozess der Kommunikation zwischen informationsverarbeitenden Systemen wird als Integration bezeichnet. Laterale Integration ist diejenige zwischen Nervenzellen innerhalb derselben kognitiven Verarbeitungsstufe, vertikale Integration (auch Rückkopplung) bezeichnet die Einwirkung höherer (oder späterer) Stufen auf jeweils vorangegangene. Beide Formen des Zusammenspiels sind wichtig, die Bedeutung der Rückkopplung wurde indes lange unterschätzt. In vielen Teilen des Gehirns ist die Zahl der rekurrenten Fasern mindestens genauso groß, wie die derer, die von niedrigen hin zu höheren Ebenen projizieren, wodurch eine hohe Flexibilität und Plastizität in der Verarbeitung erreicht wird. Laterale Integration besteht aus zwei Komponenten: Die erste ist die gegenseitige Unterstützung und Aktivierung meist in kurzen Reichweiten, die zweite ist die laterale Hemmung bzw. Inhibition. Letztere lässt sich gut an einigen Wahrnehmungstäuschungen des visuellen Systems verdeutlichen, wie der am Hermann-Gitter, den sogenannten Mach'schen Bändern sowie dem Simultankontrast. Alle haben mit der Helligkeitswahrnehmung zu tun und sind durch laterale Inhibition innerhalb der rezeptiven Felder der Retina erklärbar.

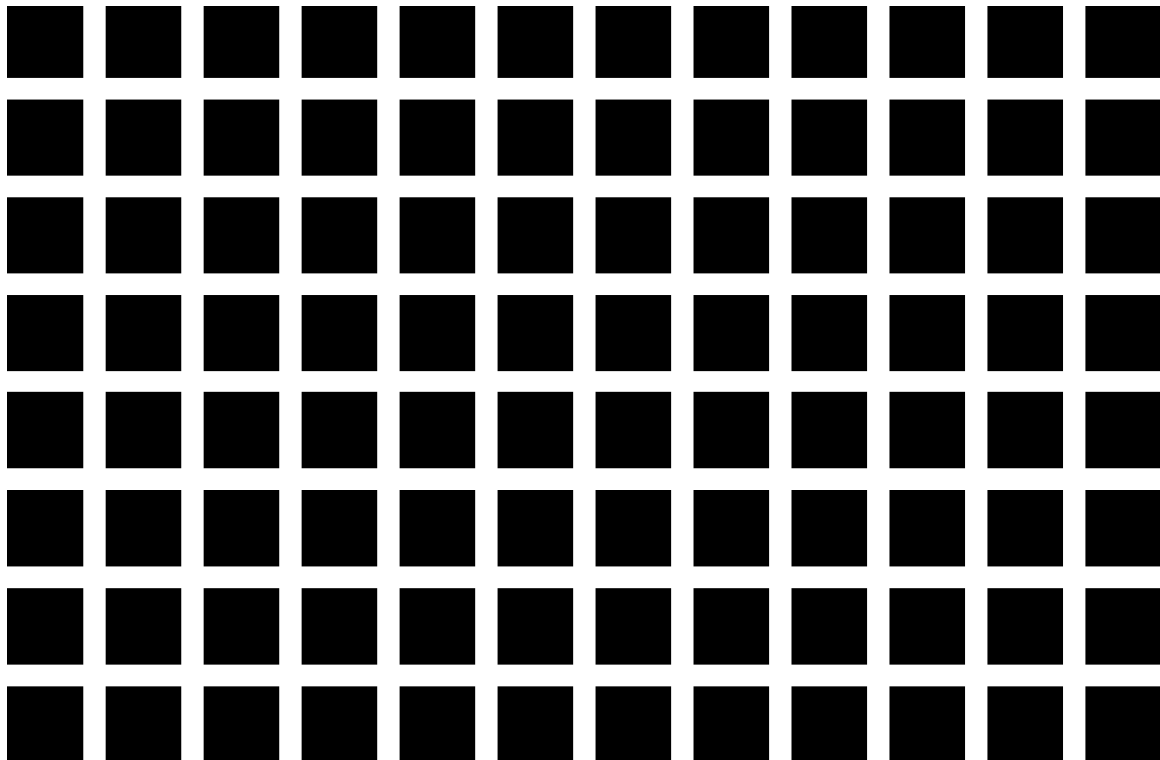


Abb. 3.11: Hermann-Gitter

Bei der Betrachtung des Hermann-Gitters fallen die grauen Punkte an den Kreuzungspunkten der weißen Korridore auf, die allerdings verschwinden, wenn der Blick genau auf diese Kreuzungspunkte gelegt wird. Möglich ist eine hemmende Wirkung der Peripherie eines rezeptiven Feldes auf sein Zentrum, die an den Kreuzungspunkten stärker ist, als in den Straßen des Gitters.

Mach'sche Bänder entstehen an fließenden Übergängen von Hell nach Dunkel, wie z.B. weichen Schatten. Das visuelle System schafft durch die laterale Inhibition eine Überhöhung des Helligkeits-Kontrasts an Kanten, dadurch konstruiert es an weichen Übergängen ein zu helles Band an der hellen Seite und ein zu dunkles an der dunklen Seite der weichen Kante.



Abb. 3.12: Simultankontrast

In der *Abb. 3.12* zum Simultankontrast erscheinen die inneren Quadrate unterschiedlich hell, obwohl sie dieselbe Menge Licht emittierten (Fahle, 2005, S.71f. /Goldstein, 2008, S.57).

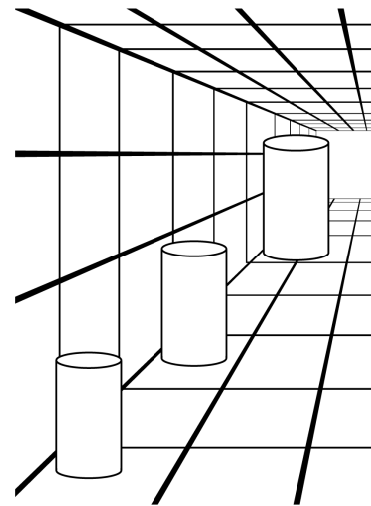
### *Wahrnehmung als Konstruktion: Informationsreduktion und partielle Synthese*

Wahrnehmungstäuschungen sind erlebbare Beispiele dafür, dass unsere Wahrnehmung die Außenwelt nicht so abbildet, wie sie ist. Informationen werden im Sinne einer schnelleren, effizienteren und zielgerichteten Verarbeitung modifiziert. Im Fall des Bildes auf der Netzhaut geschieht dies durch die Reduktion der Informationsmenge: Konturen werden extrahiert, Kontraste überhöht und gleichmäßige Flächen dabei eher vernachlässigt. Das ist für das Ziel der visuellen Wahrnehmung sinnvoll, denn Konturen entsprechen meist den Außengrenzen von Objekten. Für das Überleben ist eine möglichst schnelle Identifikation von Objekten unerlässlich, die Erfassung von kleinen Details hat dagegen weniger Bedeutung und ist daher zweitrangig. Es entsteht auf frühen Stufen als kortikales Abbild eine Art Cartoon (was auch der Grund dafür sein könnte, warum Cartoons so erfolgreich sind!). Aus diesen reduzierten Informationen wird dann unter Zuhilfenahme von Erfahrungen (Prinzipien und Vorannahmen) eine möglichst konsistente Repräsentation der Außenwelt konstruiert. Synthese bzw. Konstruktion ist demnach neben der Reduktion von Information ein wesentliches Prinzip der Wahrnehmung. Visuell unvollständige Konturen, die das Gehirn zu Formen zusammensetzt, werden als „kognitive Konturen“ bezeichnet. Diese treten automatisch, also unwillkürlich auf. Ihre Bildung kann daher auf frühe Stufen der Verarbeitung zurückgeführt werden und findet vermutlich im sekundären visuellen Cortex statt.

Das visuelle System reagiert besonders empfindlich auf bewegte Kanten. Diese erzeugen in den betroffenen rezeptiven Feldern der Netzhaut eine schnelle Änderung der Reizintensität. Versuche zeigen, dass bewegte „kognitive Konturen“ ähnlich starke neuronale Aktivitätsmuster hervorrufen, wie Vergleichstest mit realen Konturen. Dies ist ein deutliches Indiz dafür, dass diese illusorischen Konturen auf frühen Ebenen gebildet werden (Fahle, 2005, S.73ff.).

### *Bottom-Up und Top-Down: Die Rolle des Wissens in der Wahrnehmung*

Die Repräsentation der Außenwelt ist zu einem Teil von den Signalen der Sinnesorgane und zum anderen von der Erwartungshaltung des Gehirns abhängig. Das in die Situation eingebrachte Wissen kann dabei vor langer Zeit, oder aber erst kurz zuvor erlernt sein. Die Einbeziehung von Wissen in den Wahrnehmungsprozess wird als Top-Down-Verarbeitung, oder auch wissensbasierte Verarbeitung bezeichnet. Diese ermöglicht z.B. durch die perspektivische Interpretation einer Szene, dass real gleich große Objekte in unterschiedlichen Entfernungen gleich groß erscheinen, obwohl ihre optische Abbildung auf der Netzhaut kleiner ist (Größenkonstanz). *Abb. 3.13* zeigt eine typische optische Täuschung, die durch Top-Down-Verarbeitung der Perspektive hervorgerufen wird.



*Abb. 3.13: Größenkonstanz*

Wahrnehmungstäuschungen entstehen meist aus Diskrepanzen zwischen subjektiv konstruierter Repräsentation und einbezogenem Wissen. Bottom-Up-Verarbeitung (oder reizgesteuerte Verarbeitung) ist demgegenüber die Verarbeitung, die auf eingehenden, sensorischen Informationen beruht (Fahle, 2005, S. 77ff.; Goldstein, 2008, S.8).

### *Gehirnaktivität und Wahrnehmung*

Die Aktivität des Cortex und seiner sensorischen Areale hängt nur zu einem Teil von äußeren Sinneseindrücken ab, der andere Teil ist sogenannte intrinsische Aktivität. Es gibt Beobachtungen an Patienten mit Netzhautdefekten, die starke visuelle Halluzinationen haben. Diesen Patienten ist bewusst, dass es sich um Illusionen handelt, anders als beispielsweise bei Schizophrenen, bei denen Betroffene die Realität nicht von ihren Halluzinationen unterscheiden können. Es ist davon auszugehen, dass die durch Schädigungen der Retina verursachten Halluzinationen vom visuellen Cortex anstelle der verlorengegangenen Eindrücke generiert werden. Ferner gibt es Experimente, in denen Menschen in einen Zustand ohne sensorische Reize versetzt wurden, d.h. in absolute

Dunkelheit, Stille und in körperwarmes Wasser, um auch taktile Reize zu minimieren. Testpersonen begannen bei diesen Versuchen innerhalb weniger Minuten stark zu halluzinieren. Das Gehirn scheint in solchen Fällen verlorengegangene sensorische Reizung zu ersetzen, indem es eigene aus abgespeichertem Wissen erzeugt. Auch aus anderen Versuchen ist bekannt, dass ein großer Teil der Gehirnaktivität intrinsische ist und nur zu einem geringen Teil von äußerer Reizung abhängt. Übertragen auf das Sehen bedeutet dies, dass die Aktivität im visuellen Cortex durch die Netzhautbilder nur moduliert wird (Fahle, 2005, S.87ff.).

### *Wahrnehmung und Aufmerksamkeit*

Um tatsächlich bewusst wahrnehmen zu können, bedarf es der Aufmerksamkeit. Sie ist der erste große Filter, da nur ein Bruchteil der einströmenden Sinnesreize tatsächlich verarbeitet werden kann. Es lässt sich an vielen Beispielen deutlich machen, dass große Teile der äußeren Eindrücke das Bewusstsein nicht erreichen. Eindrucksvoll sind in diesem Zusammenhang Versuche von Simons und Chabris, die den Begriff Unaufmerksamkeitsblindheit prägten. Testpersonen wurde ein Film eines Basketballspiels präsentiert und sie erhielten dabei die Aufgabe, die Ballwechsel im Kopf mitzuzählen. Rund die Hälfte der Probanden bemerkte dabei nicht den quer durchs Bild laufenden Gorilla (Person im vollständigen Gorilla-Kostüm) (Goldstein, 2008, S.137).

Trotzdem spielt auch die unbewusste Wahrnehmung eine Rolle. So konnte nachgewiesen werden, dass Ereignisse, die der bewussten Wahrnehmung entgehen, darauffolgende Erlebnisse beeinflussen können. In Versuchen mit kurz dargebotenen Hinweisreizen (Pfeile) und einer unmittelbar folgenden Maske misslang es den Testpersonen, die Richtung der Hinweisreize zu identifizieren. Dennoch beeinflussten sie die Reaktionszeiten bei der Beurteilung des darauffolgenden Zielreizes der Versuchspersonen massiv (Fahle, 2005, S.91ff.).

### *Stufen der Verarbeitung*

Die Verarbeitung sensorischer Reize läuft auf verschiedenen kognitiven Stufen ab. Dies ist zunächst einmal ein abstraktes Bild von einer hierarchischen Informationsverarbeitung des Gehirns. Anhand von Experimenten zu Wahrnehmung und kognitiven Leistungen, sowie durch Personen mit partiellen

Schädigungen des Cortexes ist erkennbar, dass es derartige Stufen geben muss. Prinzipiell ist es dabei so, dass die ‚Höhe‘ der Stufe die Komplexität der zu lösenden Aufgabe widerspiegelt. Dies sei am Beispiel des Drei-Stufen-Modells verdeutlicht: Die erste Stufe ist die sensorische. Hier wird physikalische Energie bemerkt und in neuronale Energie und sensorische Erfahrungen umgewandelt. Die zweite Stufe ist die Wahrnehmungsorganisation. Auf dieser Ebene verbinden perzeptuelle Prozesse die sensorischen Erfahrungen zu einem kohärenten Ganzen und liefern die Wahrnehmung von Objekten und Mustern. Die dritte Stufe ist die Identifikation und Wiedererkennung. Hier werden die Perzepte der Objekte und Muster mit Erfahrungen und Gedächtnisrepräsentationen verglichen, um als vertraute und bedeutungsvolle Objekte erkannt zu werden. (Zimbardo, 2008, S.157).

Dieses Forschungsgebiet unterliegt aufgrund des steten Wissenszuwachses einem ständigen Wandel. Es gibt eine Vielzahl von Modellen, die sich teils überschneiden, divergieren oder aber auch noch weiter differenzieren. Für das grundlegende Verständnis soll das vorgestellte Modell an dieser Stelle reichen.

### *Zusammenfassung*

Wahrnehmung ist ein hoch komplexer Vorgang, der große Anforderungen sowohl an Rezeptoren als auch an das Zentralnervensystem stellt. Sie ist der Vermittler zwischen der Außenwelt und dem Ich und ist somit elementar für das Überleben und Agieren in der Umwelt. Wahrnehmung ist dabei stets ein aktiver Prozess, an dem viele kognitive Prozesse beteiligt sind und funktioniert stets unter Zuhilfenahme von Wissen und Erfahrungen. Daher ist die ‚Sicht auf die Welt‘ immer subjektiv gefärbt. In diesem Kapitel wurden die Eigenschaften und Einschränkungen der menschlichen akustischen und visuellen Wahrnehmung ebenso erläutert, wie die Prinzipien der kognitiven Informationsverarbeitung. Im folgenden Kapitel wird besprochen, wie aus den Eindrücken der jeweiligen Modalitäten ein einheitliches Erleben konstruiert wird und welche Form der Interaktionen und gegenseitigen Beeinflussungen bei der Integration von akustischer und visueller Wahrnehmung auftreten können.

## 4 MULTISENSORISCHE PROZESSE

In Kapitel 3 wurde dargelegt, dass Wahrnehmung ein aktiver Prozess der Synthese und Konstruktion ist, der sowohl auf sensorischem Erleben, als auch auf Erfahrungen und Wissen basiert. Das Gehirn verarbeitet dabei die eintreffenden Informationen parallel, mittels auf bestimmte Aufgaben spezialisierter Areale, wodurch ein rascher Wahrnehmungsprozess erreicht wird. Um die vielen modularen Analyseergebnisse zu konsistenten Objekten zu machen und zu einem Gesamterleben zu integrieren, bedarf es intensiver Kommunikation aller Areale. Demzufolge kommt es zu Interaktion und gegenseitiger Beeinflussung, auch zwischen den Sinnesmodalitäten. Dieses Kapitel untersucht Interaktionen zwischen der akustischen und visuellen Wahrnehmung. Im Mittelpunkt stehen dabei solche, die für die Analyse von audiovisuellen Medien relevant sind. Multimodale Sprachwahrnehmung wird dabei weitestgehend ausgeklammert.

### 4.1 Forschungsgebiete und Untersuchungsmethoden

Vorgestellt werden Untersuchungen aus dem Bereich der Kognitionswissenschaften, einem relativ jungen Wissenschaftszweig, dessen Ziel es ist, kognitive Fähigkeiten wie die Wahrnehmung, das Denken, das Lernen, die Motorik und die Sprache zu erforschen. Die Wahrnehmung wurde lange Zeit als modulare Funktion angesehen, in der die Sinne getrennt und unabhängig voneinander funktionieren. Erste Untersuchungen multisensorischer Prozesse fanden zwar bereits Anfang des 20. Jahrhunderts statt, ein wirklich breites Interesse und damit systematische Erforschung erfuhren diese allerdings erst ab den 1970er Jahren (Shams/Kamitani/Shimojo, 2004).

Viele Untersuchungen auf dem Gebiet der multisensorischen Wahrnehmung sind empirische Tests, in denen Probanden Reize in einer oder mehreren Modalitäten präsentiert werden. Entsprechend einer bestimmten Fragestellung werden Reaktion, Verhalten oder Interpretation der Reize gemessen und damit die Wahrnehmung analysiert. Durch statistische Erhebungen können dann allgemeingültige Rückschlüsse über Wahrnehmungsprozesse gezogen werden. In der jüngeren Vergangenheit kamen bildgebende Verfahren wie die

Computertomografie (CT), Magnetresonanztomografie (MRT) sowie die sogenannten funktionellen bildgebenden Verfahren, wie z.B. die Positronen-Emissionstomografie (PET) als Untersuchungsmittel hinzu. Diese Methoden ermöglichen das Messen und Lokalisieren von Hirnaktivität. Dadurch lässt sich bestimmen, welche Areale bei welchen kognitiven Aufgaben wie stark aktiv sind.

Die meisten Untersuchungen konzentrieren sich jeweils auf einen kleinen Teilaspekt von Wahrnehmung oder Kognition. Erkenntnisse in einem größeren Rahmen sind somit das Ergebnis unterschiedlichster Forschungen, die in ihrer Gesamtheit betrachtet Strukturen und Funktionsweisen von Wahrnehmung erkennen lassen (Calvert/Spence/Stein, 2004).

#### 4.1.1 Kreuzmodale Prozesse bei der exogenen Lenkung der Aufmerksamkeit

Aufmerksamkeit ist Grundvoraussetzung für bewusstes Wahrnehmen. Sie kann durch einen Sinnesreiz gelenkt werden, was zur Folge hat, dass sich sowohl die gereizte Modalität, als auch alle anderen Sinne dem jeweiligen Bereich zuwenden und die Sensitivität dort verstärkt ist. In einer Vielzahl von alltäglichen Situationen ist dies der Fall, sei es eine Mücke, die sich auf den Arm setzt, oder ein hupendes Auto. Innerhalb kürzester Zeit sind das Sehen, Hören und Fühlen in erhöhter Alarmbereitschaft und auf den Ort des Geschehens fokussiert. Dies sind beispielhafte Situationen, in denen die Reizung einer Modalität Auswirkungen auf andere Sinne hat.

Bei Betrachtung der Funktionsweise des menschlichen Auges ist es leicht ersichtlich, dass Aufmerksamkeit unerlässlich ist. Ein Objekt muss von den Augen fixiert werden, um richtig gesehen und erkannt zu werden, da der Bereich des scharfen Sehens (Fovea) beim Menschen wenige Grad klein ist. Erstaunlich ist hingegen, dass sich das Hören und Fühlen in einem Punkt intensiviert, wenn dieser visuell betrachtet wird. Dies ist bereits der Fall, sobald die bloße Absicht besteht, dorthin zu schauen (Spence/McDonald, 2004).

Demzufolge gibt es zwei Arten der Aufmerksamkeitsverlagerung: Die offene, von außen durch Bewegung der Augen, des Kopfes oder des ganzen Körpers sichtbare, sowie die verdeckte, die sich nicht durch äußere Anzeichen oder Veränderungen feststellen lässt. Es gilt, dass einer offenen Verlagerung unter

normalen Umständen stets eine verdeckte vorausgeht. Andersherum muss jedoch nicht jede verdeckte eine offene Verlagerung zur Folge haben.

Bereits frühe Studien wie z.B. von Jonides (1981) oder Posner und Cohan (1984) stellen fest, dass räumlich nicht vorhersagbare visuelle oder akustische Hinweisreize für einen kurzen Zeitraum Reaktionen auf Zielreize in gleicher Richtung und Modalität (intramodal) fördern, sowohl mit, als auch ohne äußere Verlagerungen. Da die Hinweisreize in diesen und ähnlichen Studien nicht vorhersagbar waren, schlussfolgerten die Wissenschaftler, dass die Aufmerksamkeit unfreiwillig, oder zumindest auf eine gewisse Art automatisch gelenkt wird. Dieser Argumentation folgend bezeichnen Spence und McDonald sie als „*exogenous attention orienting*“ und unterscheiden diese von der „*endogenous attentional orienting*“, welche auf die Präsentation eines räumlich vorhersagbaren Hinweisreizes folgen (Spence/McDonald, 2004).

Eine Vielzahl von Studien gingen seit Mitte der 1990er Jahre der darauf folgenden Frage nach, ob verdeckte als auch offene Aufmerksamkeitsverlagerungen auch zwischen den Sinnen, also intermodal auftreten. Nach vielen zunächst widersprüchlichen Ergebnissen verschiedener Arbeitsgruppen ist mittlerweile der empirische Nachweis dafür erbracht, dass ein nicht vorhersagbarer akustischer, visueller, oder taktile Hinweisreiz, die Wahrnehmung eines Zielreizes, unabhängig in welcher Modalität, am selben Ort fördert. Experimentell konnten intermodale Aufmerksamkeitslenkungen zwischen den Sinnen Sehen, Hören und Fühlen in allen neun Kombinationen erzeugt werden. Dabei war die Art der Verlagerung von den jeweiligen Eigenschaften und Fähigkeiten der Sinnesorgane abhängig: So ist z.B. die räumliche Orientierung infolge visueller Hinweisreize exakter, als bei akustischen. Auffällig war in allen Versuchen, dass der Effekt der erhöhten Aufmerksamkeit auf einen Zielreiz abnimmt, je weiter dieser räumlich vom Hinweisreiz entfernt wird. Darüber hinaus konnte durch bildgebende Verfahren gezeigt werden, dass sich durch einen Hinweisreiz die Hirnaktivität in Bereichen erhöht, die der Modalität des Zielreizes zugeordnet werden können. Sinnesreize erzeugen demnach für einen gewissen kurzen Zeitraum Aufmerksamkeit in allen Modalitäten. Dies betrifft das gesamte Wahrnehmungssystem, äußere Sinnesorgane, als auch kognitive Stufen der Verarbeitung. Die Studien zeigten auch einheitlich, dass Top-Down-Verarbeitung Einfluss auf die Stärke jeder

Aufmerksamkeits-Verlagerung haben kann. Diese Erkenntnis ist zwar wenig überraschend, dennoch ist der experimentelle Nachweis relativ jung. Die zu Grunde liegenden Mechanismen werden indes noch diskutiert. Die von Martha Farah (1989) vorgeschlagene Hypothese eines supramodalen Aufmerksamkeits-Systems scheint dabei wieder an Popularität zu gewinnen, da es mittlerweile gelungen ist, kreuzmodale Verlagerungen zwischen den genannten drei Sinnen nachzuweisen (Spence/McDonald, 2004).

Zusammenfassend lässt sich festhalten: Die menschliche Wahrnehmung ist darauf spezialisiert, die Aufmerksamkeit auf einen Ort zu fokussieren. Reizungen einer Modalität können zu einer erhöhten Sensitivität anderer Sinne in einem bestimmten Punkt führen. Diese kann verdeckt erfolgen, ohne äußerlich sichtbare Veränderungen, oder offen, z.B. durch Hinwenden zu einem Ereignis. Dabei kann es durch die Reizung einer Modalität zu Aktivierungen von Hirnarealen anderer Modalitäten kommen. Jede Verlagerung hat eine Intensivierung der Wahrnehmung in diesem Bereich zur Folge, jedoch nur für einen gewissen kurzen Zeitraum.

#### 4.1.2 Modulation der visuellen Wahrnehmung durch Klang

Der Einfluss anderer Modalitäten auf das Sehen wird erst seit einiger Zeit systematisch erforscht, da wie eingangs erläutert lange Zeit das Sehen als die dominante Modalität galt. Im Folgenden werden einige Studien und Ergebnisse zum Einfluss des Hörens auf das Visuelle vorgestellt, was von besonderem Interesse für die Betrachtung der Wahrnehmung audiovisueller Medien ist.

##### *Der Einfluss akustischer Reize auf temporäre Aspekte des Sehens*

Akustische Reize können die zeitliche Dimension der visuellen Wahrnehmung deutlich beeinflussen. So ist es möglich, die zeitliche Auflösung des Sehens durch akustische Reize zu verbessern oder auch herabzusetzen, wie eine Untersuchung von Scheier, Nijwahan und Shimojo (1999) zeigten. Ihre Versuchsanordnung bestand aus zwei Lampen auf einer horizontalen Achse, die nacheinander eingeschaltet wurden. Die Verzögerungszeit  $t$  zwischen den Einschaltvorgängen war variant von 60ms bis -60ms. Testpersonen sollten nun

die Reihenfolge bestimmen, in denen die Lampen eingeschaltet wurden. Simultan zu deren Einschaltvorgang wurden Piep-Töne dargeboten. Im Folgenden wird der akustische Reiz A und der visuelle Reiz V genannt. Das Ergebnis war: Testpersonen beurteilten die Einschaltreihenfolge der Lampen im Durchschnitt besser, wenn die Reize in der Reihenfolge A-V-V-A präsentiert wurden, verglichen mit Versuchen ohne Töne. Dagegen verschlechterte sich die Beurteilungsgenauigkeit, wenn die Töne zwischen den Einschaltvorgängen der Lampen lagen (V-A-A-V), im Vergleich zu Versuchen ohne Töne. Ähnliches fanden Fendrich und Corballis (2001) in ihrem Versuch heraus, bei denen Testpersonen einen Blitz früher wahrnahmen, wenn diesem ein Ton vorausgeht und später, wenn der Ton dem Blitz folgt, verglichen mit der synchronen Darbietung beider Reize.

Offenbar ist es so, dass akustische Reize die visuelle Wahrnehmung auf zeitlicher Ebene unterstützt. Dies deckt sich mit der „*modality appropriateness hypothesis*“, die besagt, dass diejenige Modalität, die im Kontext einer bestimmten Aufgabenstellung verlässlicher und sicherer ist, die Führung über die Wahrnehmung bei der Lösung dieser Aufgabe übernimmt (Choe et al., 1975; Fisher, 1968; Howard/Templeton, 1966; Kaufmann, 1974; Welch/Warren, 1986). So hat das visuelle System ein höheres räumliches Auflösungsvermögen als das Hören und dominiert somit die Wahrnehmung in entsprechenden Situationen, wohingegen die akustische Wahrnehmung ein höheres zeitliches Auflösungsvermögen besitzt und somit bei dieser Art der Aufgabenstellung leitet (Shams/Kamitani/Shimojo, 2004).

#### *Der Einfluss akustischer Reize auf andere Aspekte des Sehens*

Der Einfluss der akustischen auf die visuelle Wahrnehmung beschränkt sich jedoch nicht nur auf zeitliche Aspekte. Akustische perzeptuelle Organisation kann die visuelle perzeptuelle Organisation verändern. Ein plötzlicher Klang kann die Identifikation eines an selber Stelle folgenden Blitzes erhöhen. Die erlebte Intensität eines visuellen Stimulus kann durch die Anwesenheit von Klang erhöht werden. Ein abrupter Klang kann die Identifikation eines synchronen visuellen Ziel-Reizes innerhalb dargebotener Stör-Reize verbessern. In einer Studie von Sekuler und Lau (1997) ergab sich eine veränderte Interpretation eines mehrdeutigen visuellen Bewegungsereignisses durch die

Anwesenheit von Klang. Beim Anblick zweier identischer Objekte, die sich aufeinander zu bewegen und in einem Punkt durchlaufen, empfanden die meisten Testpersonen diese Objekte als sich unterlaufend. Abb.4.1 zeigt die Phasen eines möglichen Versuchsaufbaus.

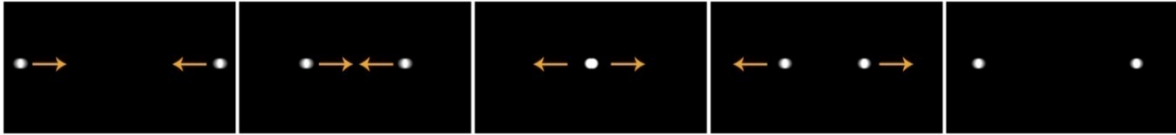


Abb.4.1: Phasen eines Displays, mit verschieden interpretierbarem visuellen Stimulus (weiße Kreise).

Wurde zu demselben visuellen Stimulus zum Zeitpunkt des Aufeinandertreffens beider Objekte ein kurzer Ton präsentiert, interpretierte die Mehrzahl der Befragten dies als Kollision der Objekte. Die Beschaffenheit des Geräuschs hatte dabei gewichtigen Einfluss. So stellte sich dieser Effekt nur dann ein, wenn das Geräusch etwas „Scharfes“, Plötzliches hatte. Bemerkenswert ist auch, dass die gleiche Veränderung der Interpretation nicht nur durch Töne hervorgerufen wurde, sondern ebenso durch taktile Reize oder einen kurzen Blitz erzeugt werden kann (Shams, Kamitani/Shimojo, 2004).

#### Manipulation der visuellen Struktur durch das Hören

Der Einfluss des Hörens auf die visuelle Wahrnehmung kann über den modulierenden Charakter hinausgehen: Akustische Reize können die visuelle Wahrnehmung grundlegend in ihrer Struktur verändern. Dies zeigt sich beim „illusory flash effect“ (Shams/Kamitani/Shimojo, 2000 & 2002), bei dem die Wahrnehmung eines eindeutigen visuellen Stimulus durch akustische Reize manipuliert wird. In einer Versuchsanordnung wurde Testpersonen ein einzelner Lichtblitz zusammen mit mehreren kurzen Piep-Tönen präsentiert, was dazu führte, dass diese mehrere Lichtblitze wahrnahmen. Hierbei handelt es sich tatsächlich um ein perzeptuelles

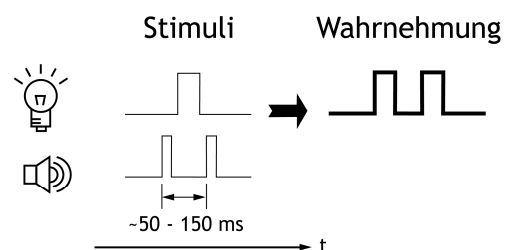


Abb. 4.2: „Illusory flash effect“. Schematische Darstellung der dargebotenen Reize und der resultierenden Wahrnehmung

Phänomen, denn der Effekt hielt auch unter verschiedensten Kontrollbedingungen an und sogar dann, wenn die Testpersonen den Versuchsaufbau kannten. Ferner unterschied sich der illusorische zweite Blitz für die Testpersonen kaum von dem tatsächlich vorhandenen. Die zeitliche Grenze, bis zu dem dieser Effekt auftritt, liegt bei etwa 100ms Abstand zwischen den Piep-Tönen. Bei größeren Verzögerungszeiten nimmt er stark ab. Interessant ist auch, dass der Effekt nicht umkehrbar ist, das heißt zwei oder mehr visuelle Blitze werden nicht als einer wahrgenommen, wenn sie mit einem einzelnen Piep-Ton dargeboten werden.

Ähnliche Mechanismen könnten bei einer früheren Studie von Saldaña und Rosenblum von 1993 zu Grunde gelegen haben. Hier wurde ein kontinuierlich gestrichener Cello-Ton durch das gleichzeitig präsentierte Bild eines gezupften Cellos moduliert. Andersherum wurde jedoch der Ton eines gezupften durch das Anschauen eines gestrichenen Cellos nicht moduliert.

Kamitani und Shimojo (2001) erweiterten das „illusory-flash“ Experiment. Sie boten in ihrer Versuchsanordnung zwei räumlich und zeitlich versetzte visuelle Reize dar und erzeugten somit einen sogenannten „*apparent motion stream*“, eine Art simulierte Bewegung zwischen zwei Punkten. Zu jedem dieser visuellen Ereignisse ertönte synchron ein Piep-Ton. Einen dritten Piep-Ton schalteten sie zeitlich zwischen die beiden synchronisierten audiovisuellen Reize. Dieser moduliert die visuelle Wahrnehmung in der Form, dass ein dritter Blitz räumlich zwischen den tatsächlich vorhandenen wahrgenommen wurde, also auf dem virtuellen Bewegungspfad zwischen den beiden tatsächlich vorhandenen visuellen Reizen. Damit wiesen sie nach, dass beim „*illusory flash effect*“ tatsächlich ein eigenständiges visuelles Ereignis induziert wird (Shams, Kamitani/Shimojo, 2004).

### ***Klang verändert die visuelle Bewegungswahrnehmung***

In der Vergangenheit beschäftigten sich viele Versuche vor allem mit dem Einfluss von visuell dargebotener Bewegung auf die Bewegungswahrnehmung im Klang. Sie alle kamen zu dem Ergebnis, dass die Bewegung eines akustischen Reizes besser wahrgenommen wird, wenn sie durch einen visuellen Reiz in der gleichen Richtung begleitet wird. Jüngere Studien widmeten sich dann der Frage, inwieweit bewegte akustische Reize Einfluss auf die visuelle

Bewegungswahrnehmung haben. Meyer und Wuerger (2001) installierten dazu ein Display, auf dem sich viele Punkte willkürlich bewegten. Die Kohärenz zwischen den Bewegungen der einzelnen Punkte war steuerbar, so dass ein mehr oder weniger starker Eindruck einer globalen Bewegungsrichtung erzeugt werden konnte. Simultan zu diesem visuellen Reiz wurden akustische „*apparent motion streams*“ in horizontaler Richtung präsentiert. Meyer und Wuerger stellten fest, dass durch die so erzeugte akustische Bewegung, visuelle nicht-kohärente Reize als bewegt wahrgenommen wurden, und zwar in Richtung der jeweiligen akustischen Bewegung. In einer anderen Studie verwendeten Shams, Allman und Shimojo (2001) einen ähnlichen Versuchsaufbau wie Meyer und Wuerger, jedoch nicht mit einem mehrdeutigen visuellen Stimulus, sondern einem kurzen, stationären Reiz (Blitz). Dabei fanden sie heraus, dass ein bewegter akustischer Reiz eine Bewegung in dem stationären visuellen Reiz hervorruft. Ein bewegter Blitz hatte in diesen Versuchen hingegen keinen Einfluss auf die Wahrnehmung eines stationären akustischen Stimulus. Diese Ergebnisse sind wichtig, da die Richtung dieser kreuzmodalen Interaktionen nicht eindeutig mit der „*modality appropriateness hypothesis*“ vorhersagbar ist (Shams/Kamitani/Shimojo, 2004).

### *Neuronale Mechanismen*

Shams, Kamitani und Shimojo (2001) führten *Visual-Evoked Potential* (VEP) Messungen mit dem Versuchsaufbau des „*illusory flash effect*“ durch, also Messungen von Hirnströmen, die durch visuelle Reize ausgelöst werden. Damit wollten sie der Frage nachgehen, auf welchen Ebenen der kognitiven Verarbeitung die Beeinflussungen des Visuellen durch das Akustische stattfinden. Die Forscher hatten zuvor beobachtet, dass „*illusory-flash effect*“ in der Peripherie des Sichtfeldes wesentlich stärker ist, als in der Fovea, weshalb sie Messungen für beide Fälle durchführten. Sie fanden heraus, dass beim zweiten illusorischen Blitz der visuelle Cortex aktiviert wird und die Potential-Kurven, die ein Illusions-Blitz hervorrief, denen eines real vorhandenen zweiten Blitzes stark ähnelten. Außerdem waren die Hirnströme stärker, je deutlicher die visuelle Illusion war, was sie durch Modifikationen im Versuchsaufbau steuern konnten. Diese Beobachtungen legen den Schluss nahe,

dass es sich beim „*illusory flash effect*“ um ein perzeptuelles Phänomen handelt und nicht durch kulturell erlernte oder semantische Relationen verursacht wird. Es gibt eine Menge anderer Studien, die ähnliches beobachteten: So wurde z.B. Aktivität im auditiven Cortex durch bloßes stummes Lippenlesen nachgewiesen (Calvert, 1997), oder durch Berührung hervorgerufene Aktivierung des visuellen Cortexes (Sathian et al.; 1997). Untersuchungen an Affen ergaben darüber hinaus starke Projektionen vom primären auditiven Cortex sowie von multimodalen Regionen zu peripheren Regionen des visuellen Cortexes. Das deckt sich mit den Beobachtungen, dass der „*illusory flash effect*“ wesentlich stärker in Randbereichen des Sichtfeldes auftritt (Shams/Kamitani/Shimojo, 2004).

Zusammenfassend lässt sich sagen, dass akustische Reize in vielerlei Hinsicht Einfluss auf die visuelle Wahrnehmung haben können. Es ist möglich, dass sie das Sehen auf zeitlicher Ebene unterstützen, da dort die Auflösung der akustischen Wahrnehmung höher ist, als die der visuellen. Die empfundene Intensität eines visuellen Reizes lässt sich durch Klänge erhöhen. Ferner können akustische Reize helfen, visuelle zu identifizieren, oder in mehrdeutigen Situationen eine Veränderung der Interpretation der Situation hervorrufen. Neben diesen unterstützenden oder verstärkenden Effekten sind auditive Stimuli darüber hinaus in der Lage, die Struktur der visuellen Wahrnehmung grundlegend zu verändern. Der „*illusory flash effect*“ zeigt, wie durch Klänge visuelle Wahrnehmungen auf perzeptueller Ebene hervorgerufen werden. Auch auf die visuelle Bewegungswahrnehmung haben Klänge Einfluss. So können bewegte akustische Reize visuell wahrgenommene Bewegung hervorrufen, wenn diese in ihrer Bewegung mehrdeutig oder statisch sind.

#### 4.1.3 Modulation der akustischen Wahrnehmung durch das Sehen

Das Sehen als die dominierende Modalität der menschlichen Wahrnehmung anzusehen, war und ist eine auch unter Wissenschaftlern immer noch verbreitete Ansicht. Daher gibt es viele Studien aus der Anfangszeit der multisensorischen Kognitionsforschung, die den Einfluss der visuellen Wahrnehmung auf andere Modalitäten untersuchen (Shams/Kamitani/Shimojo, 2004). Im Folgenden

werden zwei Effekte vorgestellt, die eine Modulation der akustischen Wahrnehmung durch visuelle Reize zeigen.

### *Der Bauchredner-Effekt (Ventriloquist-Effekt)*

In der Regel werden aus den vielen Eindrücken der verschiedenen Sinne konsistente Objekte zusammengesetzt. Sind die jeweiligen Informationen einer oder mehrerer Modalitäten widersprüchlich, kann das zur Synthese zweier oder mehr Objekte führen. „Die Art, wie die Informationen mehrerer Sinne zu geschlossenen Objekten zusammengeführt werden ist faszinierend.“ (Woods/Recanzone, 2004, S.35) Genaue Regeln, wie diese Integration auf neuronaler und perzeptueller Ebene abläuft, sind jedoch bisher weitgehend unbekannt.

Der Bauchredner-Effekt beschreibt die räumliche Anbindung eines akustischen Reizes an eine visuelle Quelle. Ohne diesen Effekt wäre die Illusion in Film, Fernsehen und vielen anderen audiovisuellen Medien nicht möglich. Denn er bewirkt, dass z.B. die Stimme des Sprechers aus dessen Mund kommend wahrgenommen wird und nicht etwa aus dem neben dem Bildschirm platzierten Lautsprecher. Die Faktoren, die die Stärke des Bauchredner-Effekts beeinflussen, wurden in Versuchen mit der Wiedergabe von menschlichen Stimmen in Verbindung mit einem Videobild eines sprechenden Menschen (Driver, 1996; Thurlow/Jack, 1973; Warren/Welch/McCarthy, 1981), als auch mit Puppen unterschiedlicher Detailtreue untersucht (Thurlow/Jack, 1973). Diese Studien zeigten, dass je „angemessener“ der visuelle Reiz ist, also je mehr er einer realen Situation nahekommmt, desto stärker ist der Bauchredner-Effekt. Auch der räumliche Abstand zwischen Schallquelle und Bild spielt eine Rolle. Kleine Distanzen fördern den Effekt, größere verringern ihn. Auch zeitliche Komponenten spielen eine Rolle. Thurlow und Jack (1973) fanden heraus, dass Asynchronitäten zwischen Bild und Ton von einigen 100ms den Bauchredner-Effekt verhindern (Woods/Recanzone, 2004). Im Folgenden wird eine jüngere Studie vorgestellt, die diese Zusammenhänge genauer quantifiziert.

### *Die räumliche und zeitliche Abhängigkeiten des Bauchredner-Effekts*

Slutsky und Recanzone (2001) testeten in Experimenten die räumliche und zeitliche Abhängigkeit des Bauchredner-Effekts mit dem Licht einer LED und Rauschen bzw. einem 1kHz Tons. In einem ersten Schritt sollten Testpersonen mit einem Schalter angeben, ob beide Reize gleichzeitig, oder zeitlich versetzt auftreten. Ein erstes wichtiges Ergebnis ist, dass es hierbei keine Unterschiede in den Ergebnissen zwischen Versuchen mit Rauschen und solchen mit Ton gab. Ein weiteres ist, dass bei asynchroner Darbietung von Licht und Ton bzw. Rauschen die Reihenfolge eine wichtige Rolle spielt; Wurde der akustische Reiz vor dem visuellen dargeboten, nahmen die Testpersonen dies bei kleineren zeitlichen Differenzen wahr, als wenn der akustische dem visuellen Reiz folgte. Gegenüber einer Asynchronität zwischen Bild und Ton ist die menschliche Wahrnehmung also empfindlicher, wenn der Ton dem Bild vorausgeht, als wenn er dem Bild folgt. Der Unterschied in der Reizschwelle liegt hier bei etwa 50ms, das bedeutet ein Ton-vor-Bild Versatz von 100ms wird genauso wahrgenommen wie ein Bild-vor-Ton Versatz von 150ms. Das dritte wesentliche Ergebnis ist, dass bei einem zeitlichen Versatz von 250ms (200ms bei akustisch vor visuell) die Fähigkeit die Reize getrennt wahrzunehmen bei annähernd 100% lag (Woods/Recanzone, 2004).

Ein zweites Experiment zeigte, dass es eine deutliche Interaktion zwischen räumlichen und zeitlichen Parametern gibt. Slutsky und Recanzone (2001) formulierten zunächst die Hypothese, dass die Ortbarkeit des akustischen Stimulus den Bauchredner-Effekt maßgeblich beeinflusst. Ist die Schallquelle deutlich aus einer anderen Richtung als das Bild kommend wahrnehmbar, so die Vermutung, wird die Interaktion wesentlich geringer sein, als wenn die Richtung der Schallquelle nicht gut von der des Bildes zu unterscheiden ist. Um diese Hypothese zu bestätigen, arbeiteten sie mit zwei akustischen Stimuli, einem genau vor den Testpersonen, und einem weiteren links oder rechts in verschiedenen Auslenkungen positionierten Reiz. Ein visueller Stimulus war statisch auf der azimuthalen Achse zu sehen, jedoch in zeitlich variablen Abständen zu den akustischen Reizen. Die Testpersonen wurden zunächst in unimodalen Tests gefragt, ob sie die beiden akustischen Reize aus einer oder unterschiedlichen Richtungen wahrnehmen. In anschließenden multisensorischen Tests wurden in veränderlichen zeitlichen Abständen akustische und visuelle

Reize gemeinsam präsentiert. Die Stärke des Ventriloquist-Effekts wurde dann dadurch gemessen, wie sehr der visuelle Stimulus die Lokalisierung des akustischen Reizes beeinflusste. Es bestätigte sich die zuvor aufgestellte Hypothese: Schallquellen, die deutlich als nicht aus der Richtung des Bildes kommend wahrgenommen wurden, zeigten so gut wie keinen Ventriloquist-Effekt. Bei Schallquellen hingegen, die nicht gut vom visuellen Reiz lokal trennbar waren (0-24% korrekte Lokalisierung der Schallquellen beim Test ohne visuellen Stimulus), stellte sich der Bauchredner-Effekt sehr gut ein. In diesem Fall war selbst bei einer Verzögerungszeit von 250ms die Anhaftung noch bei rund 50% der Versuche gegeben (Woods, Recanzone, 2004).

Es kann also angenommen werden, dass die Grundvoraussetzung für den Bauchredner-Effekt eine nicht eindeutig lokalisierbare Schallquelle ist. Dies hängt vom Ort der Schallquelle, aber auch wesentlich von der Beschaffenheit des Schalls ab (vgl. Kap. 3.1, S. 28: Richtungshören). Synchronität zwischen Akustischem und Visuellem spielt ebenfalls eine wichtige Rolle insofern, als dass sie die Anhaftung fördert, während zeitliche Differenzen den Effekt hemmen. Je größer die Asynchronität, desto hemmender für die Anhaftung. Zeitliche und räumliche Aspekte stehen in einem engen Verhältnis zueinander.

#### *Der Bauchredner-Nacheffekt*

Unter gewissen Umständen kann der Bauchredner-Effekt die Wahrnehmung des akustischen Raums auch nachhaltig verändern. Wenn Testpersonen für ca. fünf Minuten visuelle und akustische Reize synchron, aber räumlich versetzt dargeboten bekommen und sie nach dieser Phase aufgefordert werden, die gleichen akustischen Signale zu lokalisieren, jedoch ohne einen visuellen Reiz, so geben diese in der Regel einen falschen Ort als Schallquelle an. Ihr Wahrnehmungssystem hat sich durch kurzes ‚Training‘ angepasst und übernimmt die Verschiebung, die ihnen ihr visuelles System vorgegeben hat. Dieser Effekt lässt sich ebenso bei Affen beobachten und ist der Nachweis dafür, dass die Anhaftung des Akustischen an das Visuelle über die gemeinsame Darbietung hinaus andauern kann, das Wahrnehmungssystem sich also an die Täuschung gewöhnt (Woods, Recanzone, 2004).

### *Der McGurk-Effekt*

Dieser Effekt beschreibt das Phänomen, dass die Wahrnehmung bzw. Interpretation eines akustischen Signals durch die gleichzeitige Beobachtung von Lippenbewegungen beeinflussbar ist. Im klassischen Versuchsaufbau von McGurk und MacDonald (1976) werden die Silben „ga“ und „ba“ verwendet. Versuchspersonen wird ein Video gezeigt, in dem das Gesicht einer Person zu sehen ist, die „ga“ sagt. Der Ton wird allerdings substituiert durch ein gesprochenes „ba“. Dies hat zur Folge, dass ein Großteil der Testpersonen die Silbe „da“ versteht (Massaro, 2004).

Dies ist ein Beispiel dafür, dass die visuelle Wahrnehmung das Hören in seiner Struktur beeinflussen und verändern kann. Die Sprachwahrnehmung ist indes ein sehr weites und komplexes Forschungsgebiet; Es gibt viele verschiedene Theorien und Untersuchungen, die kognitive Mechanismen und die Bedeutung multisensorischer Integration behandeln. Im Hinblick auf die Fragestellung dieser Diplomarbeit wird darauf verzichtet, diese eingehend zu erläutern, weshalb es an dieser Stelle bei dem Beispiel bleiben soll.

Zusammenfassend lässt sich für den Einfluss der visuellen Wahrnehmung auf das Akustische sagen, dass sowohl die Lokalisation von Schallquellen beeinflussbar ist (Bauchredner-Effekt) als auch die Struktur des Gehörten durch visuelle Reize moduliert werden kann (McGurk-Effekt). Letzterer beschreibt dies für die Sprachwahrnehmung, der komplexe Verarbeitungsmechanismen zu Grunde liegen.

#### 4.1.4 Multisensorische Integration von Bewegungsinformation

Die Fähigkeit, Bewegungen von Objekten richtig einzuschätzen und zu bewerten, ist von außerordentlicher Bedeutung, denn sie verleiht einem Organismus eine hohe Anpassungsfähigkeit. Die Umwelt ist voll von bewegten Dingen, fehlende Bewegungswahrnehmung würde daher das tägliche Leben erschweren und sogar gefährlich sein. Es ist offensichtlich, dass das Einbinden verschiedener Sinne in den Prozess des Erkennens und Bewertens von Bewegungen von hohem Nutzen ist, da dies die Genauigkeit wesentlich erhöht (Soto-Farako/Kingstone, 2004).

Der folgende Abschnitt stellt einige Studien zur multisensorischen Integration von Bewegungsinformation vor. Dies ist insofern für die Betrachtung von audiovisuellen Medien von Interesse, da in Filmen oft die Illusion von Bewegung durch ein raffiniertes Zusammenspiel von Bild und Ton erzeugt wird.

#### *Multisensorische Integration zwischen statischen und bewegten Ereignissen*

Erste Studien zu diesem Thema fanden bereits 1927 statt. All die frühen Experimente bis 1954 kamen zu dem Schluss, dass akustische und taktile Reize die Bewegungswahrnehmung von visuellen „*apparent motion streams*“ verstärkt. Allen und Kolars (1981) sowie Ohmura (1987) überprüfen diese Ergebnisse. Ihre Versuche ergaben übereinstimmend, dass unbewegte akustische Reize keinen Einfluss auf die Reizschwelle von visuellen „*apparent motion streams*“ haben. Widersprüchliche Ergebnisse lieferten ihre Untersuchungen dagegen über den Einfluss statischer visueller Blitze auf die Wahrnehmung von bewegten Schallquellen. Während Allen und Kolars zu dem Schluss kamen, dass ein Blitz simultan zum ersten Geräusch die Bewegungswahrnehmung hemmt und simultan zum zweiten Geräusch keinen Einfluss hat, fand Ohmura heraus, dass ein Blitz zu jedem Zeitpunkt (simultan zu einem der Geräusche oder dazwischen) die Spanne der ISIs (*interstimulus interval*: Zeit zwischen dem ersten und zweiten Reiz) erweitert, also die Wahrnehmung von akustischen „*apparent motion streams*“ als Bewegung förderte (Soto-Farako/Kingstone, 2004).

#### *Multisensorische Integration zwischen bewegten Ereignissen*

Es gibt einige frühe Experimente zu diesem Themenbereich, wie z.B. das von Zapparoli und Reatto (1969), Allen und Kolars (1981) oder Staal und Doneri (1983). Diese kommen insofern zu ähnlichen Ergebnissen, als dass sie wenig bis keinen gegenseitigen Einfluss von bewegten Reizen in unterschiedlichen Modalitäten feststellten. Allerdings muss die Vergleichbarkeit dieser Studien in Frage gestellt werden. Zum einen wegen ihrer methodischen Unterschiede und zum anderen wurden sowohl in der Studie von Allen und Kolars als auch von Staal und Doneri die visuellen Stimuli vor Kopf und die akustischen Reize über Kopfhörer dargeboten, was nach Auffassung von Stein und Meredith (1993) sowie

Welch (1999) zu einem räumlichen Versatz zwischen akustischem und visuellem Reiz führt, der für eine multisensorische Integration hinderlich ist.

In der jüngeren Vergangenheit gab es erneut Interesse an dem Thema der multimodalen Bewegungswahrnehmung. Erwähnt sei in diesem Zusammenhang noch einmal die Studie Meyer und Wuerger (2001). Sie verwendeten ein sogenanntes *random dot kinematogram* (RDK), dessen Bewegungs-Kohärenz zwischen 0% und 32% (32% bedeutet 32 von 100 Punkten bewegen sich in die vorbestimmte Richtung) variierte. Dazu präsentierten sie weißes Rauschen, das sich in die gleiche oder entgegengesetzte Richtung bewegte. In diesen Versuchen kam heraus, dass ein bewegter akustischer Reiz nur dann Auswirkungen auf die Bewegungswahrnehmung hat, wenn der visuelle Reiz in seiner Bewegungsrichtung nicht eindeutig ist, sondern verschieden interpretierbar. Sobald dies nicht mehr zutrifft und eine visuelle Bewegung eindeutig identifizierbar ist, hat der akustische Reiz keinen Einfluss mehr auf die Bewegungswahrnehmung.

Sehr interessant ist auch ein Versuch von Kitagawa und Ichihara (2002), bei dem intermodale Post-Adaptions-Effekte genutzt wurden. Ein jeweils langsam größer, bzw. kleiner werdendes Quadrat diente als visueller Stimulus. Nach der Präsentation eines größer werdenden Quadrats empfanden die Probanden einen gleichbleibenden Ton als leiser werdend und nach Betrachtung eines kleiner werdendes Quadrat als lauter werdend. Die Adaption des visuellen Systems führte also zu einem Nach-Effekt im Hören. Dagegen bewirkten lauter oder leiser werdende Töne keine Veränderung der visuellen Wahrnehmung beim Betrachten eines konstanten Quadrats (Soto-Faraco/Kingstone, 2004).

### *Kreuzmodale Vereinnahmung der Bewegungswahrnehmung*

Soto-Faraco, Lyons et al. (2002) versuchten der Frage nachzugehen, welche Modalität in der Bewegungswahrnehmung dominiert und welchen Einfluss kreuzmodale Integration dabei hat. Sie entwarfen ein Experiment mit zwei Lautsprechern und darauf angebrachten LEDs. Mit diesem Aufbau konnten sie akustische und visuelle „*apparent motion streams*“ erzeugen. Es wurde in ihren Tests akustische und visuelle Bewegung in folgender Weise miteinander kombiniert:

1. synchron-kongruent: sowohl räumliche (Richtung) als auch zeitliche Übereinstimmung zwischen Ton und Licht
2. synchron-inkongruent: zeitliche Übereinstimmung, jedoch in die entgegengesetzte Richtung
3. asynchron-kongruent: zeitlich versetzt, jedoch in dieselbe Richtung
4. asynchron-inkongruent: zeitlich versetzt und in die entgegengesetzte Richtung

Die Testpersonen waren aufgefordert, die Bewegungsrichtung des Tons zu bestimmen. Sie wurden instruiert, das Licht nicht zu beachten und die Blickrichtung konstant geradeaus zu halten. Das Ergebnis war, dass in den Settings 1, 3 und 4 die Testpersonen zu annähernd 100% die tatsächlich vorhandene Bewegung im Ton korrekt bestimmten. Lediglich im 2ten Setting, lag die Trefferquote nur noch bei ca. 50%, d.h. bei etwa der Hälfte aller Versuche wurde die akustische Wahrnehmung vom synchronen Licht der LEDs in die entgegengesetzte Bewegung getäuscht.

Soto-Faraco und Kingstone interpretierten dieses Ergebnis so, dass im Falle einer synchronen Darbietung der Reize die Möglichkeit besteht, dass die visuelle Bewegungswahrnehmung die akustische vereinnahmt. Die Richtung bewegter akustischer Reize kann also unter gewissen Voraussetzungen falsch wahrgenommen werden, wenn synchron ein bewegter visueller Reiz stattfindet. Diese These wird gestützt durch ein weiteres Experiment unter gleichen Bedingungen, in dem die Personen zusätzlich angeben sollten, wie sicher sie sich in ihrer Entscheidung waren. Dabei kam heraus, dass die Entscheidungssicherheit durchgehend hoch war, auch bei den synchron-inkongruenten Versuchen, bei denen die Fehlerquote bei 50% lag (Soto-Faraco/Kingstone, 2004).

Bemerkenswert ist an diesem Ergebnis, dass es sich im Vergleich zu anderen Untersuchungen nicht um einen mehrdeutig interpretierbaren Reiz in einer Modalität handelt und die Testpersonen somit nicht wissen konnten, auf welchen Reiz sie ihre Entscheidung bauen sollen. Daher ist dies der Nachweis dafür, dass visuelle Bewegung tatsächlich die akustische vereinnahmen kann.

Alle erläuterten Experimente arbeiteten mit „*apparent motion streams*“. Um die Gültigkeit der erzielten Ergebnisse zu bestätigen, muss geklärt werden, ob diese tatsächlich als bewegte Reize angesehen werden können. Denn die Versuche zum Ventriloquist-Effekt haben gezeigt, dass es bei statischen Reizen eine Anhaftung der akustischen an die visuelle Wahrnehmung gibt. Ist also in dem oben beschriebenen Versuchsaufbau von Soto-Faraco und Kingstone diese Anhaftung an statische Reize ursächlich für ihre Ergebnisse?

Um diese Frage zu beantworten, führten die Forscher Vergleichstests mit demselben Versuchsaufbau unter nicht bewegten Bedingungen durch, also mit je einem akustischen und visuellen Reiz. Diese Ergebnisse verglichen sie mit denen unter dynamischen Bedingungen. Versuche mit dynamischen Reizen ergaben demnach 53% Anhaftung der Tonwahrnehmung an den visuellen Reiz (Prozent richtig bei kongruenten Reizen minus Prozent richtig bei inkongruenten Reizen), 16% Anhaftung bei gleichem Test unter statischen Bedingungen. Dieses Ergebnis zeigt, dass die erzeugte Bewegung für die Anhaftung eine wesentliche Rolle spielt. Daneben gab es weitere Tests mit störenden Blitzen, Variationen im zeitlichen Abstand der Reize und mit einer Testperson, bei der die Verbindungen zwischen den Hemisphären (Hirnhälften) gestört sind. Alle bestätigten die verstärkende Wirkung von Bewegung auf die Anhaftung von akustischer Bewegungswahrnehmung an visuelle Bewegung (Soto-Faraco, Kingstone, 2004).

### *Kognitive Stufen der Verarbeitung bei kreuzmodaler Vereinnahmung*

Um der Frage nachzugehen, auf welchen kognitiven Stufen diese Vereinnahmung akustischer Bewegungswahrnehmung abläuft, entwarfen Soto-Faraco, Kingstone und Spence (2002) ein Versuchsaufbau mit „*apparent motion streams*“ in einem sogenannten psychophysischen Treppenstufensystem, welches post-perzeptuelle Einflüsse, wie z.B. Top-Down-Verarbeitung, ausschließt. In diesem wurden visuelle und akustische „*apparent motion streams*“ zeitlich synchron vorgeführt, einmal in gleichlaufender Richtung, ein andermal in entgegengesetzte Richtungen. Dabei wurde die stimulus onset asynchrony (SOA: Dauer des motion streams) beider streams stufenweise verringert und durchlief schließlich einen Umkehrpunkt, ab dem aus gleicher Richtung (kongruent) von Akustischem und Visuellem zu Beginn des Tests die entgegengesetzte Richtung (inkongruent)

wird. Gleiches wurde auch umgekehrt durchgeführt, also Inkongruenz zu Beginn und Kongruenz zum Schluss. Ein wichtiges Ergebnis dieser Versuche war, dass bei Tests, die inkongruent begannen, Testpersonen mehrheitlich bereits vor dem Umkehrpunkt die apparent motion streams als gleichgerichtet empfanden. Daraus lässt sich ableiten, dass die intermodale Anhaftung bei der Bewegungswahrnehmung auf frühen, perzeptuellen Ebenen stattfindet (Soto-Farako, Kingstone, 2004).

Zusammenfassend lässt sich festhalten, dass Bewegungswahrnehmung ein komplexer Vorgang ist, in den die Informationen mehrerer Sinne einfließen. Die genauen Mechanismen, wie diese Integration abläuft und welche Modalität in welchen Situationen dominiert, ist nicht abschließend geklärt. Die teilweise unterschiedlichen Untersuchungsergebnisse könnten dafür stehen, dass es eine starke Abhängigkeit von der Situation, sowie von Art und Beschaffenheit der Reize gibt. Unter bestimmten Voraussetzungen sind jedoch eindeutige Interaktionen nachweisbar, so z.B. der Einfluss von bewegten akustischen Reizen auf nicht eindeutige visuelle Bewegungswahrnehmung, oder Post-Adaptions-Effekte von visueller Größe zu akustischer Lautstärke. Für den Einfluss visueller auf akustische Bewegung lässt sich sagen, dass eine Vereinnahmung der Wahrnehmung möglich ist, sofern beide Modalitäten synchrone Reize dargeboten bekommen. Zur Frage, auf welchen Stufen der Verarbeitung eine solche Anhaftung stattfindet konnte nachgewiesen werden, dass frühe, perzeptuelle Ebenen definitiv eine Rolle spielen.

#### 4.1.5 Kreuzmodale Interaktionen in schnellen Klassifizierungen

Unsere Wahrnehmung kann nur einen Bruchteil der auf uns einströmenden Informationen, die uns meist über verschiedene Modalitäten gleichzeitig erreichen, verarbeiten. Aus diesem Grund gibt es eine Selektion der Informationen bevor sie das Bewusstsein erreichen. Die Frage, wie diese Selektion funktioniert und insbesondere, wann sie fehlschlägt ist von zentraler Bedeutung für das Verstehen früher Prozesse der Wahrnehmung. In diesem Kapitel wird erläutert, wie sich die Sinne auf der Ebene dieser frühen Selektion gegenseitig beeinflussen. Dies lässt sich mit Hilfe von sogenannten „*speeded*

*classifications*“ untersuchen. Das sind Experimente, in denen Testpersonen möglichst schnell bestimmte Reize klassifizieren, das heißt qualitativ beurteilen sollen. Das kann z.B. die Helligkeit eines visuellen Reizes sein. Durch gleichzeitiges Ansprechen mehrerer Sinne kann somit der Einfluss anderer Modalitäten auf die Wahrnehmung des Zielreizes gemessen werden. Kommt es zu einer Veränderung der Beurteilung durch einen weiteren Reiz, so nennt man dies „*dimensional interactions*“ oder „*failures of selective attention*“ (Marks, 2004).

### *Hintergrund, Methoden und Definitionen*

Der einfachste Versuchsaufbau einer schnellen Klassifizierungen besteht aus der Darbietung eines relevanten Reizes  $R$ , der zwei Zustände  $R_a$  und  $R_b$  annehmen kann, sowie eines irrelevanten Reizes  $I$ , ebenfalls in zwei möglichen Zuständen  $I_a$  und  $I_b$ . Daraus ergeben sich vier Kombinationen:  $R_a + I_a$ ;  $R_b + I_a$ ;  $R_a + I_b$ ;  $R_b + I_b$ . Unterschieden werden „*baseline tasks (identification)*“, bei denen der irrelevante Reiz konstant bleibt und nur der zu klassifizierende (relevante) Reiz variabel ist und „*selective tasks (classification)*“, bei denen alle Reize veränderbar sind.

Garner (1977) unterscheidet ferner zwischen „*seperable*“ und „*integral dimensions*“. Er stellte fest, dass es Dimensionen gibt (z.B. Farbe und Größe) die vollständig trennbar sind, sich also in schnellen Klassifizierungen nicht gegenseitig beeinflussen. Anders verhält es sich z.B. mit den visuellen Attributen Farbton, Sättigung und Helligkeit untereinander, die sich in der Wahrnehmung gegenseitig beeinflussen (*integral*). Für „*integral dimensions*“ gilt, dass die „*performance*“ (Fähigkeit, die Reize unabhängig voneinander zu klassifizieren) in „*selective tasks*“ schlechter ist, als in „*baseline tasks*“. Mislingt es den Testpersonen, die Attribute einer irrelevanten Dimension bei der Beurteilung einer relevanten zu ignorieren, so bezeichnet man dieses Übersprechen als „*Garner interference*“. Es stellte sich heraus, dass Garner Interferenzen bei intramodalen Interaktionen stärker ausgeprägt ist, als bei intermodalen, sich also tendenziell Dimensionen unterschiedlicher Modalitäten besser separieren lassen, als Dimensionen innerhalb einer Modalität. Hier besteht allerdings einen klarer Unterschied zwischen schnellen Klassifizierungen und solchen, die nicht auf Schnelligkeit angelegt sind („*unspeded task*“): So

sind z.B. die Lautstärke eines Tons und die Helligkeit eines Lichts „*seperable*“ in „*unspeded tasks*“ während es in „*speded tasks*“ zu einem deutlichen Übersprechen zwischen ihnen kam, es also dort „*integral dimensions*“ sind (Marks, 2004).

Zusätzlich zur „*Garner interference*“ kommt es in „*speded classifications*“ zu einer weiteren Form der Interaktion. Der sogenannte Kongruenz-Effekt („*congruence effect*“) beschreibt die Beziehung mehrerer Dimensionen zueinander, die auf eine bestimmte Weise besonders passend zueinander verlaufen. Als kongruent werden solche Dimensionen bezeichnet, die gemeinsam dargeboten schneller klassifiziert werden, als es bei ihrer alleinigen Darbietung der Fall ist. Sie scheinen demnach die Wahrnehmung der jeweils anderen zu unterstützen. Kongruenz-Effekte treten in der Regel zusammen mit Garner Interferenzen auf, also selten allein, wohingegen dies bei Interferenz-Effekten durchaus der Fall ist. Das gilt sowohl für unimodale, als auch für multimodale Stimuli. Doch was genau bedeutet „kongruent“? Es gibt verschiedene Ansätze, sich diesem Begriff zu nähern. Es kann die Übereinstimmung von physikalischen Eigenschaften sein, wie z.B. bei Clark und Brownell (1975). Sie führten schnelle Klassifizierungen von Pfeilen durch, deren Zeigerichtung (nach oben/ nach unten) bestimmt werden sollte. Dabei fanden sie Interaktionen zwischen der Pfeilrichtung und der vertikalen Position der Pfeile. Andere Konzepte sind dagegen abstrakter. Kornblum, Hasbroucq und Osman (1990) nutzen den Begriff „*dimensional overlap*“, und bezogen sich dabei auf Korrespondenzen zwischen der räumlichen Anordnung von Stimuli und den räumlichen Mustern der gemessenen Reaktionen. Verallgemeinert lässt sich formulieren: Kongruenzen bezeichnen Gemeinsamkeiten in relativen (seltener absoluten) Eigenschaften oder Wahrnehmungs-Attributen zwischen verschiedenen Dimensionen von Stimuli. Typischerweise treten sie bei polaren Dimensionen auf, wie z.B. Lautstärke, Größe, Helligkeit, Sättigung (Gegenbeispiel: Farbton) (Marks, 2004).

### ***Kreuzmodale selektive Aufmerksamkeit***

Interferenz- und Kongruenz-Effekte treten sowohl bei unimodalen, als auch bei multimodalen Stimuli auf. Für unimodale Stimuli wurde vielfach vorgeschlagen, dass diese Effekte das Ergebnis einer holistischen frühen perzeptuellen Verarbeitung sind. Diese Annahme kann als fragwürdig angesehen werden,

insbesondere für Interaktionen zwischen multimodalen Stimuli scheint sie unzutreffend zu sein. Denn laut dieser These werden Informationen über die Intensität einzelner Dimensionen, wie z.B. die Tonhöhe und die Lautstärke beim Hören oder die Helligkeit und die Sättigung beim Sehen, erst relativ spät in der Verarbeitungskette differenziert. Auf frühen Stufen der Verarbeitung sind die Reize dagegen noch sehr undifferenziert, so die Hypothese. Diese Annahmen können allerdings für die Effekte in *speeded classifications* keine schlüssige Erklärung liefern, da sie davon ausgehen, dass auf den frühen Wahrnehmungsebenen noch keine Trennung zwischen den Sinnen vorhanden ist. Viele Studien belegen jedoch, dass genau das der Fall ist. Trotzdem können die Interaktionen auf frühe Ebenen zurückzuführen sein. Anzunehmen ist jedoch vor allem bei den Kongruenz-Effekten, dass diese ihre Ursache eher auf späteren Ebenen der Re-Kodierung und Entscheidung haben (Marks, 2004).

**KREUZMODALE KONGRUENZEN** Eine einfache empirische Methode Kongruenz-Beziehungen herauszufinden ist die einer einfachen Befragung; Testpersonen werden Reize einer Modalität *A* in zwei Zuständen sowie einer Modalität *B* in zwei Zuständen präsentiert. Anschließend sollen die Personen bestimmen, welcher Stimulus aus *A* zu welchem aus *B* passt. Versuche von Marks, Hammeal und Bornstein (1987) fanden auf diese Art deutliche Kongruenzen zwischen akustischen und visuellen Reizen in den Dimensionen Tonhöhe-Helligkeit sowie Lautstärke-Helligkeit bei Erwachsenen und bis vier Jahre alten Kindern. Visuell ‚hell‘ verhält sich demnach kongruent zu akustisch ‚hoch‘ sowie ‚laut‘, visuell ‚dunkel‘ zu akustisch ‚tief‘ und ‚leise‘. Dabei ist es wichtig zwischen der Helligkeit eines Lichts und der einer Oberfläche zu unterscheiden, denn nur die des Lichts wird mit Tonhöhe und Lautstärke in Verbindung gebracht. Die Helligkeit einer Oberfläche erwies sich dagegen nur als kongruent zur Tonhöhe, jedoch nicht zur Lautstärke. Anzumerken ist auch, dass mit den Dimensionen Lautstärke, Tonhöhe sowie Helligkeit akustische und visuelle Reize niemals erschöpfend beschrieben sind. Ferner können die Dauer, die räumliche Ausdehnung und Position eine Rolle spielen. Räumliche Position stellt sich als besonders interessante Dimension heraus, denn die Bezeichnungen ‚hoch‘ und ‚tief‘ finden sich in fast allen Sprachen der Welt als Attribute für Tonhöhe und Position (Marks, 2004).

**KREUZMODALE INTERAKTIONEN** Von kreuzmodalen Interaktionen in schnellen Klassifizierungen und insbesondere solchen, die Kongruenz-Effekte beinhalten, wurden wahrscheinlich zuerst durch Bernstein und Edelstein (1971) berichtet. Sie ließen die vertikale Position visueller Reize beurteilen (relevanter Reiz). Dabei stellten sie eine Verbesserung der „*performance*“ durch simultane Darbietung von kongruenten Tönen (hoch/tief) fest. Melara und O’Brien (1987) erweiterten diese Versuche. Sie bestätigten die Befunde von Bernstein und Edelstein und stellten darüber hinaus fest, dass das willkürliche Variieren des irrelevanten (akustischen) Reizes Garner Interferenzen verursacht, also die Wahrnehmung des (visuellen) Ziel-Reizes behindert, was ein zusätzliches Indiz für die Kongruenz-Beziehung zwischen Tonhöhe und Position ist. Ben-Artzi und Marks (1995) führten ebenfalls „*speeded classifications*“ zwischen Position und Tonhöhe durch. Zusätzlich variierten sie auch die physikalische Differenz zwischen den Reizen. Dadurch machten sie zwei neue Entdeckungen. Erstens: Die Tonhöhendifferenz hatte einen wesentlich höheren Einfluss auf Kongruenz- und Interferenz-Effekte, als die Höhendifferenz der visuellen Stimuli. Zweitens: Interferenz- sowie Kongruenz-Effekte waren stärker, wenn die Testpersonen die Tonhöhe beurteilten sollten (relevanter Reiz), als wenn die Klassifizierung der Position gefragt war. Dabei ist zu beachten, dass in diesen Studien von Melara und O’Brien, Ben-Artzi und Marks (1995), als auch einer weiteren von Patching und Quinlan (2002), die Performanz in „*baseline tasks*“ im Visuellen besser war als im Akustischen. Das heißt, die Beurteilung der vertikalen Position verlief im Durchschnitt schneller, als die der Tonhöhe. Dieser Unterschied könnte mit der Asymmetrie in den „*selektive tasks*“ zusammenhängen. Melara und Mounts (1993) stellten heraus: Wenn Testpersonen in „*baseline tasks*“ schneller auf eine der beiden Reiz-Dimension reagieren, gibt es die Tendenz, dass Interaktionen in „*selective tasks*“ stärker sind, wenn die ‚ärmere‘ Dimension beurteilt wird (relevante Dimension). Der Grund hierfür könnte sein, dass die in „*baseline tasks*“ schneller klassifizierbare Dimension als irrelevante Dimension in „*selective tasks*“ stärkeren Einfluss ausübt, als die langsamer klassifizierbare Dimension.

Insgesamt wurden durch „*speeded classifications*“ Interaktionen zwischen folgenden Dimensionen ermittelt:

1. Tonhöhe und visuelle Position (hoch-tief)
2. Tonhöhe und Farbe (schwarz und weiß)
3. Tonhöhe und Helligkeit
4. Lautstärke und Helligkeit
5. Tonhöhe und Form (rund und eckig)

(Marks, 2004).

#### KONGRUENZ-EFFEKTE, SYNÄSTHESIE UND KREUZMODALE ÄHNLICHKEIT

Zu vielen der festgestellten kreuzmodalen Interaktionen gibt es analog auftretende Synästhesien. Das ist ein Phänomen, bei dem Betroffene Wahrnehmungen in einer Modalität haben, die durch die Reizung einer anderen ausgelöst wird, sie z.B. Farben sehen, wenn sie Töne hören (Anm.: Synästhesien werden ausführlich in Kap. 5 besprochen). So gibt es bei eben jenen Synästheten, bei denen visuelles Erleben durch Klänge verursacht wird, oft einen starken Zusammenhang zwischen der Tonhöhe und der Helligkeit der bei ihnen hervorgerufenen Farbe. Hohe Töne verursachen hellere Farben als tiefe Töne.

Es gibt demnach drei Bereiche, in denen kreuzmodale Kongruenz-Beziehungen auftreten: In kreuzmodalen Entsprechungen und dazu analogen Messungen der Ähnlichkeit, die dazu dienen, Kongruenz-Beziehungen zu definieren, in Synästhesien und in „*speeded classifications*“. Einige dieser Kongruenzen konnten sogar bei 4-jährigen Kindern, teilweise sogar bei einem Monat alten Säuglingen nachgewiesen werden, wie die zwischen Helligkeit und Lautstärke. Diese Beobachtungen stützen die Hypothese, dass manche Kongruenz-Verhältnisse, wie die zwischen Tonhöhe, Lautstärke und Helligkeit, auf einer intrinsischen Ähnlichkeit von Dimensionen unterschiedlicher Modalitäten beruhen. Darüber hinaus könnten solche erlebten Ähnlichkeiten, zumindest zum Teil, von ähnlichen neuronalen Prozessen herrühren. So ist z.B. über das Wahrnehmen von Intensität bekannt, dass diese durch die Feuerungsrate von Neuronen (Anzahl der Nervenimpulse) abgebildet wird, unabhängig von der beteiligten Modalität. Helligkeit, Lautstärke als auch Tonhöhe könnten demzufolge in Form von Reiz-Intensitäten abgebildet werden. Doch es ist stets Vorsicht vor zu schnellen, einfachen Schlüssen geboten: Selbst wenn Prinzipien der neuronalen Kodierung als Ursache von kreuzmodalen Ähnlichkeiten in

Betrachtet kommen, die Kontextbezogenheit von kreuzmodalen Kongruenzen bei Erwachsenen impliziert, dass ein simpler neuronaler Code wie die Feuerungsrate als alleiniges Prinzip nicht ausreicht (Marks, 2004).

### *Kongruenz-Interaktionen zu Grunde liegende Mechanismen*

Aus den bisherigen Erkenntnissen lassen sich zunächst drei Grundprinzipien für die Verarbeitung multisensorischer Stimuli ableiten:

1. In Situationen, in denen die Stimulation einer Modalität relevant im Sinne der Aufgabenstellung ist und die einer anderen Modalität irrelevant, wird dennoch auch die irrelevante verarbeitet. Wäre dem nicht so, wären sie nicht in der Lage, die Wahrnehmung der relevanten Reize zu beeinflussen und zu verändern.
2. Die Stimulation einer irrelevanten Modalität kann die Verarbeitung einer relevanten besonders beeinflussen, wenn es Kongruenz-Beziehungen zwischen ihnen gibt.
3. Kongruenz-Beziehungen sind typischerweise kontextabhängig. Jeder Reiz wird in seiner Qualität relativ zu einem anderen beurteilt. Die Bewertung eines Lichtes geschieht im Vergleich mit einem anderen Licht, eine Tonhöhe oder Lautstärke wird stets im Vergleich zu anderen Tönen bestimmt. Daher ist es auch schlüssig, dass Kongruenz-Effekte nur in „*selektive attention tasks*“ auftreten, in denen sich beide Stimuli von Versuch zu Versuch ändern und nicht in „*baseline tasks*“, bei denen nur ein Stimulus variabel ist, während der andere konstant bleibt.

Wie bereits geschildert, kann die Hypothese einer ganzheitlichen, frühen Wahrnehmungsverarbeitung („*early holistic hypothesis*“) nicht die Erklärung für die dargestellten Effekte sein. Schlüssiger scheint die Vorstellung von einem Übersprechen zwischen Quellen von Information. Es gibt drei mögliche Ansätze dazu, wie das ablaufen könnte. Die erste Möglichkeit sind konkurrierende, oder überlappende Stimulationen unterschiedlicher Sinne, die sich gegenseitig in ihrer perzeptuellen Repräsentation beeinflussen. So könnte z.B. ein Ton die Helligkeit eines Lichts verstärken oder abschwächen, je nachdem, ob der Ton relativ laut oder leise ist. Die zweite Möglichkeit ist, dass die Repräsentationen der verschiedenen Stimuli unabhängig voneinander, ohne kreuzmodale Interaktionen verarbeitet werden. Beeinflussungen träten dann auf, wenn die

perzeptuell gewonnenen Informationen auf höheren Stufen um-kodiert werden, z.B. semantische oder linguistische. Oder drittens könnte es bei unabhängiger Verarbeitung und Kodierung auf perzeptuellen und semantischen Ebenen zu einem Übersprechen auf der Entscheidungsebene kommen. Diese drei Möglichkeiten sind allerdings nicht getrennt voneinander zu betrachten, vielmehr laufen für unterschiedliche multimodale Stimuli verschiedene Mechanismen ab, meist auch auf mehr als einer Ebene.

Robert D. Melara (1989) versuchte einen Beleg für Interaktionen der perzeptuellen Repräsentationen, also auf frühen Stufen zu finden. Dafür entwarf er ein Experiment, in dem Probanden multimodale Stimuli aus unterschiedlich hellem Licht und verschieden hohen Tönen klassifizieren sollten. Diese Tests waren nicht auf Schnelligkeit angelegt, also keine „*speeded tasks*“. Melaras These war, dass sich Tonhöhe und Helligkeit kongruent verhalten und es zu Interaktionen dieser Dimensionen auf perzeptueller Ebene kommt. Die Ergebnisse seiner Versuche bestätigen diese These allerdings nicht. Vieles spricht eher dafür, dass Beeinflussungen, die auf Kongruenz basieren, durch fehlerhafte oder eingeschränkte Separabilität auf Entscheidungsstufen basieren. Nach einem solchen Modell werden die Qualitäten der Stimuli in den Modalitäten weitestgehend getrennt verarbeitet. Bei kongruenten Stimuli fällt es den Entscheidungsmechanismen dann leichter, die Stimuli zu klassifizieren als bei inkongruenten Stimuli. Das würde auch erklären, warum Kongruenzen in erster Linie in schnellen Klassifizierungen auftreten (Marks, 2004).

### *Kreuzmodale Kongruenz-Interaktionen: Welche Rolle spielen sprachliche Prozesse?*

Eine der Hauptfragen ist, in welcher Form Informationen zu den Entscheidungsstufen gelangen, insbesondere, ob sie semantischer, oder gar sprachlicher Natur sind. Auch wenn Kongruenz-Effekte nicht immer explizite sprachliche Kennzeichnungen benötigen, so könnten sie dennoch auf impliziten Kennzeichnungen der Perzepte beruhen, auch wenn diese, auf die verschiedenen Reize angewandten Kennzeichnungen, nicht identisch sind. Es könnte sein, dass die Eigenschaften von Stimuli in abstrakten Dimensionen, wie z.B. Intensität, kodiert werden, welche übergeordnete Klassifizierungen darstellen und über die Repräsentationen der einzelnen Modalitäten

hinausgehen. Diese These findet Bestätigung in den Versuchen von Walker und Smith (1984), die schnelle Klassifizierung mit einem sprachlichen und einem nicht-sprachlichen Stimulus durchführten. Dabei fanden sie Kongruenz-Effekte zwischen tiefen und hohen Tönen mit Worten wie *blunt* und *sharp*. Melara und Marks (1990) untersuchten systematisch Interaktionen innerhalb und zwischen visuellen und akustischen Stimuli, deren Attribute binnen einer Dimension variierte, die mit „hoch“ und „tief“ beurteilt werden konnte. Dabei verwendeten sie sowohl unimodale Versuchsanordnungen, wie z.B. die akustische Darbietung der Wörter „low“ und „high“ in Verbindung mit Variationen der Tonhöhe oder der vertikalen Position der Schallquelle, als auch multimodale, wie die visuelle Darbietung der Zeichenkette „LO“ und „HI“ zusammen mit Tönen unterschiedlicher Frequenz. In allen Fällen fanden sie sowohl Garner Interferenzen als auch Kongruenz-Effekte (Marks, 2004).

**HYPOTHESE DER SEMANTISCHEN KODIERUNG**                      Aufbauend auf diesen Erkenntnissen formulierten Martino und Marks (1999) die Hypothese, dass Kongruenz-Effekte entstehen, weil Stimuli abstrakte Repräsentationen aktivieren, deren Dimensionen oder Eigenschaften (linguistisch oder semantisch) zumindest teilweise überlappen. Einige dieser Kongruenz-Effekte scheinen in der Natur der Wahrnehmung zu liegen, andere wiederum eher das Ergebnis von erlernten Assoziationen zu sein, wie z.B. die Verbindung von tiefer werdenden Tönen mit größer werdenden Objekten, oder die Assoziationen von Farben mit Temperaturen. Bei den beiden letztgenannten Beispielen kommen Untersuchungen mit Kindern zu anderen Ergebnissen, als mit Erwachsenen, was für eine erlernte Verbindung spricht. Kongruenz-Effekte zwischen Lautstärke und Helligkeit sowie Tonhöhe und Helligkeit sind dagegen auch bei Kindern nachweisbar, was für eine Verbindung auf frühen, perzeptuellen Ebenen spricht. Auch wenn einige Kongruenzen auf perzeptuelle Prozesse zurückzuführen sind, so überwiegen mit den Jahren der Erfahrung eines Menschen in Wahrnehmung und Sprache eher die semantisch und linguistisch erlernten Verbindungen. Diese führen dazu, dass Menschen gewisse Reize mit Hilfe eines anderen schneller identifizieren können, selbst wenn sie versuchen, den irrelevanten Reiz zu ignorieren (Marks, 2004).

#### DAS ENTSCHEIDUNGS-MODELL KREUZMODALER INTERAKTIONEN BEI SELEKTIVER AUFMERKSAMKEIT

Die Hypothese der semantischen Kodierung kann innerhalb eines Entscheidungs-Modells angewandt werden, welches auf den Prinzipien der „*signal detection theory*“ beruht. Dieses Modell macht zwei Hauptannahmen: Die erste ist eine über die Art, wie sich die aufgenommene Information, die durch die Stimulation einer relevanten Dimension hervorgerufen wird, über die Zeit ändert. Die zweite ist eine über die Weise, wie ein irrelevanter Reiz diesen Aufnahmeprozess verändert. Das Modell geht davon aus, dass unmittelbar nach Auftreten des Reizes die Entscheidung, ob z.B. ein Ton hoch oder tief ist, noch relativ ungenau ist. Je mehr Zeit vergeht, desto besser kann der Reiz qualitativ beurteilt werden. Ein irrelevanter Reiz (z.B. ein helles oder dunkles Licht) erzeugt dabei eine Art Verschiebung, so dass die Entscheidung länger dauert, wenn die Reize gegenläufig sind und kürzer, wenn die Reize kongruent sind. Dieses Modell geht demnach von Entscheidungskriterien aus und nicht von perzeptuellen Mechanismen.

Anzumerken ist, dass es hier um die Interpretation von Ergebnissen aus schnellen Klassifizierungen geht, also Tests, in denen die Subjekte so schnell wie möglich „entscheiden“ sollen. Studien, die nicht auf Schnelligkeit angelegt sind (z.B. Melara, 1989; Odgaard, 2003) beruhen dagegen auf Wahrnehmungs-Repräsentationen, die in frühen Phasen der Verarbeitung stattfinden (Marks, 2004).

#### *Kreuzmodale Interaktionen: Perzeptuelle Kongruenz und Stimulus-Kongruenz*

Abschließend wird die Perzeptuelle Kongruenz, welche alle bisher aufgeführten Experimente und Untersuchungen betrifft, von der Stimulus-Kongruenz unterschieden. Perzeptuelle Kongruenz ist diejenige, die auf psychophysische Prozesse zurückzuführen ist, wie z.B. die zwischen Tonhöhe und Lautstärke oder Helligkeit und Tonhöhe.

Stimulus-Kongruenz ist hingegen auf Eigenschaften der Stimuli selbst zurückzuführen, wenn also Ähnlichkeit oder Gleichheit in einer oder mehreren Eigenschaften besteht. Ein Beispiel ist die räumliche Position: Visuelle, akustische und taktile Reize an derselben Position (somit unter Umständen von derselben Quelle ausgehend), können als räumlich-kongruent definiert werden (vgl. Kap. 4.1.4, *Kreuzmodale Vereinnahmung der Bewegungswahrnehmung*).

Marks formuliert die Hypothese, dass perzeptuelle Kongruenzen auf späte, vielleicht sogar auf Entscheidungsprozesse zurückzuführen sind, wohingegen Effekte, die durch Stimulus Kongruenz ausgelöst werden, auf frühe, perzeptuelle Prozesse zurückzuführen sind. Denn die Unterschiedlichkeit der Untersuchungsergebnisse, vor allem zwischen perzeptueller und räumlicher Kongruenz, legen den Schluss nahe, dass in schnellen Klassifizierungen intra- und intermodale Integration auf vielen verschiedenen Ebenen stattfindet, sowohl früh-perzeptuell, als auch entscheidungs- und reaktions-basiert.

### *Zusammenfassung*

Bei Prozessen der Aufmerksamkeit kommt es zu Interaktionen zwischen Dimensionen von Sinnesqualitäten, sowohl unimodal, als auch multimodal. Diese können sich in Form von gegenseitiger Hemmung (Garner Interferenzen) oder Kongruenzen äußern. Eine Möglichkeit, solche Interaktionen zu messen, bieten schnelle Klassifikationen. Dies sind Studien, in denen Testpersonen (bei multimodalen *tasks*) einen relevanten Reiz in Anwesenheit eines irrelevanten Reizes qualitativ beurteilen sollen, folglich ihre Aufmerksamkeit selektiv auf eine Sinnesqualität oder Dimension richten. Mit dieser Methode konnten folgende kreuzmodale Interaktionen zwischen akustischen und visuellen Dimensionen nachgewiesen werden: Tonhöhe und Position, Tonhöhe und Farbe (schwarz/weiß), Tonhöhe und Helligkeit, Lautstärke und Helligkeit sowie Tonhöhe und Form (rund/eckig). Bei der Suche nach zu Grunde liegenden Mechanismen lässt sich zunächst feststellen, dass es zu vielen der bei Normalpersonen gefundenen Kongruenzen analoge Synästhesien gibt. Ferner scheinen sprachliche Prozesse sowie semantische Kodierung von Information eine Rolle zu spielen. Da es sich in schnellen Klassifizierungen um Entscheidungen handelt (z.B: ob ein Reiz hell oder dunkel ist), wird davon ausgegangen, dass Top-Down-Verarbeitung eine Rolle spielt, ebenso wie Interaktionen früher perzeptueller Ebenen. Ein Modell hierzu ist das Entscheidungsmodell in dem die Hypothese der semantischen Kodierung angewandt wird.

## 4.2 Zusammenfassung Multisensorische Prozesse

In diesem Kapitel wurden Studien vorgestellt, die die gegenseitige Beeinflussung von akustischer und visueller Wahrnehmung untersuchen. Es hat sich gezeigt, dass es viele verschiedene Arten der Interaktion gibt und alle Stufen der Verarbeitung in multisensorische Prozesse involviert sind. Die Lenkung von Aufmerksamkeit auf ein Ereignis ist in der Regel ein multisensorischer Prozess, welcher erhöhte Sensitivität in einem Bereich zur Folge hat. Das bedeutet, es kommt zu offenen, äußeren Verlagerungen z.B. in Form von Bewegungen des Kopfes, aber auch zu erhöhter kognitiver Aktivität, sowohl in Arealen des Hinweis- als auch des Zielreizes.

Es gibt Modulationen der Wahrnehmung durch akustische und visuelle Reize auf die jeweils andere Modalität. Das kann temporäre Aspekte betreffen, ebenso wie die perzeptuelle Organisation, die Bewegungswahrnehmung, die räumliche Anhaftung an einen anderen Reiz, als auch grundlegende Veränderungen der Wahrnehmungsstruktur.

Weitreichend sind auch Veränderungen der Wahrnehmung von Reizen in schnellen Klassifizierungen durch gleichzeitiges Präsentieren eines irrelevanten Reizes. An solchen Interaktionen bei selektiver Aufmerksamkeit lassen sich Kongruenz-Beziehungen zwischen Dimensionen von Reizen erfassen, die auf Ähnlichkeiten in der Repräsentation durch neuronale Strukturen, als auch auf Übersprechen innerhalb semantischer Kodierungsprozesse und Entscheidungsstufen beruhen.

## 5 SYNÄSTHESIE

Synästhesie (Grischisch, *syn* = zusammen & *aisthesis* = Wahrnehmung) ist die unfreiwillige physische Erfahrung kreuzmodaler Assoziationen. Das heißt, die Stimulation einer sensorischen Modalität verursacht zuverlässig Wahrnehmungen in einer, oder mehreren weiteren Modalitäten. Es gibt einige demografische und kognitive Auffälligkeiten im Bezug auf Synästhesien: Sie treten vermehrt unter Frauen und Linkshändern, sowie gehäuft in Familien auf. Die Gedächtnisleistung von Synästheten ist tendenziell über dem Durchschnitt, mathematische und Orientierungsfähigkeiten dagegen eher unterdurchschnittlich. Synästhesien sind anormal nur in dem Sinne, dass sie statistisch selten sind. Es handelt sich faktisch um normale Hirnprozesse, die bei nur einem geringen Teil der Bevölkerung in die bewusste Wahrnehmung gelangen (Cytowic, 1995).

In der jüngeren Vergangenheit ist das Interesse der Forschung an dem Phänomen Synästhesie stark gestiegen. Es wurde erkannt, dass es sich um ein perzeptuelles Phänomen handelt und Betroffene die synästhetischen Eindrücke tatsächlich erleben und sie nicht bloß ein besonders ausgeprägtes Assoziationsvermögen haben. Daher erhoffen sich Forscher über Untersuchungen an Synästheten Erkenntnisse über neuronale Grundlagen von allgemeinen Wahrnehmungs- und Kognitionsprozessen zu erlangen. Dies betrifft gerade auch multisensorische Prozesse, da Synästhesien eine besondere Form der kreuzmodalen Interaktion sind.

Der amerikanische Neurologe Vilayanur S. Ramachandran betreibt Forschungen an Synästheten und hat in einigen Aufsätzen Thesen über mögliche zu Grunde liegende neuronale Prozesse entwickelt. Ausgehend von der Beobachtung, dass Synästhesien deutlich häufiger unter Künstlern, Poeten und Schriftstellern vorkommen, als in der Gesamtbevölkerung, spekuliert er über Zusammenhänge zu kreativem Denken, künstlerischem Schaffen und Metaphern. Er argumentiert, dass das Verstehen von Synästhesien ein experimentelles Mittel bereithalten kann, um die neuronalen Mechanismen von Metaphern zu entschlüsseln, da diese scheinbar unverwandte Konzepte miteinander verknüpfen, analog zu

Synästhesien, die scheinbar unverwandte Perzepte verbinden (Ramachandran/Hubbard, 2001a & 2001b).

Heyrman geht indes noch weiter und formuliert die Hypothese, dass synästhetische Phänomene der Ursprung allen künstlerischen Schaffens sind. Er argumentiert, dass Kunst als synästhetische Erfahrung und die synästhetischen Erfahrungen von Synästheten das grundlegende Konzept gemein haben, neue Verbindungen zwischen den Sinnen zu schaffen. Auch er schlägt eine Brücke zu Metaphern und nennt sie einen symbolischen Ausdruck, der immer einen Transfer zwischen zwei verschiedenen Kontexten beinhaltet. Sie seien deshalb nicht nur eine sprachliche, sondern auch eine gedankliche Figur. In der Kunst sind Synästhesien und Metaphern vereint, durch sie werden synästhetische Erfahrungen mitteilbar, so Heyrman (Heyrman, 2005).

In Hinblick auf die die Fragestellung dieser Diplomarbeit sind Synästhesien von Interesse, da es sich um kreuzmodale Interaktionen handelt, die sich dadurch kennzeichnen, dass Betroffene sie wie reale Sinnesreize erleben. Das macht es verhältnismäßig einfach, die Interaktionen zwischen den Sinnen zu beobachten und zu messen. Aus solchen Untersuchungen lassen sich indirekt Zusammenhänge auch über kreuzmodale Prozesse zwischen akustischer und visueller Wahrnehmung bei normalen Personen ableiten. Die Forschungsarbeit von V. S. Ramachandran ist in diesem Zusammenhang von besonderer Bedeutung, da er aus Erkenntnissen über Synästhesie Thesen von allgemeiner Gültigkeit ableitet.

In diesem Kapitel werden zunächst allgemeine Grundlagen zur Synästhesie geklärt. Im Anschluss werden einige der Studien von Ramachandran et al. vorgestellt und seine Hypothesen zu neuronalen Mechanismen, Verbindungen zu Metaphern sowie zu Kreativität und Kunst erläutert.

## 5.1 Allgemeine Grundlagen

Synästhesien sind individuell verschieden und in der Regel lebenslang vorhanden. Betroffene berichten, dass sie diese Form der Wahrnehmung haben, solange sie sich zurückerinnern können. Sie unterliegen nicht der willentlichen Kontrolle und sind weder vom Kontext, noch der kognitiven Einstellung abhängig. Frauen und Linkshänder sind dominierend, wobei es unterschiedliche Schätzungen zum Männer-Frauen-Verhältnis gibt, 3:1 bei Cytowic (1989) und 6:1 bei Baron-Cohen et al. (1996). Auch über die Häufigkeit in der Bevölkerung variieren die Schätzungen: Cytowic (1989) gibt 1:20.000 an, Ramachandran (2001a) 1:200.

Synästhesien sind keine Anomalitäten, aus klinischer Sicht sind Synästheten mental ausgewogen. Sie verfügen allgemein über gute Gedächtnisleistungen und sind überdurchschnittlich intelligent. Allerdings weisen sie häufig mathematische Defizite auf, leiden öfter unter Allocheirie (Rechts-Links-Verwechslung) und Orientierungsschwierigkeiten.

Die Diversität von Synästhesien ist indes so groß wie die Möglichkeiten der Verknüpfung zwischen den Sinnen, es gibt eine starke individuelle Streuung, jedoch große Konstanz im Individuum. Mit Abstand am häufigsten sind Graphem-Farben Synästhesien mit rund 65%, gefolgt von Zeiteinheiten-Farben mit 23%, musikalische Klänge-Farben (19%) sowie allgemeine Klänge-Farben mit 15%. Alle weiteren der 60 Formen, die Sean Day in seinem Synästhesie-Forum laut Befragung von Betroffenen aufzählt, liegen bei Häufigkeiten unter 10% (Day, 2009). Auffallend ist, dass die visuelle Wahrnehmung meist die angesprochene Modalität ist, nicht die auslösende.

Synästhesien treten vermehrt in Familien auf: Ein Drittel der von Baron-Cohen (1996) Befragten gab an, in der Familie Synästheten zu haben, was die Vermutung nahelegt, dass Synästhesien vererbbar sind und somit durch genetische Dispositionen ausgelöst werden könnten (Cytowic, 1989; Ramachandran/Hubbard, 2001a & 2001b; Haverkamp, 2003; Day, 2009).

## 5.2 Forschung

Im folgenden Abschnitt werden einige Forschungsergebnisse und Hypothesen von Ramachandran et al. vorgestellt. Zunächst erfolgt eine Rückschau auf einige zurückliegende Experimente, die deutlich zeigen, dass Synästhesien grundsätzlich perzeptuell sind. Anschließend werden einige scheinbar nicht verwandte Fakten über Synästhesie sowie andere neurologische Funktionsstörungen betrachtet und daraus eine neue, kohärente theoretische Perspektive initiiert. Es wird die Relevanz dieses Schemas für einige rätselhafte Aspekte der menschlichen Natur diskutiert, wie metaphorisches Denken, Kunst und die Evolution der Sprache.

Folgende bekannte Fakten bilden die Grundlage der theoretischen Überlegungen:

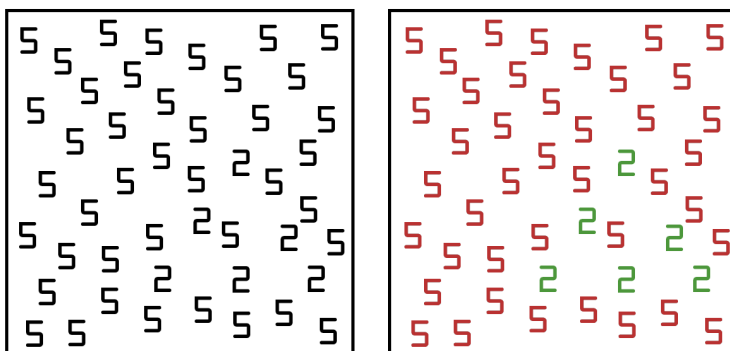
1. Synästhesien treten gehäuft in Familien auf
2. Synästheten berichten oft über außergewöhnliche Farben, die sie in der realen Welt nicht sehen können, sondern nur synästhetisch erleben. Eine Person, die eine Synästhesie hat, hat mit großer Wahrscheinlichkeit auch eine zweite, oder dritte.
3. Es gibt große Heterogenität zwischen den Synästhesien.
4. Patienten mit einem Defekt des linken Gyrus angularis leiden unter Dyskalkulie: sie können selbst einfache Rechenoperationen nicht durchführen, aber immer noch Zahlen erkennen.
5. Der Gyrus angularis ist der Sitz der multimodalen Konvergenz von sensorischen Informationen
6. Defekte des Gyrus angularis führen darüber hinaus zu Anomia und erstaunlicherweise auch zum Verlust der Fähigkeit, Metaphern zu verstehen.
7. Der alltägliche Sprachgebrauch ist reich an synästhetischen Metaphern, wie ‚schrille Farbe‘, oder ‚heiße Braut‘. Das wirft die Frage auf, was die dahinter liegende Verbindung - falls vorhanden - zwischen Metaphern und Synästhesie ist.
8. Synästhesien sind häufiger unter Künstlern, Poeten und allgemein kreativen Menschen anzutreffen.
9. Viele Synästheten berichten, dass sie auch eine Farbe sehen, wenn sie sich das entsprechende Graphem visuell vorstellen. Die Farben sind dann oft

lebendiger, als beim Anschauen des Graphems. Bei anderen Synästheten ist die Farbe eines vorgestellten Graphems dagegen weniger lebendig (Ramachandran/Hubbard, 2001b).

### 5.2.1 Sind Synästhesien perzeptuell oder kognitiv?

Lange Zeit galten Synästhesien als Phantasien, übertriebene bildhafte Vorstellungen, Assoziationen aus der Kindheit, oder als Folge von Drogenmissbrauch. Ramachandran und Hubbard (2001a & 2001b) gelang bei zwei Graphem-Farben Synästheten der experimentelle Nachweis, dass es sich um perzeptuelle Prozesse handelt und nicht um kognitive Assoziationen, oder eine lebhaftere Erinnerung. Dazu verwendeten sie verschiedene Versuchsanordnungen:

**POP-OUT** Synästhetische Farben können zu „*Pop-Out*“ Effekten führen. Dazu wurden den Synästheten sowie normalen Kontrollpersonen Anzeigen mit Graphemen präsentiert, wie z.B. in *Abb. 5.1* zu sehen ist. Unter den computergenerierten 2en befand sich eine versteckte Form (Quadrat, Rechteck, Dreieck, oder Raute), die aus gespiegelten 2en, also 5en gebildet war. Diese Displays wurden jeweils für eine Sekunde gezeigt, danach konnten sich die Versuchspersonen entscheiden, welche der vier möglichen Formen sie in dem Bild erkannt hatten. Die Synästheten erreichten Trefferquoten um die 80%, während die Kontroll-Personen im Durchschnitt auf 60% richtig erkannte Formen kamen. Die höhere Trefferquote der Synästheten zeigt, dass es ihnen leichter fällt, die Formen zu identifizieren, da sie sich in ihrer Wahrnehmung auch farblich abheben.



*Abb. 5.1: Display für Versuche zum Pop-Out Effekt(links) und wie es von einem Synästheten wahrgenommen wird*

**PERCEPTUAL GROUPING**

Getestet wurden die beiden Synästheten sowie 20 nicht-synästhetische Referenzpersonen. Ihnen wurden 7x5-Matrizen aus jeweils 35 Zeichen gezeigt (*Abb. 5.2a* und *5.2b*). Sie sollten angeben, ob sie die Zeichen in den Matrizen als horizontal oder vertikal gruppiert ansehen. Der Zweck dieses Tests war ihnen dabei nicht bekannt. Die Wahrscheinlichkeit in einer solchen Matrix die Zeichen entweder horizontal oder vertikal zu gruppieren liegt normalerweise bei ca. 50%. Die Grapheme waren allerdings so ausgewählt, dass in einer Richtung eine Ähnlichkeit der Formen bestand und in der anderen eine Ähnlichkeit der hervorgerufenen Farben.

3	8	3	8	3	8	3
7	0	7	0	7	0	7
3	8	3	8	3	8	3
7	0	7	0	7	0	7
3	8	3	8	3	8	3

3	8	3	8	3	8	3
7	0	7	0	7	0	7
3	8	3	8	3	8	3
7	0	7	0	7	0	7
3	8	3	8	3	8	3

*Abb. 5.2a (links): Schema einer Testmatrix für perceptual grouping*

*Abb. 5.2b (rechts): Testmatrix, wie sie die getesteten Synästheten sehen.*

Die Matrizen waren so gestaltet, dass die Synästheten entweder rote und grüne oder gelbe und blaue Wahrnehmungen hatten, also möglichst komplementäre Farben. Sie wurden nach einer Art Pseudo-Zufallsprinzip präsentiert, so dass alle möglichen Gruppierungsmöglichkeiten durchgespielt wurden. Jede der getesteten Personen bekam jede Tafel einmal zu sehen. So ergaben sich insgesamt 144 Tests. Das Ergebnis war, dass die Synästheten zu rund 90% nach denen von ihnen wahrgenommenen Farben gruppierten, während die normalen Testpersonen in der Regel nach der Form der Grapheme gruppierten.

„*Pop-out*“ und „*perceptual grouping*“ sind oft verwendete diagnostische Tests zur Ermittlung, ob es sich um perzeptuelle oder höhere kognitive Prozesse handelt.

**CROWDING** In einem weiteren Experiment zeigten sie, dass selbst ‚unsichtbare‘ Grapheme Farben erzeugen können. Einzelne Grapheme in der Peripherie präsentiert, können leicht identifiziert werden. Sind sie allerdings durch andere Zeichen flankiert, fällt es im Allgemeinen wesentlich schwerer, sie zu erkennen. Dieser Effekt wird als „*crowding*“ bezeichnet und ist nicht auf zu geringe Auflösung oder Unschärfe in der Peripherie zurückzuführen, denn die

Grapheme waren groß genug und konnten ohne ‚Begleiter‘ gut erkannt werden. Es zeigte sich, dass auch das flankierte Graphem die entsprechende Farbe bei den Synästheten hervorrief. Die Synästheten sagten, sie könnten den mittleren Buchstaben nicht sehen, aber es müsse ein ‚O‘ sein, da es blau erscheine. Dies zeigt, dass die sich erzeugten Farben auf perzeptuellen Stufen bilden und nicht auf kognitiven.

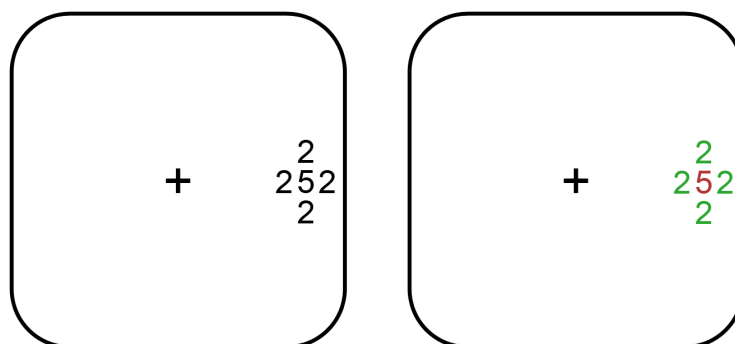


Abb. 5.3: Versuchsaufbau zu ‚crowding‘ (links) und wie ein Synästhet ihn sieht (rechts)

**PERIPHERAL VISION** Es wurde festgestellt, dass Grapheme, die mehr als 11° außerhalb der Fovea gezeigt und mit zunehmendem Winkel vergrößert wurden (Abb. 5.4), ihre Farbe verloren, auch wenn sie immer noch klar erkennbar waren.



Abb. 5.4: Illustration des Nachlassens synästhetischer Farben mit größer werdender Auslenkung

**ALTERNATING** Wenn zwei Grapheme alternierend präsentiert wurden, so wurden von den Synästheten auch die entsprechenden Farben abwechseln wahrgenommen, allerdings nur bis zu einer Frequenz von 4Hz. Ab höheren Frequenzen bis zu 10Hz wurden zwar die Grapheme noch erkannt, jedoch keine Farben mehr wahrgenommen. Bei einer dritten synästhetischen Person begannen die Farben bei einer sehr viel geringeren Frequenz von zwei- bis dreimal pro Sekunde zu alternieren.

**NUMERICAL CONCEPT** Römische Zahlen und Ansammlungen von Punkten zur visuellen Simultanerfassung (spontane Mengenerfassung, z.B. die sofortige Wahrnehmung von drei Äpfeln als ‚3 Äpfel‘, ohne sie abzählen zu müssen) lösten keine Farben aus. Folglich ist das visuelle Graphem entscheidend, nicht das Konzept der Zahl (Ausnahmen werden weiter unten besprochen). Taktile und akustische Zahlen waren ebenfalls wirkungslos, solange sich die getesteten Personen die Zahlen nicht vorstellten.

Dies war ein erstes starkes Indiz dafür, dass es sich (zumindest bei den getesteten Synästheten) um tatsächlich wahrgenommene Farben handelt und nicht etwa um Erinnerungen oder Assoziationen.

Zusammengenommen zeigen diese Ergebnisse, dass (zumindest bei den getesteten Synästheten) die Farbwahrnehmungen durch perzeptuelle Prozesse ausgelöst werden (Ramachandran/Hubbard, 2001a & 2001b).

### 5.2.2 Die cross-activation hypotheisis

Die Idee, Synästhesien seien das Ergebnis einer Form von kreuzenden Verbindungen im Gehirn, existiert seit über 100 Jahren, wurde jedoch bisher stets sehr vage formuliert und die anatomische Lokalisation nicht richtig erforscht. Es sollen nun konkret nachweisbare Vorschläge zur Lokalisation und dem Ausmaß von Kreuz-Aktivität mit den eingangs genannten Fakten in Verbindung gebracht werden, um daraus ein detailliertes Modell über das Entstehen und die neuronalen Mechanismen von Synästhesien zu formulieren. Die am häufigsten vorkommende Synästhesie ist die zwischen Graphemen und Farben. Die nachfolgenden Überlegungen beziehen sich alle auf diese Form, wobei diese Argumentationslinie auch für andere Formen gelten könnte (Ramachandran/Hubbard, 2001b).

Untersuchungen in der jüngeren Vergangenheit haben gezeigt, dass sowohl die Areale für die Farbwahrnehmung (V4; Lueck et al., 1989; Zeki/Marini, 1998; V8 Hadjikhani et al., 1998), als auch die für das Erkennen von Graphemen (Allison et al., 1994; Nobre et al., 1994; Pesenti et al., 2000) unmittelbar nebeneinander im *fusiform gyrus* (im unteren Bereich des Temporallappens) liegen. Dieses legt die Vermutung nahe, dass es bei Graphem-Farb-Synästheten

eine Form von Übersprechen zwischen diesen Regionen gibt, was umso wahrscheinlicher scheint, da diese Form die weit verbreitetste ist. Weiterhin ist bekannt, dass Synästhesien gehäuft in Familien auftreten, woraus sich ableiten lässt, dass eine Mutation eines einzelnen Gens die Ursache für eine exzessive Bildung von Kreuz-Verbindungen, oder aber ein fehlerhaftes Ausdünnen solcher Verbindungen, sein könnte. Das kindliche Gehirn enthält wesentlich mehr Verbindungen innerhalb und zwischen Arealen, als das eines Erwachsenen. Im Laufe der Entwicklung werden einige dieser Verbindungen ausgedünnt, andere bleiben erhalten. Ein genetischer Defekt, der den Mechanismus des Ausdünnens solcher Verbindungen einschränkt, könnte demnach der Grund für ausgeprägte Verbindungen zwischen Arealen im Erwachsenen-Alter bei Synästheten sein. Die tatsächlichen, konkreten Zuordnungen müssen jedoch erlernt sein, da kein Mensch mit dem Wissen über Farben, Buchstaben und Zahlen zur Welt kommt. Das übermäßige Vorhandensein von Kreuz-Verbindungen schafft lediglich die Voraussetzungen für derartige Verknüpfungen. Auf diese Weise ließe sich auch die große individuelle Streuung erklären, denn derselbe Buschstabe löst bei verschiedenen Synästheten unterschiedliche Farben aus (Ramachandran/Hubbard, 2001b).

Die *Cross-Activation Hypothesis* geht demnach von übermäßigen Verbindungen zwischen Hirnarealen aus, die in Folge einer genetischen Mutation durch fehlerhaftes Ausdünnen solcher Verbindungen beim Übergang ins Erwachsenenalter verursacht werden. Diese übermäßigen Verbindungen bilden die Voraussetzung für das kreuzmodale Erleben eines Synästheten.

Die *Cross-Activation Hypothesis* ist konform mit einer Reihe von Befunden und Fakten. Sie kann z.B. erklären, dass bei einigen der untersuchten Synästheten eine Farbwahrnehmung nur im zentralen Sehfeld induziert wird, da V4 (tertiäre Bereiche des visuellen Cortexes für Farbwahrnehmung) hauptsächlich das Sehzentrum repräsentiert. Die Beobachtung, dass bei ortsgleichen, alternierenden Graphemen ab einer Frequenz von über 4Hz keine Farben mehr induziert werden, deckt sich mit dem Befund, dass zentral-perzeptuelle Phänomene alle relativ langsam ablaufen. Ferner bietet sie eine Erklärung dafür, dass bei den getesteten Synästheten nur die tatsächliche Ziffer eine Farbe auslöst, nicht jedoch die entsprechende römische Zahl, denn im *fusiform gyrus* werden Grapheme repräsentiert, nicht deren Konzepte. Zuletzt kann die

Hypothese auch begründen, dass Synästheten meist mehrere verschiedene Synästhesien haben; Auch wenn nur ein einzelnes Gen an der fehlerhaften Ausdünnung beteiligt ist, so mag sich dies sehr unregelmäßig ausdrücken, was das Ausmaß und die anatomische Lokalisation angeht (Ramachandran/Hubbard, 2001b).

### 5.2.3 Top-Down-Verarbeitung bei Synästheten

Auch wenn es sich um perzeptuelle Vorgänge handelt, können Synästhesien trotzdem von Top-Down-Verarbeitung, wie z.B. Aufmerksamkeit beeinflusst werden. Dies lässt sich anhand von Experimenten mit Graphem-Farb-Synästheten demonstrieren.

In einem ersten Test präsentierte Ramachandran eine große 5, die aus kleinen 3en zusammengesetzt ist (Abb.5.5). Die zwei untersuchten Synästheten konnten dabei zwischen der Farbe der 5 und der Farbe für die 3 hin- und herwechseln, je nachdem, auf welche Zeichen sie ihre Aufmerksamkeit lenkten. Das zeigt, dass der Effekt durch Top-Down-Verarbeitung, wie in diesem Fall durch Aufmerksamkeit, beeinflusst werden kann.

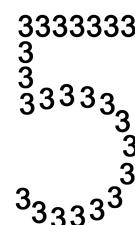


Abb. 5.5: Hierarchisches Gebilde zur Demonstration von Top-Down Verarbeitung bei Synästhesien.

In einem weiteren Versuch bekamen dieselben Personen Darstellungen wie z.B. ‚THE CAT‘ (Abb.5.6) gezeigt. Sie gaben an, die jeweils „richtigen“ Farben für den Buchstaben A und H zu sehen, obwohl die Zeichen für beide identisch waren. Das heißt, auch wenn die Formen benötigt werden um Farben auszulösen, ist die Art, wie diese klassifiziert werden, wichtig für die ausgelöste Farbe.

**THE CAT**

Abb. 5.6: Mehrdeutiger Stimulus zur Demonstration von Top-Down Verarbeitung bei Synästhesien (‚THE CAT‘).

**IV**

Abb. 5.7: Mehrdeutiger Stimulus zur Demonstration von Top-Down Verarbeitung bei Synästhesien (‚römische Vier‘).

Einem dritten Synästheten wurde das Zeichen einer römischen Vier (*Abb. 5.7*) gezeigt. Dieser berichtete, Farben zu sehen, wenn er die Grapheme als Buchstaben las, keine Farben jedoch, wenn er sie als Zahl interpretierte.

Zusammengenommen zeigen diese Ergebnisse, dass Top-Down-Verarbeitung starken Einfluss auf Synästhesien haben kann. Das heißt jedoch nicht, dass sie ein konzeptuelles Phänomen sind, vielmehr untermauern diese Beobachtungen, dass kognitive Verarbeitung sensorische Prozesse beeinflussen können (Ramachandran/Hubbard, 2001b).

### *„Unitization“ beeinflusst Synästhesie*

Um der Idee von Top-Down-Verarbeitung weiter auf den Grund zu gehen, führten Ramachandran und Hubbard (2001) einen Versuch mit einem Synästheten durch, in dem sie den *Missing Letter Effect* nutzten. Sie präsentierten dazu Sätze wie *„Finished files are the result of years of scientific study combined with the experience of years“* und forderten dazu auf, die darin enthaltenen f's zu zählen. Nicht-Synästheten kommen in der Regel auf drei, da sie die f's in Wörtern wie ‚of‘ übersehen. Das passiert, da häufig gebrauchte Funktionswörter als eigene lexikalische Einheit wahrgenommen werden. Ähnliches zeigte sich auch bei einer getesteten Synästhetin, die zunächst drei rote Buchstaben angab. Erst bei genauerer Prüfung, färbten sich sechs Buchstaben rot, was impliziert, dass *„unitization“* das Induzieren synästhetischer Farben bei ihr verhinderte (Ramachandran/Hubbard, 2001).

### *Synästhesie und visuelle Vorstellung*

Es hat sich gezeigt, dass bei vielen Synästheten Grapheme, die sie sich vorstellen, stärkere Farben induzieren, als real gesehene, auch wenn es einzelnen Ausnahmen davon gibt. Es könnte sein, dass das Einlassen auf eine mentale Vorstellung teilweise sowohl Kategorie-spezifische Regionen der Objekterkennung, als auch frühe visuelle Pfade aktiviert. In Abhängigkeit vom Ort der Kreuz-Aktivierung, dem Ausmaß und Ort von Top-Down-Aktivität und dem Ausmaß, wie stark tatsächliche Bottom-Up-Verarbeitung von der Retina diesen widersprechen, könnte es verschiedene Grade geben, inwieweit

Synästhesien durch Vorstellungen bei verschiedenen Individuen ausgelöst werden.

Um diesen Sachverhalt genauer zu erforschen, führten Ramachandran et al. Versuche mit Synästheten durch, in den „*Perky Effect*“ nutzten. Perky (1910) entwickelte eine Methode zur Untersuchung des schwer zu fassbaren Übergangs zwischen Vorstellung und Wahrnehmung. Im ursprünglichen Versuchsaufbau betrachten Personen eine weiße Leinwand und werden gebeten, sich einen bestimmten Gegenstand vorzustellen. Dieser Gegenstand wird dann ohne deren Wissen auf die Leinwand, in langsam steigender Intensität, rückprojiziert, so dass er zu Beginn nicht wahrnehmbar ist, mit der Zeit aber kontinuierlich deutlicher wird. Es hat sich in vielen solchen Versuchen gezeigt, dass Testpersonen die Projektion, selbst bei deutlicher Überschreitung der Reizschwelle, nicht wahrnehmen, da ihre sensorische Sensitivität durch die bildliche Vorstellung abnahm. Einen ähnlichen Aufbau führten Ramachandran et al. mit Graphemen durch. Dabei stellte sich heraus, dass die Reizschwelle der getesteten Synästheten größer war, als bei Nicht-Synästheten. Das heißt, ihre Vorstellung hemmte die Wahrnehmung noch mehr, als dies bei Normalpersonen der Fall war. Diese Befunde unterstützen die Meinung, dass synästhetische Farben in Folge von Top-Down-Aktivierung, erzeugt durch visuelle Vorstellung, wahrgenommen werden können (Ramachandran/Hubbard/Butcher, 2004).

#### 5.2.4 Higher und lower synesthets

Die bisher aufgeführten Untersuchungsergebnisse stimmen für die beiden getesteten Synästheten. Es gibt daneben allerdings auch solche, bei denen auch römische Ziffern sowie Anhäufungen von Punkten Farbreize auslösen, folglich das Konzept von Zahlen. Interessanterweise gibt es auch Synästheten, bei denen Wochentage, oder die Monate des Jahres farbig sind. Es könnte daher sein, dass es eine Region im Gehirn gibt, die abstrakte Ziffernfolgen und die Wertigkeit von Zahlen kodiert. Diese wäre dann bei jenen Synästheten vielleicht mit Regionen der Farb-Kodierung kreuzverbunden.

Als Areal für abstrakte Ziffernfolgen und Wertigkeiten kommt der Gyrus angularis in Betracht, von dem eine Beteiligung an numerischen Operationen bekannt ist. Schädigungen am Gyrus angularis führen in der Regel zu

Dyskalkulie. Interessanterweise liegen die in der Farbkodier-Hierarchie nachfolgenden Areale direkt daneben, im Gyrus temporalis inferior. Es kann somit spekuliert werden, dass bei den beschriebenen Synästheten diese Regionen kreuzverbunden sind. In der Tat könnten die Ebenen, auf denen sich die Gen-Mutation ausdrückt, ausschlaggebend für die Ebenen sein, auf denen sich übermäßige Kreuz-Verbindungen bilden. Es sei in diesem Zusammenhang lose von „*higher*“ und „*lower synesthets*“ gesprochen. Demnach wären „*higher synesthets*“ solche, bei denen es vorwiegend Kreuzverbindungen auf höheren Stufen der Wahrnehmung gibt und daher Synästhesien bei ihnen auf diesen Ebenen stattfinden. Das könnte sich z.B. darin ausdrücken, dass sie keine Farbwahrnehmungen haben, wenn Grapheme durch Störreize teilweise verdeckt sind. Allgemein bräuchten sie mehr fokussierte Aufmerksamkeit, um synästhetisches Erleben auszulösen. Zu ähnlichen Einteilungen kommen auch Grossenbacher und Lovelance (2001). Sie sprechen von „*perceptual*“ und „*conceptual synesthesia*“. Diese Aufteilungen dienen derweil eher theoretischen Überlegungen, in der Praxis sind bei genauerer Untersuchung eher diverse Mischformen zu erwarten (Ramachandran/Hubbard/Butcher, 2004).

#### 5.2.5 Synästhesie und Metaphern

Verwandt mit der Frage, ob Synästhesien ein perzeptuelles oder konzeptuelles Phänomen sind, ist das Verhältnis zwischen Synästhesien und Metaphern. Traditionell gibt es zwei grundlegende Ansichten:

- (1) Synästhesien sind kein echtes perzeptuelles Phänomen, sie sind bloß metaphorisch, und
  - (2) Die meisten Menschen, die vorgeben Synästhesien zu haben, sind metaphorisch. Nur wenige von ihnen haben echte (perzeptuelle) Synästhesien.
- Ramachandran et al. halten beide Ansichten für falsch und vertreten stattdessen die Ansicht, dass Synästhesien relativ verbreitet sind und ihnen ähnlichen neuronalen Mechanismen unterliegen, wie Metaphern. Die Aussage, Synästhesien sind bloß metaphorisch hilft nicht, das Phänomen zu erklären, da nicht bekannt ist, welche Prozesse Metaphern zu Grunde liegen. Es ist der Versuch ein Mysterium, mit einem anderen Mysterium zu erklären. Dies ist ein Ansatz, der zu keinem wissenschaftlichen Ergebnis kommen kann.

Der bisherigen Stand des Wissens legt nahe, dass das Problem stattdessen auf den Kopf gestellt und angenommen werden kann, dass Untersuchungen von Synästhesien helfen können, die schwierige Frage zu beantworten, wie Metaphern im Gehirn repräsentiert werden (Ramachandran/Hubbard/Butcher, 2004).

#### 5.2.6 Künstler, Poeten und Synästhesie

Synästhesien treten häufiger unter Künstlern, Poeten und Schriftstellern auf. Domino (1989) untersuchte 358 Studenten der bildenden Künste, von denen 23% angaben, synästhetische Erfahrungen zu machen. 61 dieser Synästheten verglich er mit 61 Nicht-Synästheten gleichen Geschlechts, Studiengang, Semester, und sprachlicher Intelligenz. Er fand heraus, dass sie Synästheten in allen getesteten kreativen Aufgaben überlegen waren.

Eines haben solche Gruppen von Menschen gemeinsam: Die Fähigkeit, zwei scheinbar nicht verwandte Bereiche zu verknüpfen um ein verborgene, tiefe Ähnlichkeit herauszustellen. Ein Beispiel einer literarischen Metapher ist „It is the east and Juliet ist the sun“ aus Shakespeares *Romeo and Juliet*. Was sind die zu Grunde liegenden Mechanismen dafür, dass üblicherweise jeder diese Zeilen als metaphorisches Konstrukt versteht und sie nicht wörtlich nimmt?

Es wurde oft vorgeschlagen, dass Perzepte im Gehirn auf die gleiche Weise repräsentiert werden, wie Konzepte. Ein Beispiel für ein relativ abstraktes Konzept ist eine Zahl. Es ist bekannt, dass daran gewisse Areale beteiligt sind, nämlich der Gyrus angularis sowie der Gyrus fusiform. Im Fall von Zahlen ist die Unterscheidung zwischen Konzept und perzeptuellem Symbol klar und eindeutig, in vielen anderen Fällen ist es dies jedoch nicht. Es gibt jedoch Hinweise, dass bestimmte Konzept-Klassen in Regionen repräsentiert werden, die unmittelbar neben denjenigen perzeptuellen Regionen liegen, die mit diesen Konzepten in Verbindung stehen. Wenn dem so ist, können sich Metaphern als das Involvieren von Kreuzaktivität zwischen konzeptuellen Hirnarealen vorgestellt werden, analog zu perzeptueller Kreuzaktivität bei Synästhesien. Dieses könnte das größere Vorkommen von Synästheten unter Künstlern erklären.

Wenn Kreuz-Verbindungen selektiv den Gyrus angularis oder den Gyrus fusiform betreffen, könnte es zu perzeptuellen Synästhesien kommen. Wenn sie eher diffus verteilt sind, könnte dies ein eher allgemein Kreuz-vernetztes Gehirn zur

Folge haben, welches den Hang und die Fähigkeit besitzt, auf kreative Weise Konzepte miteinander zu verbinden (Ramachandran/Hubbard/Butcher, 2004).

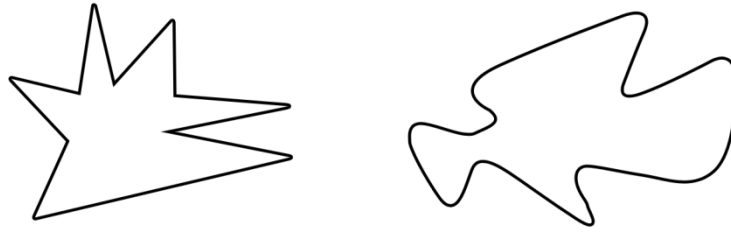
### 5.2.7 Der Gyrus angularis und synästhetische Metaphern

Es ist bekannt, dass der Gyrus angularis neben seiner Rolle bei abstrakten numerischen Operationen, auch an der Bildung von kreuzmodaler Assoziation beteiligt ist. Dies lässt die Annahme zu, dass er auch maßgeblich an der Bildung von Metaphern beteiligt ist, besonders von kreuzmodalen Metaphern. Metaphern sind –entgegen landläufiger Meinungen– nicht willkürlich! Das zeigen Untersuchungen von Lakoff und Johnson (1980). Sie stellen heraus, dass sich viele Metaphern auf den Körper beziehen und noch öfter intersensorischen oder synästhetischen Charakter haben. Ihre Untersuchungen zeigten darüber hinaus, dass es Präferenzen in der Richtung von kreuzmodalen Metaphern gibt (z.B. oft vom Gehörten zum Gesehenen: „schreiende Farbe“). Die präferierten Richtungen decken sich mit dem Vorkommen verschiedener Synästhesien. Auch hier ist das Sehen meist die angesprochene Modalität und nicht die auslösende (Heyrman, 2005). Ramachandran et al. schlagen vor, dass die Anatomie des Gehirns gewisse Kreuz-Verbindungen zulässt, gewisse andere dagegen nicht (Ramachandran/Hubbard/Butcher, 2004; Heyrman, 2005).

### *Synästhesie und die Evolution der Sprache*

Die Frage wie es zur Entwicklung von Sprache kam beschäftigt die Forschung seit langem und konnte bisher nicht abschließend beantwortet werden. Denkbar ist ein Zusammenhang mit synästhetischen Prozessen. Dazu seien einige Ideen miteinander verbunden:

Ein erster betrifft Stimuli wie den in Abb.5.8. Sie sind Bestandteil vieler Untersuchungen zu visuellen Prozessen in der Sprache. Wenn Testpersonen diese Bilder präsentiert werden und ihnen gesagt wird, einer der beiden Formen sei in einer Mars-Sprache ein ‚Kiki‘ und die andere ein ‚Bouba‘, sie sollten tippen, welche Form welches ist, so geben 95% an, die eckige Form sei ‚Kiki‘ und die runde sei ‚Bouba‘, auch wenn die die Bilder nie zuvor gesehen haben.



*Abb.5.8: Visuelle Stimuli zur Demonstration von ‚Kiki‘ und ‚Bouba‘*

Der Grund liegt darin, dass die zackige Formen und scharfen Richtungswechsel der rechten Abbildung, die scharfe Phonetik des Wortes ‚Kiki‘ widerspiegeln, ebenso wie die scharfe Bewegung der Zunge am Gaumen beim Aussprechen. Daraus lässt sich ableiten, dass es natürliche Regeln dafür geben könnte, wie Klänge Objekten zugeordnet werden.

Des Weiteren könnte eine Art sensorisch-motorische Synästhesie zur Entwicklung einer Proto-Sprache beigetragen haben. Ein plastisches Beispiel ist Tanz, wo Klänge in rhythmische Bewegungen überführt werden. Diese Synästhesie könnte demzufolge aus Kreuz-Verbindungen zwischen sensorischen Arealen (z.B. akustisch) und motorischen Arealen verursacht werden. Somit gäbe es einen natürlichen Hang bestimmte Klangkonturen bestimmten Aussprachen zuzuordnen.

Zusammengenommen liefern diese Ideen einen möglichen Ansatz zur Entwicklung der Sprache. Nach diesem gedanklichen Konstrukt werden die Repräsentationen bestimmter Lippen und Zungenbewegungen in motorischen Arealen in nicht-zufälliger Weise auf bestimmte Klang-Repräsentationen in akustischen Arealen abgebildet. Diese wiederum sind auf nicht-zufällige Art mit visuellen Repräsentationen von Objekten verknüpft (Ramachandran/Hubbard, 2001b).

### 5.3 Zusammenfassung

Synästhesien unterliegen in jüngerer Vergangenheit verstärktem Forschungsinteresse. Dies hängt auch damit zusammen, dass sich von einem besseren Verstehen der ihnen zu Grunde liegenden neuronalen Mechanismen Erkenntnisse über allgemeine kognitive Prozesse erhofft werden, insbesondere über menschliche Fähigkeiten zu Kreativität, Metaphern und künstlerischem Schaffen. Untersuchungen von Grapheme-Farben-Synästheten ergaben, dass Synästhesien (zumindest bei den getesteten Personen) ein perzeptuelles Phänomen sind und nicht ein Effekt, der auf Erinnerungs-Assoziationen aus der Kindheit oder einer vagen, metaphorischen Sprache basiert. Es konnten verschiedene Typen von Graphem-Farben Synästheten identifiziert werden. Die Thesen von Ramachandran et al. (2001a, 2001b, 2004) besagen, dass diese von einer Hyperkonnektivität zwischen Farben- und Nummern-Arealen auf verschiedenen Stufen der Verarbeitung verursacht wird. ‚Niedrige‘ Synästheten könnten eine Hyperkonnektivität innerhalb des „*fusiform gyrus*“ haben, ‚höhere‘ Synästheten dagegen eine Kreuz-Aktivierung im Gyrus angularis. Diese Hyperkonnektivität könnte durch einen genetischen Defekt verursacht werden, der eine fehlerhafte Trennung von Verbindungen zwischen Hirnarealen zur Folge hat. Wenn Kreuz-Verbindungen zwischen perzeptuellen Arealen der Auslöser für Synästhesien sind, so könnten auf ähnliche Weise Kreuz-Verbindungen zwischen konzeptuellen Arealen die Grundlage für Metaphern und kreatives Schaffen sein. Die Hypothesen von Ramachandran et al. sind spekulativ und bilden einen Ausgangspunkt für künftige Forschung.

## 6 MICHEL CHIONS AUDIO-VISION

Der französische Komponist, Filmemacher, Kritiker und Dozent Michel Chion setzt sich in seinem Buch „Audio-Vision“ ausführlich mit dem Thema audiovisueller Interaktionen auseinander und hat aus medientheoretischer Sicht viele Aspekte des Zusammenspiels akustischer und visueller Wahrnehmung im Hinblick auf die Rezeption von audiovisuellen Medien erläutert und analysiert. Er entwickelte dabei eigene Thesen, Begriffe und Definitionen, um das Zusammenspiel von Akustischem und Visuellem im Film zu beschreiben. Sein Blick ist dabei auf die medialen Repräsentanten der sensorischen Reize gerichtet, das heißt er trifft Aussagen über ‚den Ton‘ und ‚das Bild‘ und macht sie damit zu agierenden und interagierenden Subjekten. Chion ist klar, dass die Verarbeitung und Interpretation von Sinnesreizen in der Wahrnehmung des Menschen stattfindet. Daher kann diese Sichtweise einerseits als Vereinfachung angesehen werden, andererseits ist dies auch die Perspektive eines Medienschaffenden, der in seinem Alltag gestalterisch mit Ton und Bild umgeht, wie ein Handwerker mit Werkstücken. Ausgangspunkt seiner Thesen ist das Kino und der Spielfilm, diese sind jedoch durchaus auf andere audiovisuelle Medien übertragbar (Chion, 1994).

In diesem Kapitel wird eine Auswahl seiner Theorien, Analysen und Thesen vorgestellt, die im Hinblick auf die Fragestellung dieser Diplomarbeit relevant sind. Im Wesentlichen sind dies allgemeine Thesen zum Zusammenwirken von Ton und Bild, die in den ersten beiden Kapiteln von „Audio-Vision“ behandelt werden. Darauf aufbauend entwickelte Chion komplexere Zusammenhänge, die den semantischen, kulturellen und emotionalen Gehalt von akustischen und visuellen Reizen mit einbeziehen. Diese weiterführenden Gedanken werden hier nicht erläutert, da sie über die Fragestellung dieser Arbeit hinaus gehen.

Seine theoretischen Erkenntnisse leitet Chion zum größten Teil aus seinen Tätigkeiten als Komponist, Regisseur und Dozent für Audiovisuelle Beziehungen an der Universität Paris her. Er gibt selbst an, viele seiner Gedanken seien auch durch Diskussionen mit den Studenten an diversen Filmhochschulen in Frankreich und Belgien inspiriert. Ferner bezieht er sich u.a. auf Pierre Schaeffer, den Begründer der *musique concrète*, dessen Assistent er 1970 war.

Vor diesem Hintergrund erweist sich ein Vergleich mit kognitionswissenschaftlichen Studien als sehr interessant.

### 6.1 The audiovisual contract

Der erste Teil des Buches trägt den Titel „*The Audiovisual Contract*“, in dem die Beziehung zwischen Ton und Bild als eine Art Abkommen beschrieben wird, welches das Gegenteil einer natürlichen Beziehung ist, die auf einer bereits vorhandenen Harmonie der Sinne beruht. Ferner drückt Chion durch diese Bezeichnung aus, dass Ton und Bild sich in der Wahrnehmung gegenseitig beeinflussen und daher nicht getrennt voneinander betrachtet werden können, wenn audiovisuelle Medien hinsichtlich ihrer Wirkung auf den Rezipienten analysiert werden. Die Kapitel dieses ersten Teils führen eine Reihe möglicher Antworten auf, nach welchen Gesetzmäßigkeiten und Regeln das audiovisuelle Abkommen in Medien funktioniert (Chion, 1994, S.XXVI).

Einige dieser Gesetzmäßigkeiten werden nun näher vorgestellt.

#### *Added value*

Die Bildebene kann durch Klang (oder auch durch grafischen Text) in seinem Informationsgehalt und seiner Aussagekraft angereichert werden. Es wird der Eindruck erzeugt, der Gehalt entstamme „natürlicherweise“ dem Gesehenen und sei darin bereits vollständig enthalten. *Added value* bezeichnet demnach Effekte, die den Rezipienten -fälschlicherweise- glaubend machen, der Ton sei nicht nötig und dupliziere bloß den Eindruck und Gehalt des Bildes. Dabei handelt es sich tatsächlich um übersummativ Effekte, da der Gehalt des zusammen Dargebotenen größer ist, als die Summe der einzelnen Ebenen. In besonderer Weise arbeitet *added value*, wenn Ton und Bild synchron zueinander sind. Dies wird besonders deutlich an Soundeffekten im Kino, Geräusche die Schläge, Stöße, Explosionen usw. begleiten. Hier wird die Illusion erzeugt, Bild und Ton stünden in einer notwendigen, natürlichen Beziehung zueinander, obwohl diese Geräusche nachträglich hinzugefügt wurden (Chion, 1994, S.5).

### *Added value by music*

Musik ist eine von vielen Möglichkeiten der Gestaltung des akustischen Raums im Film. Unterschieden werden kann zwischen *empathetic* und *anempathetic music*. Diese bezeichnen zwei Arten, wie Musik mit den dargestellten Emotionen im Bild zusammenspielt. Ein empathischer Effekt wird demnach erzielt, wenn die Musik sich auf den emotionalen Gehalt einlässt und ihn mitträgt. *Empathic music* bedient sich kultureller Codes für den Ausdruck von Gefühlen wie Traurigkeit, Freude oder Gerührtheit. *Anempathetic music* ist dagegen solche, die isoliert und emotional unberührt neben den visuell dargestellten Gefühlen steht. Damit erzeugt sie den Eindruck, kalt und verschlossen gegenüber den Geschehnissen zu sein. Dieser Kontrast verstärkt die emotionale Dichte einer Szene. Durch den erzeugten Gegensatz werden die einzelnen Pole intensiver wahrnehmbar, als bei empathischer Musik (Chion, 1994, S.8).

### *Temporäre Aspekte der Wahrnehmung*

Visuelle und akustische Wahrnehmung sind von grundlegend verschiedener Natur, auch wenn dies meist nicht bewusst ist. Gerade in ihrer Beziehung zu Bewegung und Statik unterscheiden sie sich deutlich. Das Bild kann statisch sein, was faktisch in Filmen auch häufig der Fall ist. Oft sind nur einzelne Teile (Objekte/Personen) einer Szene in Bewegung, während der Rest unbewegt ist. Ton hingegen setzt Bewegung zwingend voraus, was in seiner Natur liegt und mit den physikalischen Eigenschaften des Schalls zu tun hat (Chion, 1994, S.9f.). Entsprechend dieser Prinzipien unterscheiden sich die Modalitäten der menschlichen Wahrnehmung in ihrem räumlichen und zeitlichen Auflösungsvermögen. Die akustische Wahrnehmung ist in temporären Aufgaben besser ausgestattet als die visuelle, wohingegen diese dem Hören in räumlichen Aspekten überlegen ist (Chion, 1994, S.10f.).

Es gibt folgende Arten, wie Klang die Bewegungswahrnehmung im Visuellen verändert:

1. Klang kann die Wahrnehmung schneller Vorgänge verbessern. Vor allem in sehr schnellen, actionreichen Szenen ist es fast immer der Ton, anhand dessen der Rezipient die zeitliche Orientierung behält. Kampfszenen oder rasante Autofahrten kann das Auge allein nicht erfassen. Geräusche und

Klangeffekte helfen hier, schnelle Bewegungsabläufe, Schläge und Stöße besser verfolgen zu können.

2. Ton kann unter gewissen Voraussetzungen visuelle Wahrnehmungen erzeugen, die das Bild nicht zeigt, bzw. nur andeutet. Es ist möglich einen Schlag oder Stoß zu sehen, obwohl keine wirkliche Berührung stattfindet. Ferner kann eine visuelle Bewegung durch Klang erzeugt werden. Ein Beispiel hierfür ist eine Szene aus der Star Wars Episode „Das Imperium schlägt zurück“: Das Bild einer geöffneten Tür wurde hart an das Bild derselben Tür im geschlossenen Zustand geschnitten. Zusammen mit einem Soundeffekt wurde der Eindruck erzeugt, die Tür würde in einer Bewegung geschlossen.

(Chion, 1994, S.11f.)

Klang hat darüber hinaus auch wesentlichen Einfluss auf die visuelle Zeitwahrnehmung:

1. Klang animiert das Bild. Je nach Gestaltung des Tons kann dieselbe Szene als zeitlich vage, variabel und ausgedehnt, oder im Gegensatz dazu als exakt, detailliert und unmittelbar empfunden werden.
2. Ton linearisiert eine Szene. Wird der Ton über Bildschnitte hinweg kontinuierlich fortgeführt, so wird diese Szene als zeitlich und möglicherweise auch als räumlich zusammenhängend wahrgenommen. Denn der sogenannte Atmo-Ton, die natürliche Geräuschkulisse einer Szene, zeigt an, in welcher Umgebung sie stattfindet.
3. Ton vektorisiert oder dramatisiert eine Szene: Er orientiert sie Richtung Zukunft und kann etwas Bevorstehendes ankündigen. Dieses kann an einem Beispiel verdeutlicht werden: Viele relativ statische Film-Einstellungen könnten rückwärts abgespielt werden, ohne dass der Zuschauer dies bemerkte. Rückwärts abgespielter Ton fällt im Vergleich dazu stärker auf und würde viel eher vom Rezipienten bemerkt werden.

(Chion, 1994, S.13f.).

Die Wirkweise der beschriebenen Effekte ist an Bedingungen gebunden, wie Bild und Ton jeweils beschaffen sind und kombiniert werden. Was die Beschaffenheit des Bildes angeht, können zwei grundlegende Szenarien unterscheiden werden: Erstens, das Bild besitzt keine eigene zeitliche Animation oder Vektorisierung,

ist also mehr oder weniger statisch. Das heißt, es enthält keine zielgerichtete, sondern nur fluktuierende Bewegung. Das kann in Form von rinnendem Wasser, oder im Wind wiegenden Blättern sein. In einem solchen Fall übernimmt die Szene die zeitliche Struktur und Beschaffenheit des Tons. Zweitens, das Bild besitzt eine eigene zeitliche Animation oder Vektorisierung in Form von bewegten Objekten oder Personen. Dann kommt es zu einem Zusammenspiel der Zeitlichkeit von Ton und Bild. Diese können dann zusammenspielen und sich so unterstützen oder leicht gegeneinander arbeiten.

Bei der Beschaffenheit des Tons lassen sich mehrere Eigenschaften nennen, die seine Zeitlichkeit und damit den Einfluss auf die der Szene ausmachen:

1. Wie gedehnt oder anhaltend ist der Klang? Ein kontinuierlicher Klang wirkt weniger animierend, als ein fluktuierender.
2. Wie vorhersagbar ist der Klang? Vorhersagbare Klänge, insbesondere solche durch musikalische Form (typische, kulturell erlernte Abfolgen), erzeugen weniger zeitliche Animation, als plötzliche und unerwartete Veränderungen und Wendungen.
3. Das Tempo. Dabei ist das tatsächliche, messbare Tempo (Schläge pro Minute) weniger ausschlaggebend, als die Art des Rhythmus, die Dynamik, die Häufigkeit harmonischer Wechsel und die Melodieführung, Faktoren, die das empfundene Tempo bestimmen.
4. Klang-Definition. Hat ein Klang viele oder starke Anteile hoher Frequenzen, erzeugt er eine gesteigerte Aufmerksamkeit, Wachsamkeit und Präsenz beim Rezipienten.

Die Zeitlichkeit einer Szene wird darüber hinaus bestimmt durch die Gestaltung der Verknüpfung von der Ton- mit der Bild-Ebene, sowie das Vorhandensein von Synchronitäts-Punkten zwischen diesen. Entscheidend scheint in jedem Fall das Spiel mit den Erwartungen der Zuschauer zu sein (Chion, 1994, S.14f.).

## 6.2 The three listening modes

Wenn Personen gefragt werden, was sie gehört haben, so sind die Antworten von einer auffallenden Heterogenität, was die Hör-Ebene angeht, auf die sie sich jeweils beziehen. Das liegt auch daran, dass es mindestens drei Arten oder

Modi des Hörens gibt, die sich auf jeweils unterschiedliche Merkmale beziehen. Dies sind *Causal*, *semantic* und *reduced listening* (Chion, 1994, S.25).

**KAUSALES HÖREN** ist die gebräuchlichste Form und besteht darin, Informationen über die Quelle bzw. Ursache eines Schallereignisses zu erhalten. Wenn die Schallquelle sichtbar ist, kann Klang zusätzliche, wichtige Informationen über das Objekt liefern. Wenn die Schallquelle jedoch nicht sichtbar ist fällt es in der Regel schwer, allein anhand des akustischen Signals das Objekt eindeutig zu identifizieren. Dies zeigt, dass das kausale Hören nicht so präzise und verlässlich für die Objektidentifikation ist, wie es den Anschein hat. Im Gegenteil, es ist der am leichtesten zu manipulierende Modus.

Kausales Hören kann auf verschiedenen Ebenen stattfinden. In manchen Fällen ist es tatsächlich möglich, das schallverursachende Objekt individuell zu identifizieren: Die menschliche Stimme ist wahrscheinlich das einzige Klang-Phänomen, das einer einzigen Person zugeordnet werden kann und seine Identifikation erlaubt. In den meisten anderen Fällen werden Schallereignisse allgemeinen Klassen zugeordnet. Solche Klassen können z.B. Attribute wie „mechanisch“ oder „tierisch“ sein. Ein Schallereignis hat indes immer mehrere Ursachen gleichzeitig. Das Schreiben mit einem Filzstift erzeugt z.B. ein Geräusch durch die Miene, eines durch das Papier, eines durch die Hand des Schreibenden und evtl. durch den Schreibenden selbst (Atmung u.Ä.). In allen Fällen hat kausales Hören immer etwas mit Erfahrung und erlerntem Wissen zu tun (Chion, 1994, S.25ff.).

**SEMANTISCHES HÖREN** bezeichnet das Erkennen und Verstehen gesprochener Sprache. Sprachwahrnehmung ist ein sehr komplexer Vorgang, der intensiv erforscht wurde und wird. Eine wesentliche Erkenntnis vieler Studien ist die, dass die Spracherkennung ein differentialer Prozess ist, in dem nicht bloß die akustischen Eigenschaften eines Phonems analysiert werden, sondern dieses als Teil eines Systems von Gegensätzen und Differenzen verarbeitet wird. Das ist der Grund, warum Menschen in der Lage sind sowohl Dialekte, als auch verschiedene Stimmen und Aussprachen zu verstehen (Chion, 1994, S.28).

**REDUZIERTES HÖREN** ist der Teil der akustischen Wahrnehmung, der am unbewusstesten abläuft. Es handelt sich um das Wahrnehmen der intrinsischen Klangeigenschaften eines Schallereignisses, ohne Berücksichtigung oder Einbeziehung der Ursache bzw. der Schallquelle. Pierre Schaeffer (1967) entwickelte das reduzierte Hören als bewusste Technik der Wahrnehmung, um die Eigenschaften von Klängen besser analysieren zu können. Denn der emotionale, physikalische und ästhetische Wert eines Geräuschs hängt nicht allein von seiner Quelle ab, sondern liegt auch in seiner Stimmung, Textur und seinen ureigenen qualitativen Eigenschaften. Es ist jedoch auffallend, dass es kaum sprachliche Mittel und Ausdrücke gibt, um Klänge auf diese Art zu klassifizieren. Meist werden eher Beschreibungen verwendet, die auf kausales Hören zurückgehen (Chion, 1994, S.29).

### *Aktive und passive Wahrnehmung*

Die klare Trennung zwischen den drei genannten Hör-Modi ist im Hinblick auf die Analyse von Klangeigenschaften sinnvoll. Sie ist jedoch eine theoretische, da sie in der Praxis überlappen und miteinander verwoben sind. Das bedeutet gleichzeitig, dass neben der aufmerksamen, aktiven Wahrnehmung auch immer die passive, unbewusste eine Rolle spielt, sowohl beim Hören, als auch beim Sehen.

Im Kino ist das Schauen gleich dem visuellen Erleben, da der Blickwinkel auf eine Szene durch die Kameraeinstellung vorgegeben ist und durch die Leinwand begrenzt wird (Anm.: das ist übertragbar auch auf andere audiovisuelle Medien). Dieser begrenzte Blick übernimmt somit stellvertretend die Funktion der Aufmerksamkeit. Für das Hören trifft dies weniger zu, da der akustische Raum umgebend und umfassend ist. Die Eigenschaft des Schalls sich in Räumen auszubreiten und die fehlende Fähigkeit, die Ohren verschließen zu können, verleihen dem Ton stets etwas Überwältigendes und Überraschendes. Der Mensch ist kaum in der Lage, sich einem Schallereignis zu entziehen. Deshalb hat Schall einen starken Einfluss auf die Wahrnehmung und kann sie leichter durchdringen und kurzschließen, als das Visuelle. Der Ton bietet daher ein starkes Mittel der affektiven und semantischen Manipulation. Einerseits, weil er direkt physisch wirken kann, zum anderen, weil er die Wahrnehmung

beeinflusst, wie durch das Konzept des *added value* gezeigt wurde (Chion, 1994, S.33f.).

### 6.3 Analyse und Vergleich mit wissenschaftlichen Studien

An vielen Betrachtungen und Thesen Chions lassen sich mehr oder weniger deutliche Parallelen zu wahrnehmungspsychologischen und kognitionswissenschaftlichen Studien feststellen. Einige allgemeine Prinzipien der Wahrnehmung, die bei Chion zur Sprache kommen, waren zum Zeitpunkt der Veröffentlichung von „Audio-Vision“ (Erstausgabe im Französischen 1990) bereits Stand der Wissenschaft und öffentlich bekannt. Manche Schlüsse, die er im Bezug auf das Kino zog, wurden allerdings erst nach seiner Veröffentlichung durch Kognitionswissenschaftler erforscht, dies betrifft insbesondere die in Kapitel 4 dieser Diplomarbeit aufgeführten Untersuchungen zu multisensorischen Prozessen.

Im Folgenden werden einige der erläuterten Thesen Chions mit den behandelten wissenschaftlichen Studien verglichen. Die Reihenfolge ist wie in den Kapiteln 6.1 und 6.2.

Chions Begriffe *audiovisual contract* und *added value* stehen stellvertretend für die allgemeine Sicht, dass Bild und Ton in Medien eine kommunikative Einheit bilden, die zusammen dargeboten eine andere Wirkung erzielen, als dies jeweils einzeln der Fall wäre.

Diese Ansicht deckt sich mit denen moderner Forschungen. Jede in Kapitel 4 besprochene Studie bestätigt, dass in der Wahrnehmung des Menschen Integrationen zwischen den Sinnesmodalitäten stattfinden. Dies geschieht auf allen Ebenen der Wahrnehmung, beginnend auf frühen, perzeptuellen Stufen (vergl. z.B. Kap. 4.1.4, *Kognitive Stufen der Verarbeitung bei kreuzmodaler Vereinnahmung*), über die Ebene der Wahrnehmungsorganisation bis hin zu Ebenen der Identifikation und Wiedererkennung (vergl. z.B. Kap. 4.1.5, *Kongruenz-Interaktionen zu Grunde liegende Mechanismen*).

Chion stellt zwei grundlegende Arten heraus, wie der emotionale Gehalt von Musik mit dem des Bildes zusammenwirken kann, *empathic* und *unempathic*. Die

Wirkweise von *unempathic music* beruht demzufolge auf dem emotionalen Kontrast zwischen bildlich Dargestelltem und der Musik: Der Gegensatz zwischen beiden verstärkt die emotionale Wirkung der Szene.

Mehrere der in Kapitel 3 und 4 vorgestellten Analysen und Studien kommen zu dem Ergebnis, dass Kontraste und ihre Verarbeitung in der Wahrnehmung eine besondere Rolle spielen. In Kapitel 3 wird die laterale Inhibition als ein allgemeines Prinzip der neuronalen Signalverarbeitung beschrieben, welches sowohl auf untersten Reiz-Ebenen (Rezeptive Felder der Retina), als auch in höheren Stufen der kognitiven Verarbeitung greift. An optischen Täuschungen wie dem Hermann-Gitter oder dem Simultankontrast kann dieses sichtbar gemacht werden. Insbesondere der Simultankontrast zeigt, dass die Intensität umgebender Reize die wahrgenommene Intensität eines Zielreizes beeinflusst. Dabei kommt es im Allgemeinen zu einer Überhöhung von Kontrasten durch das visuelle System (vgl. Kap. 3.3, *Neuronen, Synapsen, neuronale Netze* sowie *Funktionsprinzipien des Gehirns: Laterale und vertikale Integration*).

In Kapitel 4.1.5, welches sich mit kreuzmodalen Prozessen in schnellen Klassifizierungen beschäftigt, wird deutlich, dass die qualitative Beurteilung von Reiz-Dimensionen im Wesentlichen kontextuell, also im Bezug auf zeitlich oder örtlich nahe Reize erfolgt. Auch hier spielt Kontrastierung eine zentrale Rolle. Ben-Artzi und Marks (1995) stellten fest, dass die physikalische Differenz zwischen Reizen auch einen Einfluss auf Kongruenz und Interferenz-Effekte hat. Das heißt, sie sind nicht nur beim Klassifizieren entlang einer Dimension von Bedeutung (*baseline task*), sondern auch der Kontrast einer irrelevanten Dimension (*selective task*) hat Einfluss auf die Wahrnehmung und Klassifizierung des Ziel-Reizes (vgl. Kap. 4.1.5, z.B. *Kongruenz-Interaktionen zu Grunde liegende Mechanismen*).

Kontraste sind demnach ein wichtiges Element der menschlichen Wahrnehmung: Zum einen werden sie durch laterale Inhibition verstärkt, zum anderen werden sie benötigt, um Reize qualitativ beurteilen zu können.

Diese Forschungsergebnisse bilden eine Schnittmenge mit Chions Thesen zur Wirkung von *unempathic music*. Auch er sagt, dass sich die Intensität einer Emotion dadurch erhöht, dass eine andere ihr entgegensteht und somit einen Kontrast bildet.

Es sei an dieser Stelle betont, dass es sich bei den dargelegten wissenschaftlichen Untersuchungen und Chions These um zwei verschiedene Themengebiete handelt, Reizverarbeitung und Klassifizierung einerseits und emotionale Kommunikation andererseits. Diese sind nicht direkt vergleichbar und der hier hergestellte Zusammenhang ist somit eine übergreifende Interpretation. Emotionen sind ein eigenständiger Gegenstand von Forschungen, ihre Verarbeitung und Kodierung unterliegt nochmals komplexeren Mechanismen, als die Reizverarbeitung in den dargestellten schnellen Klassifikationen. Trotzdem ist eine Emotion auch ein Reiz, der ebenfalls seine Repräsentation in Form von neuronalen Aktivierungsmustern hat. Kontrastierung könnte daher ebenfalls eine Rolle spielen.

Chion behandelt temporäre Aspekte der Wahrnehmung. Er stellt fest, dass die akustische Wahrnehmung der visuellen auf zeitlicher Ebene überlegen ist, wohingegen die visuelle Wahrnehmung räumlich besser ausgestattet ist.

Dies stellt eine eindeutige Übereinstimmung mit den wissenschaftlichen Studien dar. Kap. 4.1.2 erläutert die *modality appropriateness hypothesis*, nach der diejenige Modalität, die im Kontext einer bestimmten Aufgabenstellung verlässlicher und sicherer ist, die Führung über die Wahrnehmung bei der Lösung dieser Aufgabe übernimmt. Das ist im Zeitlichen das Hören und im Räumlichen das Sehen (vgl. Kap. 4.1.2 *Der Einfluss akustischer Reize auf temporäre Aspekte des Sehens*).

Übereinstimmung in diesem Punkt findet sich auch mit den in Kap. 3.2.1 und 3.2.2 erläuterten räumlichen und zeitlichen Auflösungen von Auge und Ohr: Während das visuelle System eine maximale räumliche Auflösung von 10 Bogensekunden erreicht, liegt beim Hören der kleinste wahrnehmbare Richtungsunterschied bei etwas unter 2°. Das entspricht einem Winkel-Verhältnis von 1:720.

Die kritische Dauer (Zeit, unterhalb derer optische Reize gleicher Energie nicht unterscheidbar sind) beträgt beim Auge 40ms. Das gesunde Gehör eines Erwachsenen ist in der Lage Frequenzen bis zu 16kHz zu verarbeiten, das entspricht Schwingungen mit einer Periodendauer von 0,0625ms. Dies ist ein Zeitverhältnis zur kritischen Dauer von 640:1 (vgl. Kap. 3.2.1, *Das Auge* und Kap. 3.2.2, *Das Ohr*). Es ist demnach wissenschaftlich belegt, dass das Gehör

über die höhere zeitliche Auflösung und das visuelle System über die höhere räumliche Auflösung verfügt.

Ferner behandelt Chion das Thema der Bewegungswahrnehmung im Akustischen und Visuellen. Demnach kann Klang im Film das Erkennen schneller Vorgänge verbessern, wie z.B. in actionreichen Szenen mit rasanten Bewegungsabläufen. Dies deckt sich ebenfalls mit der „*modality appropriateness hypothesis*“ (s.o.): Im Sinne der Aufgabenstellung, nämlich zeitlich schnelle Abläufe im Film zu verfolgen, ist das Gehör dem visuellen System überlegen, da es über eine höhere zeitliche Auflösung verfügt. Ferner zeigten die Versuche von Scheier, Nijwahan und Shimojo (1999), dass die zeitliche Auflösung des Sehens durch akustische Reize verbessert werden kann. Ebenfalls relevant sind die Ausführungen von Shams, Kamitani und Shimojo (2004), nach denen ein plötzlicher Klang die Identifikation eines an selber Stelle folgenden Blitzes erhöhen kann, oder auch ein abrupter Klang die Identifikation eines synchronen visuellen Ziel-Reizes innerhalb dargebotener Stör-Reize verbessern kann (vgl. Kap. 4.1.2, *Der Einfluss akustischer Reize auf andere Aspekte des Sehens*). Dies sind experimentelle Situationen, die in ihrer Grundstruktur mit einer schnellen, rasanten Filmszene vergleichbar sind.

Der zweite Aspekt, den Chion zum Thema Bewegungswahrnehmung herausstellt, ist, dass mit Hilfe von Klängen visuelle Wahrnehmungen (vornehmlich Bewegungen, wie z.B. Schläge und Stöße) hervorgerufen werden können, die auf der Bild-Ebene nicht vorhanden sind, bzw. nur angetäuscht werden.

Auch hier gibt es ähnliche Beobachtungen in wissenschaftlichen Studien. Der *illusory flash effect* zeigt, dass unter bestimmten Voraussetzungen visuelle Wahrnehmungen durch akustische Reize erzeugt werden können. In dem Versuch von Shams, Kamitani und Shimojo (2000, 2002) wird ein einzelner Blitz durch gleichzeitige Darbietung zweier kurz aufeinander folgender Piep-Töne als Doppelblitz wahrgenommen (vgl. Kap. 4.1.2, *Manipulation der visuellen Struktur durch das Hören*). Akustische Stimuli induzierten demnach in diesen Versuchen ein visuelles Erleben, das real nicht vorhanden war. Ähnliches beschreibt Chion für das Kino.

In seinen weiteren Ausführungen erläutert Chion dann einige Bedingungen, an die die Effekte der Bewegungswahrnehmung geknüpft sind. Er skizziert den Fall, dass das Bild keine eigene zeitliche Animation oder Vektorisierung besitzt. In einem solchen Fall übernimmt das Bild (und mit ihr die gesamte Szene) die zeitliche Struktur des Tons.

Meyer und Wuerger (2001) untersuchten den Einfluss akustischer Bewegung auf die visuelle Bewegungswahrnehmung. Sie installierten ein RDK (random dot kinematogram) mit dem sie mehr oder weniger kohärente, bewegte visuelle Reize erzeugten und präsentierten dazu akustische *apparent motion streams*. Sie fanden heraus, dass visuelle nicht-kohärente Reize, also solche die keine eindeutige Bewegungsrichtung haben, durch den akustischen Reiz als bewegt wahrgenommen wurden, und zwar in Richtung der akustischen Bewegung. Hatte indes die visuelle Bewegung eine identifizierbare Richtung (Kohärenz), gab es keine Beeinflussung durch den akustischen Reiz (vgl. Kap. 4.1.2, *Klang verändert die visuelle Bewegungswahrnehmung* sowie Kap. 4.1.4, *Multisensorische Integration zwischen bewegten Ereignissen*).

Diese Ergebnisse weisen Parallelen mit Chions Aussagen insbesondere darin auf, dass Bilder ohne eigene Bewegungsrichtung die Bewegung vom Ton übernehmen. Seine Beschreibungen von fluktuierenden Bewegungen, erzeugt durch rinnendes Wasser oder rauschende Blätter, ähneln stark dem Reizmuster eines RDK mit nicht-kohärenten Bewegungen (viele kleine Objekte bewegen sich zufallsgesteuert und unabhängig voneinander).

Anschließend beschreibt Chion auch den Fall, dass das Bild eine eigene zeitliche Animation oder Vektorisierung besitzt. In einem solchen Fall kommt es zu einem Zusammenspiel der jeweiligen Zeitlichkeit von Ton und Bild. Je nach Gestaltung der Szene können Ton und Bild sich auf dieser Ebene ergänzen, oder gegenläufig funktionieren.

Parallelen sind hier zu Versuchen von Soto-Faraco, Lyons et al. (2002) erkennbar, die der Frage nachgingen, unter welchen Voraussetzungen es zu einer Anhaftung der Bewegungswahrnehmung zwischen akustischen und visuellen Reizen kommt. Sie kamen zu dem Ergebnis, dass das Hören einer bewegten Quelle durch eine visuelle Bewegung irritiert werden kann, wenn beide Stimuli synchron sind. Es kommt laut dieser Studie zu einer Interaktion

zwischen der Bewegungswahrnehmung des Sehens und des Hörens in Abhängigkeit vom zeitlichen und dem Richtungs-Verhältnis der Modalitäten zueinander (synchron-kongruent, asynchron-inkongruent, usw.) (vgl. Kap. 4.1.4, *Kreuzmodale Vereinnahmung der Bewegungswahrnehmung*).

Die Gemeinsamkeit zwischen Chion und Wissenschaft liegt demnach darin, dass es zu Vereinnahmungen der Bewegungswahrnehmung kommen kann, die von der Gestaltung der Verbindung abhängig ist.

Als letztes sei hier die Aussage Chions genannt, dass kausales Hören immer etwas mit Erfahrung und erlerntem Wissen zu tun hat.

In Kap. 3 wurde ausführlich erläutert, dass Wahrnehmung ein aktiver Prozess der Synthese und Konstruktion unter Zuhilfenahme von Wissen und Erfahrung ist. Top-Down-Verarbeitung spielt dabei eine wichtige Rolle auf den Ebenen der Organisation und Identifikation (vgl. Kap. 3.3, *Bottom-Up und Top-Down: Die Rolle des Wissens in der Wahrnehmung sowie Stufen der Verarbeitung*). Als Folge dessen ist nur ein gewisser Teil der Gehirnaktivität auf sensorischen Input zurückzuführen, ein weiterer Teil ist intrinsische Aktivität (vgl. Kap. 3.3, *Gehirnaktivität und Wahrnehmung*). Was Chion für das kausale Hören formuliert ist demnach ein allgemeines Prinzip der Wahrnehmung.

### *Zusammenfassung*

Die erläuterten Ansichten und Thesen zeigen, dass es Parallelen zwischen medientheoretischen Analysen und kognitionswissenschaftlichen Untersuchungen gibt. Ein grundsätzlicher Unterschied zwischen beiden Disziplinen ist, dass in wissenschaftlichen Experimenten einzelne Parameter isoliert herausgestellt und untersucht werden, um zu einem empirischen Nachweis kommen zu können. Die Reproduzierbarkeit eines Resultats ist oberste Prämisse, die Kontrolle von verfremdenden oder störenden Faktoren und Einflüssen ist daher notwendig. Chion dagegen betrachtet das Gesamte und stellt durch Überlegungen unter Einbeziehung von Erfahrungen Zusammenhänge her, die komplex sind und eine Vielzahl von Parametern berücksichtigen. Aus seiner Sicht ist dies notwendig, da bei der Analyse eines Film vielfältige Faktoren eine Rolle spielen: Ein Bild enthält Formen, Farben, Bewegung, Repräsentationen von Objekten, Emotionen, und vieles mehr. Gleiches gilt für

den Ton, die Verknüpfung beider Ebenen erhöht die Komplexität nochmals deutlich.

Die hier aufgezeigten Parallelen zwischen den Schlussfolgerungen, zu denen beide Disziplinen kommen, sind ein Hinweis darauf, dass medientheoretische Analysen und Thesen kognitionswissenschaftlich belegbar sind. Die moderne Forschung kommt durch neue Methoden, wie bildgebende Verfahren, zu immer tieferem Verständnis, wie die menschliche Kognition und Wahrnehmung funktionieren. Dies ist auch von grundlegendem Interesse für Medienschaffende, -theoretiker und -wissenschaftler. Denn besseres Verstehen der Wahrnehmung kann zu besserer Kommunikation verhelfen. Andersherum könnten auch psychologische und kognitionswissenschaftliche Fragestellungen durch medientheoretische oder -wissenschaftliche Analysen angeregt werden oder gar bei deren Beantwortung helfen.

## 7 PRAKTISCHE ARBEIT

Die Praktische Arbeit ‚congruence‘ ist eine Musikvisualisierung, bzw. ein Visual Music Clip. Dies bedeutet im Allgemeinen, dass Musik und bewegtes Bild in besonderer Weise aufeinander abgestimmt sind. Beide Ebenen sind also eng miteinander verbunden. Visual music ist der Versuch einer Sichtbarmachung des Gehörten, bzw. der Hörbarmachung des Gesehenen. Es geht darum, eine gemeinsame Erlebnisdimension aus Visuellem und Akustischem zu kreieren. Abstraktion ist dabei ein häufig verwendetes gestalterisches Mittel. Damit ist Visual Music ein künstlerischer Ansatz, Verbindungen zwischen der akustischen und visuellen Wahrnehmung zu finden.

In dem Visual Music Clip dieser Diplomarbeit erfolgt die Visualisierung der Musik anhand der theoretisch erarbeiteten multisensorischen Prozesse zwischen akustischer und visueller Wahrnehmung. Welche das sind und wie sie umgesetzt wurden wird in diesem Kapitel erläutert.

### 7.1 Theoretische Hintergründe

Ideengebend für den Clip war die Thematik ‚Kongruenz-Effekte in schnellen Klassifizierungen‘ (vgl. Kap. 4.1.5). Diese beschreiben, wie in empirischen Tests bestimmte Qualitäten eines Reizes schneller klassifiziert werden, wenn ein anderer Reiz synchron dazu präsentiert wird. Auf diese Weise stehen visuelle und akustische Reize folgendermaßen in Verbindung: Die Lautstärke eines Tons ist kongruent mit der Helligkeit sowie der Größe eines visuellen Reizes. Akustisch ‚laut‘ wird demnach schneller klassifiziert in Verbindung mit visuell ‚hell‘ bzw. visuell ‚groß‘; akustisch ‚leise‘ dagegen mit visuell ‚dunkel‘ bzw. ‚klein‘. Die Tonhöhe eines akustischen Reizes ist kongruent zu den visuellen Dimensionen Helligkeit, vertikale Position, Größe und Farbton. Akustisch ‚hoch‘ wird demnach schneller klassifiziert in Verbindung mit visuell ‚hell‘, ‚oben‘ und ‚klein‘, akustisch ‚tief‘ dagegen mit visuell ‚dunkel‘, ‚unten‘ und ‚groß‘.

Des Weiteren wurde die theoretischen Themen ‚kreuzmodale Assoziationen‘ bzw. ‚synästhetische Metaphern‘ (Ramachandran/Hubbard, 2001b) in der Musikvisualisierung aufgegriffen (vgl. Kap. 4.1.5 und Kap. 5.2.7). Hierbei geht es

um die Rolle der Sprache innerhalb multisensorischer Prozesse. Viele Hinweise sprechen dafür, dass sprachliche Beschreibungen wie Adverbien und Redewendungen, Ausdruck von Wahrnehmungsprozessen sind. Ein sehr eindrucksvolles Beispiel hierfür ist das der sprachlichen Beschreibung der Tonhöhe als ‚hoch‘ bzw. ‚tief‘, was sich in fast allen Sprachen der Welt wiederfindet. Die Dimension ‚Höhe‘ repräsentiert indes nicht nur akustische Phänomene, sondern mittelbar über die räumliche Wahrnehmung auch visuelle. Sie dient dort vor allem der qualitativen Beschreibung der vertikalen Position von Objekten. ‚Höhe‘ ist somit eine multisensorische Dimension. Dieses Wahrnehmungs-Prinzip findet sich u.a. in der Notenschrift wieder. Dort symbolisieren (vereinfacht ausgedrückt) oben angeordnete Punkte musikalisch hohe Töne und unten angeordnete Punkte musikalisch tiefe Töne.

Michel Chion knüpft ebenfalls an die Idee kreuzmodaler Assoziationen an. Er stellt fest, dass gerade zur qualitativen Beschreibung akustischer Reize oftmals Begriffe herangezogen werden, die originär einer anderen Modalität zugeschrieben werden können (vgl. Kap. 6.2: Reduziertes Hören). Beispiele für solche Beschreibungen von Klangeigenschaften sind ‚scharf‘, ‚rau‘, ‚spitz‘, ‚grell‘, ‚dumpf‘, ‚dunkel‘ usw. Diese Liste ließe sich weit fortführen. Dagegen gibt es weit weniger Begriffe, die in erster Linie der Beschreibung akustischer Reize dienen bzw. originär mit dieser Dimension in Verbindung gebracht werden (vgl. Chion, S.31). Auch aus diesen Beobachtungen lässt sich schließen, dass Sprache Aufschluss über die Wahrnehmung geben kann, gerade auch im Hinblick auf multisensorische Prozesse.

Ein weiteres theoretisches Thema, welches stets bei der Betrachtung von audiovisuellen Medien relevant ist und somit auch in den praktischen Teil dieser Diplomarbeit einfließt, ist der Bauchredner-, oder Ventriloquist-Effekt (vgl. Kap. 4.1.3). Er beschreibt, unter welchen Voraussetzungen ein konsistentes Objekt als Quelle von Bild und Ton wahrgenommen wird. Untersuchungen haben ergeben, dass die Ortbarkeit der Schallquelle ein wesentlicher Faktor ist. Nur wenn diese nicht eindeutig aus einer anderen Richtung als das Bild kommend wahrgenommen werden kann, ist die Anhaftung an das Visuelle möglich. Synchronität spielt ebenfalls eine wichtige Rolle; Dabei wirken sich zeitliche Differenzen zwischen Ton und Bild kritischer aus, wenn der Ton vor dem Bild

kommt, als wenn er ihm folgt. Für die Gestaltung von audiovisuellen Medien sind diese Mechanismen äußerst wichtig, da Wiedergabesysteme in aller Regel das Bild und den Ton räumlich getrennt wiedergeben. Durch das Synchronisieren beider Ebenen lässt sich jedoch trotzdem die Illusion hervorrufen, sie entsprängen ein und derselben Quelle (z.B. dem Mund des Moderators).

Die von Michel Chion beschriebene Idee des „*Reduced Listening*“, welche in Kap. 6.2 behandelt wird, fand ebenfalls Einzug in den praktischen Teil dieser Diplomarbeit. Es ist das Konzept, einen Klang rein nach seinen inhärenten Eigenschaften zu beschreiben und nicht nach seiner Ursache, was Chion im Gegensatz dazu „*Causal Listening*“ nennt.

Beispiele für die Beschreibung eines Klangs nach den Maßgaben des „*Reduced Listening*“ sind ‚hoher Ton‘ oder ‚weicher Klang‘. Beispiele für die Beschreibung eines Geräuschs nach den Maßgaben des „*Causal Listening*“ sind ‚Geigenton‘ oder ‚mechanischer Klang‘, also Beschreibungen, die sich auf die Ursache bzw. Quelle des Klangs beziehen.

Die Idee des „*Reduced Listening*“ steht indes bei der Umsetzung dieses Clips in einem engen Verhältnis zu den kreuzmodalen Assoziationen. Qualitative Beschreibungen nach Art des „*Reduced Listening*“ ziehen wie beschrieben oft Adverbien und sprachliche Wendungen heran, die auch in einer anderen sensorischen Dimension verwendet werden und sind damit per se kreuzmodale Assoziationen, also über oder zwischen den jeweiligen Modalitäten stehende Attribute.

## 7.2 Die Umsetzung

Es wird nun erläutert, wie die theoretischen Überlegungen in praktische audiovisuelle Gestaltung umgesetzt wurden. Bei manchen Aspekten geschah dies sehr direkt, andere unterliegen stärkeren Interpretationen.

### 7.2.1 Die Musik

Begonnen wurde die Umsetzung mit der Komposition und Produktion eines speziell auf die Anforderungen dieser Diplomarbeit zugeschnittenen Musikstücks

durch Fabian Schulz. Anhand der oben aufgeführten theoretischen Betrachtungen wurden dafür die folgenden Vorgaben formuliert:

1. Eine musikalische Ebene zeichne sich durch möglichst große Kontraste in Tonhöhe und Lautstärke aus. An ihr sollen die erläuterten Kongruenz-Effekte verdeutlicht werden können. Sie sei daher möglichst frei von Klang-Attributen, die von dieser Intention ablenken.
2. Eine weitere Ebene arbeite mit Klängen, die mit Attributen und Ausdrücken wie „rund“, „eckig“, „hart“, „weich“, „spitz“, „stumpf“ etc. beschreibbar sind, also vornehmlich solchen aus der visuellen Dimension der geometrischen Form. An dieser Ebene soll das Thema der kreuzmodalen Assoziationen (im Zusammenspiel mit „Reduced Listening“) verdeutlicht werden können. Sie sei daher möglichst frei von Klangattributen, die außerhalb der beschriebenen Dimension liegen.
3. Es werde ein hohes Tempo gewählt und die Dauer der Töne und Geräusche sei vorwiegend kurz. Ein Grund hierfür ist, dass viele der ermittelten Kongruenz-Effekte besonders in *speeded classifications* auftraten, demnach wenn schnelle Entscheidungen der Testpersonen gefragt waren. Zum anderen lässt sich durch ein hohes Tempo die Bedeutung der Synchronität zwischen Bild und Ton besser verdeutlichen (vgl. Kap. 7.1 und Kap. 4.1.3: „Ventriloquist-Effekt“).
4. Das Musikstück soll sich mit der Zeit derart aufbauen, dass die beschriebenen Elemente und ihre Ebenen Schritt für Schritt eingeführt werden. Dem Rezipienten sollte somit die Möglichkeit gegeben werden, jedes akustische und das dazugehörige visuelle Element aufmerksam wahrnehmen und kennenlernen zu können.
5. Es soll trotz aller theoretischen Vorgaben Musik sein, die als solche wahrgenommen wird und dementsprechend den Gestaltungsprinzipien eines Musikstücks folgt. Das Genre (Klassik/Jazz/Electro o.ä.) ist dem Komponisten freigestellt.

Entstanden ist ein elektronisches Musikstück von drei Minuten Länge, welches sich in Tempo und Rhythmus den Genres „Electro“ und „Drum n’Bass“ zuordnen

lässt. Es basiert auf den Akkorden A-Dur und F-Dur, das Tempo liegt bei 180 bpm. Die oben beschriebenen Vorgaben wurden folgendermaßen umgesetzt:

1. Die musikalische Ebene, an der die Kongruenz-Effekte deutlich gemacht werden sollten, wurde durch Sinustöne umgesetzt. Der Grund hierfür ist, dass Sinustöne keinerlei Obertöne und damit keine Klangfärbungen aufweisen. Ein Sinuston besteht aus einer einzelnen Frequenz, seiner Grundfrequenz. Daher lassen sich Sinustöne durch die beiden Attribute Tonhöhe und Lautstärke eindeutig beschreiben. Damit wurde die Vorgabe ideal umgesetzt, eine Ebene zu schaffen, deren Töne sich hauptsächlich durch diese beiden Dimensionen beschreiben lassen und andere Klangattribute möglichst ausschließt. Sie sei im Folgenden die **tonale Ebene**.
2. Die **Rhythmus-Ebene** des Stücks repräsentiert die Umsetzung der kreuzmodalen Assoziationen. Sie enthält Klänge, die sich vor allem durch Attribute der visuellen Dimension der geometrischen Form beschreiben lassen. Dagegen besitzen die Klänge dieser Ebene keine definierbare Tonhöhe. Sie sind deshalb weitestgehend frei von Klangattributen der ersten Ebene (außer der nicht vermeidbaren Lautstärke).
3. Tempo des Stücks ist mit 180bpm sehr hoch. Hinzu kommen der ungleichmäßige Rhythmus (gebrochener Drum N' Bass Beat) sowie die kurze Dauer der meisten Sinustöne (Achtelnoten), die dem Stück ein hohes Maß an Dynamik und Schnelligkeit verleihen.
4. Der musikalische Aufbau geschieht in Etappen, fast jedes neue Klangelement wird einzeln eingeführt. Dadurch ergibt sich eine stetige Steigerung der Intensität. Eine ausführliche Beschreibung des Aufbaus erfolgt weiter unten.
5. Das Musikstück ist als solches erkennbar und folgt gängigen musikalischen Gestaltungsprinzipien. Es ist daher prinzipiell auch ohne die dazugehörige Visualisierung konsumierbar.

Zur genaueren Analyse wird die tonale Ebene in fünf weitere Unterebenen geteilt; die erste ist die der *Random-Beeps*. Diese haben jeweils die Dauer einer Achtelnote und sind die musikalischen Töne a', cis' und c'. Ihre Abfolge ist

durch einen Zufallsgenerator bestimmt, wobei jede Achtelnote mindestens ein und maximal zwei *Random-Beeps* gleichzeitig zu hören sind. Die zweite tonale Unterebene ist die der *Sinus-Sweeps*. Sie besteht aus einem von ca. 440Hz (Kammerton a') bis 24kHz (Darstellungsgrenze bei 48kHz Abtastrate) ansteigenden und einem von ca. 264Hz (Ton c') bis 0Hz (untere Darstellungsgrenze variiert je nach Wiedergabesystem) abschwelldem *Sweep*. Die dritte tonale Unterebene ist die der *Sinus-Chords*. Diese Ebene besteht aus den Tönen a, c', cis', e' und f'. Dabei erklingen jeweils die Töne a, cis' und e' zusammen zu einem A-Dur Akkord sowie a, c' und f' zu einem F-Dur Akkord. Die vierte tonale Unterebene ist die der *E-Oktaven*. Sie besteht aus Viertelnoten der Tonhöhe e in fünf Oktaven abfallend, also e<sup>3</sup>, e<sup>2</sup>, e', e, E, alternierend gefolgt von einem Dis sowie einem F. Die fünfte Unterebene ist die des *Sinus-Breakdowns*. Sie besteht aus den Tönen a, cis und c in sechs absteigenden Oktaven beginnend mit einem a<sup>5</sup> und endend bei einem c'. Es sind somit fünf Arpeggien in Achtelnoten.

Die Rhythmus-Ebene des Stücks lässt sich in zwei Unterebenen aufteilen. Die erste ist die der *Rhythm-Clicks*. Sie besteht aus verschiedensten Klopf-, Kratz-, Rassel- und ähnlichen Geräuschen. Diese folgen zwar einer sich wiederholenden rhythmischen Form, die Art der Klänge (Samples) ist jedoch ebenfalls durch einen Zufallsgenerator gesteuert, so dass sich keiner der Sounds wiederholt. Die zweite rhythmische Unterebene ist die des *Rhythm-Beat*. Diese besteht aus einem für das Genre Drum N'Bass typischen, schnellen, gebrochenen Rhythmus. Die hierfür gewählten Klänge sind trotz einer gewissen synthetischen Anmutung als Bass Drum und Snare Drum zu identifizieren.

Im Folgenden wird das Stück chronologisch analysiert. Festzustellen ist zunächst eine deutliche Klimax: Beginnend mit sehr wenigen, rhythmisch angeordneten Sinustönen kommen im Verlauf stetig neue Ebenen hinzu, so dass der zu Beginn des Stücks großzügig vorhandene musikalische Raum zunehmend mit Klängen gefüllt wird und somit gefühlte Intensität und Tempo stetig zunehmen.

Es wird folgende chronologische Aufteilung vorgenommen:

- Intro: Dauer 32 Takte (ca. 42sec.); Tonale Ebene: *Random-Beeps* nur in a' in unterschiedliche Lautstärken, 1 langer Sinuston c' über 8 Takte in der Lautstärke anschwellend, 1 langer Sinuston c' über 8 Takte in der Lautstärke abschwelld, *Sinus-Sweeps* wie oben beschrieben über 8 Takte. 1 Sinuston c' über 2 Takte in seiner Lautstärke abschwelld; Rhythmus-Ebene: keine
- 1. Vers: Dauer 32 Takte (ca. 42sec.); Tonale Ebene: *Random-Beeps* alle 4 Takte Wechsel zwischen a' - cis' und a' - c'; Rhythmus-Ebene: *Rhythm-Clicks*
- Mittelteil: Dauer 16 Takte (ca. 21sec.); beginnt mit weißem Rauschen, zwei mal über 4 Takte von 0dB konstant abschwelld; Tonale Ebene: über dem Rauschen zwei mal *E-Oktaven*, ein mal mit einem F abschließend, einmal mit einem Dis, durch Zusammenklang von E und F bzw. E und Dis (harmonisches Intervall: Sekunde) kommt es zur Schwebung, danach zwei mal, in Lautstärke pulsierendes e<sup>5</sup> über 4 Takte, einmal mit E und F abschließend, einmal mit tiefem E und Dis abschließend, *Sinus-Breakdown* über die letzten 2 Takte, überlagert von weißem Rauschen
- 2. Vers: Dauer 16 Takte (ca. 21sec.); Tonale und rhythmische Ebene wie im 1. Vers plus tiefem E-F und E-Dis Wechsel (wie im Mittelteil), *Random-Beeps* verschieben sich jeweils nach 4 Takten um eine Oktave nach oben, weißes Rauschen über die letzten 4 Takte, Ankündigung der Beats durch unregelmäßige Bass Drum Schläge, die sich im Tempo zum Ende des Takts in 16tel Noten steigern
- 3. Vers: Dauer 32 Takte (ca. 42sec.); Tonale Ebene wie im 2. Vers, *Random-Beeps* verschieben sich weiter nach oben bis in die sechsgestrichene Oktave (in dieser ist der höchste Ton technisch bedingt ein g<sup>6</sup> statt einem a<sup>6</sup>), danach alle vier Takte Verschiebung der *Random-Beeps* eine Oktave nach unten, bis die eingestrichene Oktave wieder erreicht ist, nach 16 Takten Einsatz der *Sinus-Chords*; Rhythmische Ebene: Einsatz des *Rhythm-Beat*, Reduzierung der *Rhythm-Clicks*
- Ende Dauer 4 Takte (ca. 5sec.): Plötzliche Stille (Achtelpause), danach in der Lautstärke an- und wieder abschwelldes ,A (Kontra-Oktave)

### 7.2.2 Die Visualisierung

Die Visualisierung wurde anhand der oben beschriebenen theoretischen Überlegungen sowie der Analyse des Musikstücks vorgenommen.

Die **tonale Ebene** des Musikstücks wird im Bild durch horizontale Balken repräsentiert, die in ihrer vertikalen Position, ihrer Dicke („Größe“), ihrer Helligkeit und im Farbton variieren. Synchron zu jedem Sinuston erscheint jeweils ein Balken, dessen visuelle Parameter an die akustischen des repräsentierten Tons gekoppelt sind. Alle genannten visuellen Eigenschaften sind über die beschriebenen Kongruenz-Effekte an die Tonhöhe und die Lautstärke der Töne gebunden, mit Ausnahme der Farbtöne, deren Wahl weiter unten erläutert wird.

Die Umsetzung der Kongruenz-Beziehung zwischen Lautstärke und Helligkeit geschah wie folgt: Je lauter ein Sinuston ist, desto heller ist der dazugehörige Balken. Dies ist besonders gut im Intro zu sehen, wo derselbe Ton a' in vielen verschiedenen Lautstärken nacheinander zu hören ist. Passend zur Lautstärke variiert hier nur die Helligkeit des Balkens, während alle anderen Parameter konstant bleiben.

Die Umsetzung der Kongruenzbeziehungen zwischen Tonhöhe und vertikaler Position sowie Größe ist indes sehr viel komplexer. Prinzipiell ist es so, dass der Balken eines höheren Tons vertikal über dem eines tieferen Tons angeordnet und schmaler ist als der des tieferen Tons. Dies ist sehr gut zu sehen im 1. Vers, wenn die *Random-Beeps* einsetzen. Hierzu wurde eine Einteilung des Bildbereichs anhand der Grundoktave des Stücks (eingestrichene Oktave) vorgenommen, also jener, in denen die *Random-Beeps* des 1. Verses liegen.

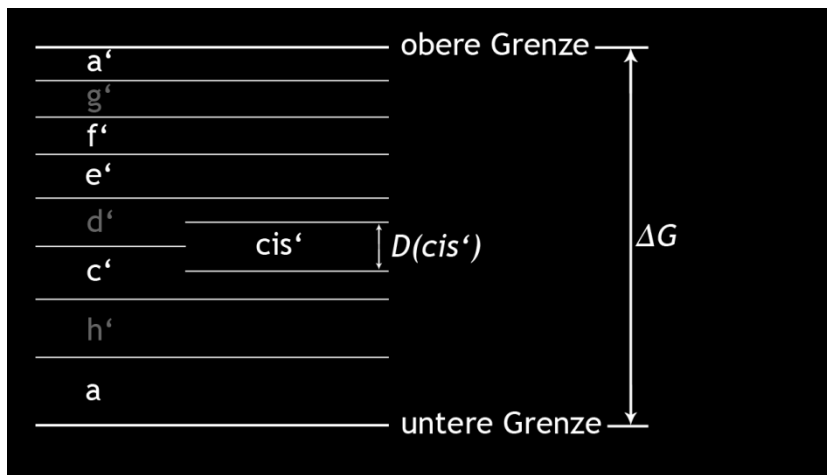


Abb. 7.1

Abb. 7.1 zeigt die vertikale Aufteilung des Bildbereichs (Videobild HD 720p 25). Zunächst wurde eine obere und eine untere Grenze festgelegt, außerhalb derer zunächst keine Balken zu sehen sein sollen. Diese wurden rein nach ästhetischen Gesichtspunkten gewählt und haben keinen wissenschaftlichen Hintergrund. Der zur Verfügung stehende Bereich ( $\Delta G$ ) wurde nun in acht Abschnitte geteilt, entsprechend acht Ganztonschritten von ,a bis a. Dabei ist der Balken des tiefen a genau doppelt so dick, wie der des a (entsprechend dem Frequenzverhältnis 2:1 eines Oktavintervalls). Die Dicke  $D$  der Balken dazwischen berechnet sich entsprechend der linearen Funktion  $D(N) = 1 \frac{N-1}{7} D(1)$ .  $D$  sei die Dicke des Balkens,  $N$  die fortlaufende Nummerierung der Balken von oben nach unten ( $N(a') = 1$ ;  $N(g') = 2$ ;  $N(f') = 3$ ;  $N(e') = 4$ ;  $N(d') = 5$ ;  $N(c') = 6$ ;  $N(h') = 7$ ;  $N(a) = 8$ ).

Neben diesen Ganztonschritten kommt ein Halbtonschritt von  $c'$  nach  $cis'$  vor. Der Balken des  $cis'$  liegt (ähnlich der schwarzen Tasten der Klaviatur) bei der vorgenommenen Einteilung genau zwischen dem der Töne  $c'$  und  $d'$ . Es sei für die Berechnung der Dicke dieses Balkens festgelegt:  $N(cis') = 5 \frac{1}{2}$ . Seine Position ist vertikal zentriert auf der Grenze zwischen den Balken von  $c'$  und  $d'$ . Wie in Kap. 7.2.1 beschrieben, variiert jedoch im Verlauf des Musikstücks die Tonhöhe der *Random-Beeps* über sechs Oktaven, das heißt im 2. Vers verlassen sie die eingestrichene Oktave und wechseln alle 4 Takte in die nächst höhere. In der sechsgestrichenen Oktave bleiben sie über acht Takte, bevor sie wieder im vier-Takte-Rhythmus den Abstieg in die eingestrichene Oktave beschreiten. Diese Oktavierungen wurden durch eine Stauchung der oben geschilderten Aufteilung umgesetzt. Die Position des jeweils höheren a bleibt dabei stets

ortskonstant, während die Anordnung aller Balken pro Oktavierung nach oben um genau die Hälfte vertikal zusammengestaucht wird. Gleichzeitig mit einer Ortsverschiebung tritt damit eine Halbierung der Dicke aller Balken der *Random-Beeps* ein.

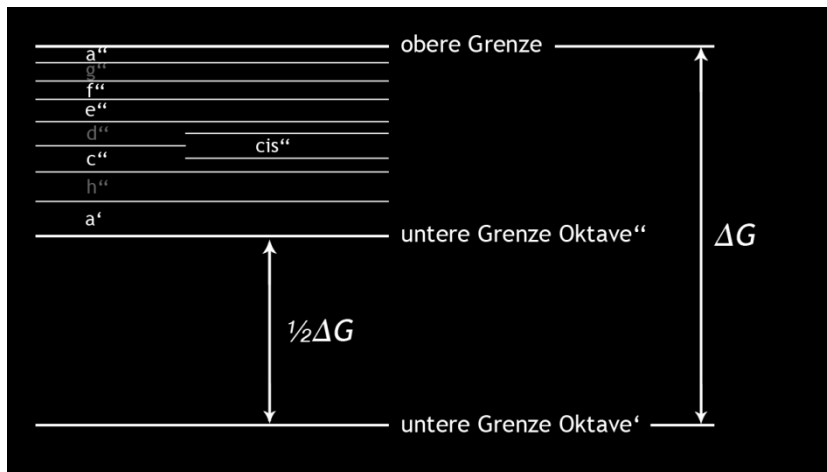


Abb. 7.2a

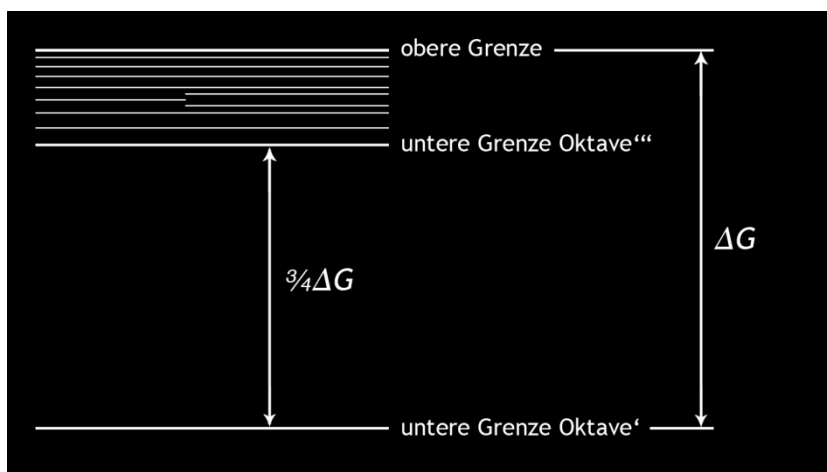


Abb 7.2b

Abb. 7.2a und 7.2b zeigen die daraus entstehenden Positionen und Dicken der Balken der *Random-Beeps* der zweigestrichenen und dreigestrichenen Oktave. Der Grund für diese Umsetzung liegt in dem insgesamt sehr großen Tonumfang des Musikstücks, der bei über sechs Oktaven liegt. Eine ‚lineare‘ oder ‚absolute‘ Umsetzung der Kongruenz-Beziehung zwischen Tonhöhe und vertikaler Position könnte z.B. bedeuten, das zur Verfügung stehende Videobild (HD 720p 25) gleichmäßig in 72 horizontale Balken (6 Oktaven à 12 Töne), ähnlich einer Klaviatur, aufzuteilen. Dies würde zu einer äußerst kleinteiligen Skala und somit zu teils sehr geringen Unterschieden zwischen einzelnen Balken führen. Eine

solche Aufteilung gäbe jedoch nicht den subjektiv empfundenen Tonhöhenunterschied der nacheinander präsentierten Töne (*Random-Beeps*) wieder. Denn die Wahrnehmung klassifiziert Reize stets relativ (vgl. Kap. 4.1.5), also im Verhältnis zueinander. Ein Ton wird demnach als ‚hoch‘ oder ‚tief‘ im Bezug auf ihm zeitlich vorausgegangene Töne empfunden. Die *Random-Beeps*, die sehr schnell aufeinander folgen, bewegen sich stets mehrere Takte am Stück innerhalb einer Oktave. Das führt dazu, dass die einzelnen Töne wesentlich im Verhältnis zu den ihnen unmittelbar vorausgegangenen klassifiziert werden. Deshalb macht es Sinn, die Tonhöhen-Differenzen innerhalb einer Oktave als Bezugsgröße für die Umsetzung in die vertikale Position der Balken (‚Höhe‘) heranzuziehen. Durch die beschriebene Stauchung der visuellen Anordnung pro Oktavverschiebung ergeben sich mehrere Parallelen zur akustischen Ebene des Clips. Zum einen verhält sich die Dicke der Balken proportional zur Frequenz der zugehörigen Töne und zum anderen verhält sich der Abstand zwischen den Balken proportional zur Frequenzdifferenz zwischen den zugehörigen Tönen.

Die beschriebenen Aufteilungen des Bildes richten sich vor allem nach den tonalen Unterebenen *Random-Beeps* sowie *Sinus-Chords*. Anknüpfend an dieses Konzept wurden Position und Dicke der *E-Oktaven* und *Sinus-Sweeps* gelöst. Die Balken der E-Oktaven haben im Bild alle dieselbe vertikale Position, wie das e', bezogen auf ihren Mittelpunkt. Die fünf Oktavierungen werden ausschließlich durch die Dicke der Balken, sowie ihre Helligkeit repräsentiert. Dadurch wird deutlich gemacht, dass es sich um dieselben musikalischen Töne in verschiedenen Oktaven handelt.

Die Balken der *Sinus-Sweeps* im Intro sind ebenfalls ortskonstant, bezogen auf ihre geometrische Mitte. Ihre Position richtet sich nach der Tonhöhe, in der sie beginnen. Das ist beim ersten, ansteigenden *Sweep* ein a' und beim zweiten, abfallenden ein c'. Zum Zeitpunkt ihres Einsetzens werden diese Sweeps durch Balken repräsentiert, die identisch mit denen der *Sinus-Beeps* a' und c' sind. Mit zunehmender bzw. abfallender Tonhöhe verändern sich ihre Dicke sowie Helligkeit.

### **Die Farbwahl der Balken**

Kognitionswissenschaftliche Untersuchungen stellten fest, dass die Kongruenz-Beziehung zwischen Farbton und Tonhöhe weniger stark ausgeprägt sind, als die

zwischen Helligkeit, Größe, vertikaler Position und Tonhöhe sowie Lautstärke. Es konnte bisher auch nur ein sehr undifferenzierter Zusammenhang hergestellt werden, der grundsätzlich nur zwischen den Farbtönen Rot und Blau unterscheidet. Es wird angenommen, dass diese Kongruenz-Beziehung zwischen Tonhöhe und Farbton erlernt ist und entfernt mit Temperatur-Assoziationen zu tun haben könnte.

Eine sehr feine Differenzierung der Farbtöne findet sich dagegen bei Synästheten. Über 70% der Synästhesien sind solche, in denen eine Farbwahrnehmung durch einen anderen Sinnesreiz ausgelöst wird, davon sind bei rund 20% akustische Reize der Trigger (Day, 2009). Die Betroffenen schildern dabei meist sehr differenziert die von ihnen erlebten Farben. Dies betrifft alle Eigenschaften von Farbe, Helligkeit, Sättigung und Farbton. Die synästhetischen Farbwahrnehmungen sind zwar sehr differenziert, allerdings individuell sehr unterschiedlich und heterogen. Es lassen sich hier keinerlei Tendenzen feststellen, die eine Verallgemeinerung zulassen würden. Einzig die Tatsache, dass hohe Töne tendenziell zur Wahrnehmung heller Farben und tiefe Töne zur Wahrnehmung dunkler Farben führen, ist bei fast allen Betroffenen gleich, was eine Übereinstimmung mit der Kongruenz-Beziehung zwischen Tonhöhe und Helligkeit darstellt (Marks, 2004, 93; vgl. Kap. 4.1.5). Überzufällige Zusammenhänge zwischen akustischen Eigenschaften und Farbtönen scheint es demnach nicht zu geben. Sie sind weder bei Synästheten, noch bei Nicht-Synästheten nachweisbar. Aus diesem Grund wurde bei der Wahl der Farbtöne ein Ansatz gewählt, der sich indirekt an den wissenschaftlichen Ergebnissen orientiert. Dazu wurde die Kongruenz-Beziehung zwischen Temperatur und Farben (warm ist kongruent zu gelb/rot, kalt zu blau/grün) (Marks, 2004) als Inspiration genutzt.

Als Grundlage diente die Harmonik des Musikstücks, also der Zusammenklang der Töne. Die Idee ist, eine gestalterische Analogie zwischen der musikalischen Wirkung des Zusammenklangs der Töne einerseits und dem Zusammenspiel der Farben andererseits herzustellen. Dies deckt sich mit der Idee von Johannes Itten, dass die Wirkung einer Farbe wesentlich durch ihr gemeinsames Erscheinen mit einer anderen Farbe entsteht (Itten, 2001). Hier besteht eine starke Analogie zur Musik, denn auch erst das Zusammenklingen von Tönen

verleiht ihnen im Wesentlichen ihre Wirkung. Einzelne Töne haben dagegen meist weniger Aussagekraft.

Als Hintergrundfarbe wurde schwarz gewählt. Das ist deshalb wichtig zu erwähnen, weil auch schwarz eine Farbe ist, die im Zusammenspiel mit anderen Farben eine bestimmte Wirkung hat, und keineswegs als farblich neutral angesehen werden sollte (Itten, 2001). Der Grund für diese Wahl ist, dass schwarz die Abwesenheit von Licht repräsentiert, was das Pendant zur akustischen Stille ist, also die Abwesenheit von Klang. Da bei dieser Musikvisualisierung eine sehr enge Bindung von Musik und bewegtem Bild realisiert wurde, hat jedes Ton-Element einen Repräsentanten auf der Bild-Ebene, somit auch die Stille.

Im 1. Vers erklingen die *Random-Beeps* in den Tönen a', cis' und c', wobei alle 4 Takte ein Wechsel zwischen a'-cis' und a'-c' stattfindet. Obwohl die Töne alle einzeln erklingen, können diese beiden Intervalle aufgrund des hohen Tempos und des unmittelbaren Aufeinanderfolgens als zusammenklingend betrachtet werden. Diese beiden Zweiklänge sind streng betrachtet kein Akkord. Sie können allerdings eine akkordische Funktion erfüllen, wenn ein dritter Ton dazu interpretiert wird. Aufgrund westlicher Hörgewohnheiten ist das in diesem Fall ein a als Oktave zum a'. Damit entsteht aus a', cis' und einem ‚virtuellen‘ a ein A-Dur, sowie aus a', c' und einem ‚virtuellen‘ a ein A-Moll.

Der erste Höreindruck von Dur wird oft als „fröhlich, hell, klar“ beschrieben, wogegen Moll oft als „traurig, dunkel, weich“ bezeichnet wird. Diese Ideen des Zusammenklangs wurden bei der Wahl der Farbtöne aufgegriffen. Ausgehend vom a' als konstantes Element in beiden Zweiklängen wurde eine Farbe für den Balken des cis' gesucht, die zusammen mit der des a' eher „hell, klar und fröhlich“ wirkt, sowie eine Farbe für den Balken des c', die zusammen mit der des a' eher „traurig, dunkel und weich“ wirkt.

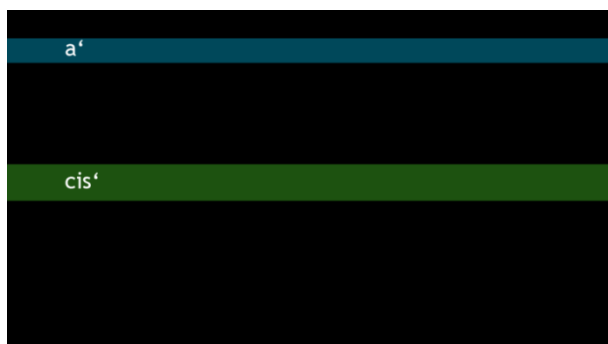


Abb. 7.3a



Abb. 7.3b

Abb. 7.3a zeigt die Farben der Balken der Töne a' und cis', Abb. 7.3b die der Töne a' und c'. Für das a' wurde aus rein subjektiven Erwägungen heraus ein relativ dunkles, gedecktes blau gewählt. Es ist eine Farbe, die auf schwarz recht unaufdringlich wirkt und wenig Strahlkraft hat. Das findet sich auf musikalischer Ebene in sofern wieder, als dass der Ton a' als reine Sinusfrequenz allein recht unscheinbar wirkt und wenig musikalische Aussagekraft hat. Zu diesem a' erklingt im 1. Vers bei den *Random-Beeps* zunächst ein cis', was nach obiger Ausführung vom Charakter zu einem Dur-Akkord tendiert. Der Balken des cis' erscheint daher in einem hell und fröhlich wirkendem grün, besonders die Kombination der Farben beider Balken hat eine freundliche, helle Ausstrahlung. Der Balken des c' erscheint in einem etwas gedeckteren grün, welches mehr Blau-Anteile hat, also Richtung türkis geht. Dieser Farbton unterscheidet sich nicht grundlegend von dem des Balkens des cis', da die dazugehörigen musikalischen Töne nur einen Halbtonschritt auseinander liegen. Zusammen mit der Farbe des Balkens für a' wirken beide Farben etwas gedeckter bzw. dunkler als die Kombination der Farben von a' und cis', der Farbkontrast ist wegen des höheren Blau-Anteils etwas weicher. In die Charakterisierung der Farben floss die oben genannte Kongruenz-Beziehung zwischen Temperatur und Farbe mit ein. „Warme Farben“ (rot/gelb) wurde in Verbindung mit „fröhlich“, „hell“ und „klar“ gebracht, „kalte Farben“ (blau/grün) mit „traurig“, „dunkel“ und

„weich“. Diese Idee beeinflusste die Wahl der Farbtöne, indem eine Farbe, die eher fröhlich wirken soll tendenziell mehr Rot- oder Gelbanteile erhielt und eine Farbe, die eher traurig wirken soll, mehr Blau- oder Grünanteile bekam. Aufbauend auf diese harmonischen und farblichen Beziehungen wurden nun Farbtöne für die anderen Tonhöhen gesucht. Dazu wurden zunächst die *Sinus-Chords* herangezogen, die zwar erst im 3. Vers einsetzen, jedoch maßgeblich für die Harmonik des Musikstücks sind. Dies ist einmal ein e', welches im Zusammenklang mit a' und cis' ein A-Dur Akkord ergibt, sowie ein f', welches zusammen mit a' und c' zu einem F-Dur Akkord wird. Gesucht sind also Farben, die die oben beschriebenen (Farb-)Zweiklänge jeweils zu einem Dreiklang ergänzen. Dabei ist zu beachten, dass durch das Hinzufügen eines e' zu a' und cis' die Erwartungshaltung erfüllt wird, ein A-Dur Akkord hören zu wollen. Hingegen bekommt der Zweiklang a'-c' durch das Hinzufügen eines f' eine andere Richtung als die erwartete: aus A-Moll wird F-Dur. Dies ist sowohl eine andere Tonart als auch ein anderes Tongeschlecht. Die Farbe des Balkens zu e' sollte demnach eine sein, die den Charakter des bestehenden Farb-Zweiklangs unterstreicht. Gewählt wurde ein warmes Gelb, welches eine hohe Leuchtkraft hat und zusammen mit den Farben Hellgrün des cis' und dem Blau des a' fröhlich, hell und klar wirkt.

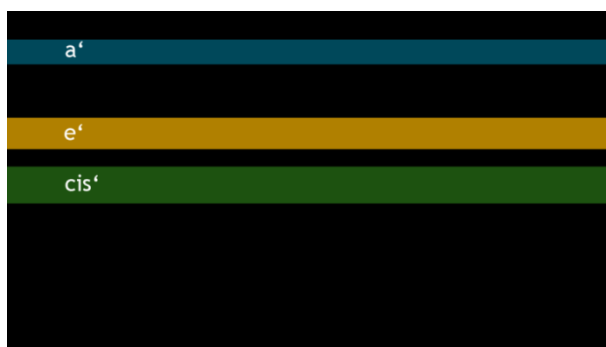


Abb. 7.4a

Der Balken des f' bekam ein leuchtendes Orange-Rot, um dem Farbzweiklang zwischen den Balken des a' und c' eine deutlich andere Richtung und mehr Leuchtkraft zu verleihen. Damit wird der Diskrepanz zwischen der Erwartungshaltung der Rezipienten und dem tatsächlichen Tongeschlecht und Tonart Rechnung getragen.



Abb. 7.4b

Somit ist jeder im Musikstück vorkommenden musikalischen Tonhöhe ein Farbton zugeordnet. Gleiche Töne in anderen Oktaven erhalten Balken in demselben Farbton, jedoch in veränderter Helligkeit und Sättigung. So hat das tiefe a einen Balken, dessen Farbton derselbe ist, die der des a', jedoch um einiges dunkler. Die Farben der Balken *Sinus-Beeps* verändern sich bei den Oktavierungen nach oben ebenfalls in Helligkeit und Sättigung, behalten jedoch denselben Farbton wie in der eingestrichenen Oktave.

Die **Rhythmus-Ebene** des Musikstücks wird im Bild durch weiße, animierte, abstrakte, geometrische Formen repräsentiert. Diese folgen der Idee kreuzmodaler Assoziationen. Dazu wurde jeder Klang dieser Ebene nach den Maßgaben M. Chions „Reduced Listening“ analysiert und anschließend versucht, ihn mit Attributen zu beschreiben, die originär der Dimension ‚geometrische Form‘ zugerechnet werden können. Dies sind vor allem Beschreibungen wie „rund, spitz, eckig, dick, dünn, groß, klein“ usw. Anhand dieser Beschreibungen wurden dann für jeden Klang eine zweidimensionale Formen entworfen, die eine visuelle Repräsentanz seiner Klangeigenschaften darstellt.

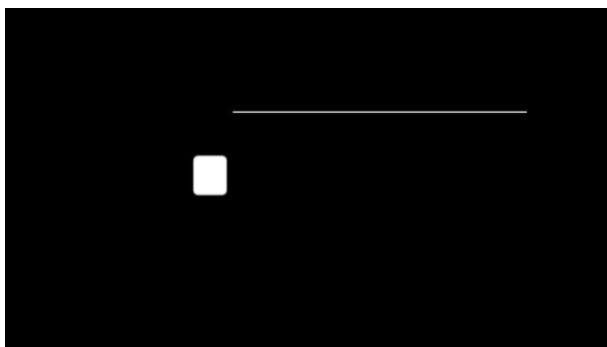


Abb. 7.5a

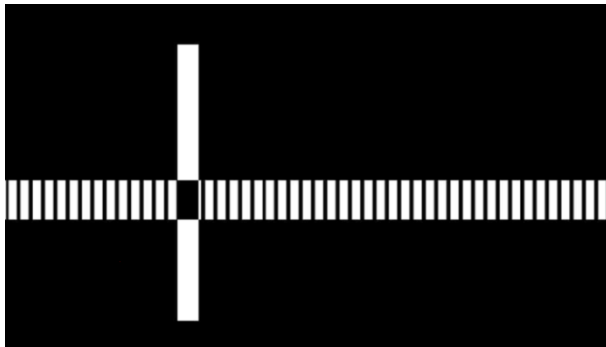


Abb. 7.5b



Abb. 7.5c



Abb. 7.5d

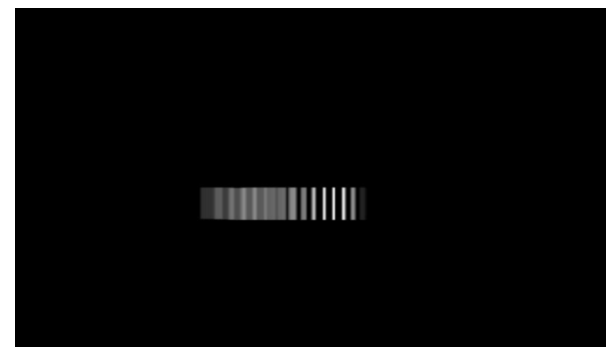


Abb. 7.5e



Abb. 7.5f

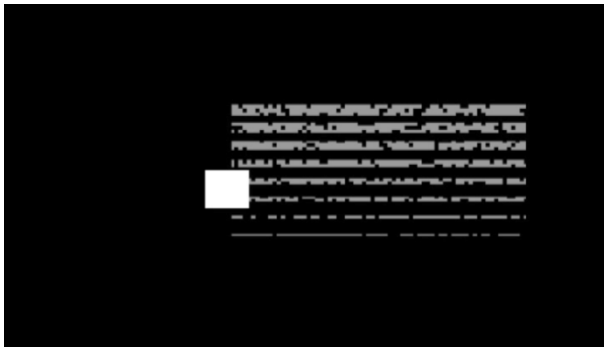


Abb. 7.5g

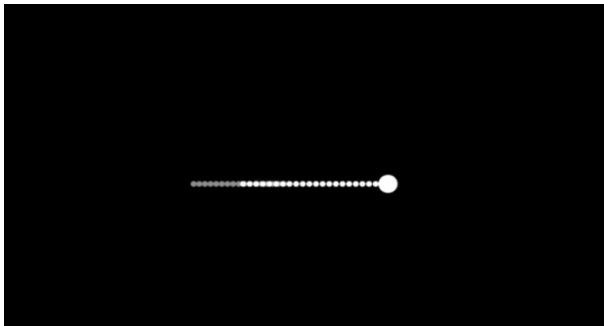


Abb. 7.5h

Abb. 7.5a-h zeigen verschiedene visuelle Umsetzungen der *Rhythm-Clicks*. Diese Formen erscheinen entweder für die Dauer dieses Klangs (meist nur für einen oder wenige Frames), oder sie sind animiert, d.h. verändern sich über die Zeit. Dies ist der Fall, wenn sich der dazugehörige Klang ebenfalls über die Zeit verändert.

Die Elemente dieser Ebene sind bewusst in Weiß bzw. verschiedenen Graustufen gehalten. Damit grenzen sie sich nicht nur ästhetisch von der visuellen Umsetzung der tonalen Ebene ab, sondern folgen auch dem theoretischen Konzept, die Rhythmus- und die tonale Ebene frei von Einflüssen der jeweils anderen zu halten (vgl. Kap. 7.2.1, Tonale Ebene und Rhythmus-Ebene).

Wie oben beschrieben teilt sich die Rhythmus-Ebene in zwei Unterebenen, die der *Rhythm-Clicks* und die des *Rhythm-Beats*. Die Erste zeichnet sich vor allem dadurch aus, dass kein Klang-Element im Musikstück zweimal identisch vorkommt, d.h. jeder Klang dieser Ebene ist einzigartig. Dementsprechend sind auch die visuellen Elemente dieser Ebene Unikate, was ihre Form, Helligkeit, Animation und Position betrifft. Die musikalische Ebene des *Rhythm-Beats* besteht hingegen aus immer denselben Klängen für Bass Drum, Snare Drum und HiHat. Dementsprechend wiederholen sich auch die diesen Klängen

zugeordneten Formen, womit auch ihre räumliche Position im Bild konstant bleibt.

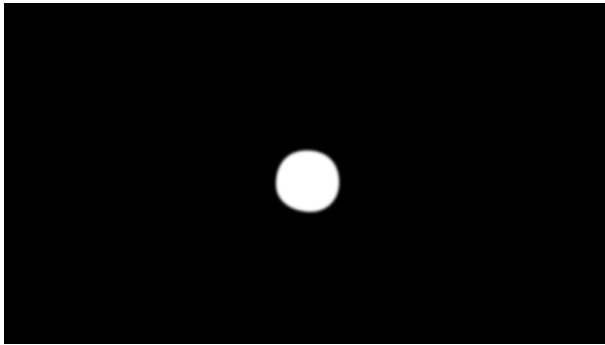


Abb. 7.6a



Abb. 7.6b

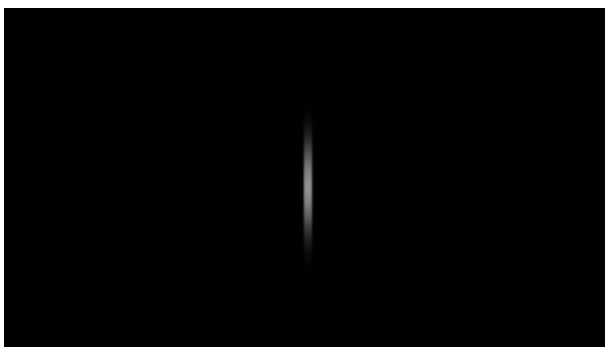


Abb. 7.6c

Abb. 7.6a-c zeigen je eine Bewegungsphase der visuellen Elemente des *Rhythm-Beats* bestehend aus Bass Drum, Snare und HiHat (von oben nach unten). Der Grund, warum alle drei Formen im Bild dieselbe Position haben, liegt vor allem in dem hohen Tempo des Beats und dem schnellen Aufeinanderfolgen der einzelnen Schläge. Durch die räumliche Deckung werden diese drei Elemente als zu einer Unterebene gehörend wahrgenommen und sind daher viel besser identifizierbar und den von ihnen repräsentierten Klängen zuzuordnen, als wenn sie unterschiedliche Positionen einnehmen. Dies haben Tests mit verschiedenen Anordnungen ergeben.

### 7.3 Zusammenfassung

Der praktische Teil dieser Diplomarbeit besteht aus dem Clip ‚*congruence*‘, einer Musikvisualisierung von drei Minuten Länge. Sie ist damit ein praktischer, gestalterischer Versuch, Beziehungen und Interaktionen zwischen akustischer und visueller Wahrnehmung in Form von Ton und Bild als deren mediale Repräsentationen darzustellen. In den Gestaltungsprozess dieses Clips flossen einige der zuvor wissenschaftlich dargelegten Erkenntnisse über multisensorische Prozesse ein, das entstandene Werk bildet diese demzufolge ab. Dies betrifft sowohl die Musik als auch das Bild: Beide Ebenen wurden bereits im Entstehungsprozess als interagierend angesehen und dementsprechend gestaltet. Das heißt, der Clip konzentriert sich vorrangig auf Interaktionen zwischen akustischer und visueller Wahrnehmung und damit zwischen Ton und Bild. Die jeweiligen Ebenen haben daher einzeln betrachtet sehr viel weniger Gehalt, als gemeinsam dargeboten.

Die Ergebnisse folgender wissenschaftlicher Untersuchungen und theoretischer Betrachtungen wurden aufgegriffen: (1) Kongruenz-Effekte in schnellen Klassifizierungen, (2) kreuzmodale Assoziationen (synästhetische Metaphern), (3) Ventriloquist-Effekt und (4) *Reduced listening*. Jeder dieser Themen wird durch eine gestalterische Ebene im Clip wiedergegeben: Die Kongruenz-Effekte spiegeln sich in der tonalen Ebene wieder. Entsprechend der Kongruenz-Beziehungen zwischen akustischen und visuellen Perzepten sind Helligkeit, Dicke und vertikale Position der Bildelemente auf Tonhöhe und Lautstärke der Töne der Musik abgestimmt. Kreuzmodale Assoziationen werden in der Rhythmus-Ebene repräsentiert. Entsprechend der Klangeigenschaften der *Rhythm-Clicks* und des *Rhythm-Beats* sind die visuellen Elemente dieser Ebene in ihrer geometrischen Form gestaltet. *Reduced listening* spielte hierbei eine wichtige Rolle: Diese Methode wurde benutzt, um explizit die originären Klangeigenschaften umzusetzen und nicht die visuellen Eigenschaften einer möglichen Schallquelle. Der Ventriloquist-Effekt ist insofern wichtig, als dass eine Zuordnung der visuellen und akustischen Elemente zueinander nur dann gewährleistet ist, wenn sie synchron dargeboten werden.

Der Clip arbeitet mit Abstraktion, die Bildebene indes mehr, als der Ton. Es sind zwar ausschließlich synthetisch generierte Klänge zu hören, was besonders deutlich an der tonalen Ebene auffällt, da reine Sinustöne am stärksten den

Hörgewohnheiten widersprechen; Allerdings wurde auf musikalische Formen bewusst nicht verzichtet, um die Aufmerksamkeit des Rezipienten nicht zu sehr an unerwartete akustische Ereignisse zu binden und damit von der Interaktion zwischen Ton und Bild abzulenken. Das Bild ist indes in allen Aspekten abstrakt, es enthält keinerlei Realsequenzen, keine Animations- oder Comicsequenzen, die an reale Szenen oder Objekte angelehnt sind. Ferner gibt es keine Bildschnitte und keinen perspektivisch abgebildeten Raum. Es wird lediglich mit geometrischen Formen und Farben gearbeitet. Der Grund dafür liegt darin, dass mit der praktischen Arbeit in erster Linie perzeptuelle Interaktionen der Wahrnehmungs-Modalitäten aufgegriffen werden sollten und keine konzeptuellen, denn diese standen im Mittelpunkt der Fragestellung dieser Diplomarbeit. Auf perzeptuellen Stufen der Verarbeitung existieren, wie erläutert wurde, zunächst einmal nur Reize. Erst auf höheren Stufen kommt es zur Integration von Objekten und dann zur Identifikation und Wiedererkennung. Unwillkürliche Assoziationen der Rezipienten dieses Clips sind dennoch nicht auszuschließen, denn besonders im Visuellen können selbst einfachste geometrische Formen bereits Wahrnehmungskonzepte von Objekten ansprechen. Diese zu verhindern war indes nicht das Ziel.

## 8 FAZIT UND AUSBLICK

Ziel dieser Diplomarbeit war es, Zusammenhänge und Interaktionen zwischen akustischer und visueller Wahrnehmung zu finden, die bei der Rezeption audiovisueller Medien eine Rolle spielen. Das Forschungsinteresse lag dabei vor allem auf frühen, perzeptuellen Wahrnehmungsebenen, auf denen semantische, narrative und linguistische Konzepte keine, oder eine untergeordnete Rolle spielen. Gibt es Zusammenhänge zwischen Klangeigenschaften und Bildeigenschaften, und sind sie auf perzeptuelle Prozesse zurückzuführen? Es sollte ferner gezeigt werden, dass Ästhetik ein Bestandteil menschlicher Wahrnehmung, und damit der Konstruktion von Erleben ist.

Der praktische Teil der Diplomarbeit setzt einige der zuvor dargelegten theoretischen Ergebnisse in einem Visual Music Clip gestalterisch um. Damit werden Wahrnehmungszusammenhänge in mediale Ton-Bild-Relationen übertragen. Ziel dieser Transformation war es herauszufinden, ob es eine Form der perzeptuellen Ästhetik gibt. Eine solche Ästhetik wäre eine, die sich auf das Zusammenspiel der Sinne bezieht und nicht, oder möglichst wenig, auf höhere konzeptuelle Repräsentationen, wie z.B. von Objekten, Situationen und Emotionen. Sie wäre damit unabhängig von Interpretationen und von Bedeutungen, die der jeweilige Rezipient den Reizen zuschreibt.

### *Aufbau der Diplomarbeit*

Nach der Einleitung in Kapitel 1 beschäftigte sich Kapitel 2 mit Multimedia, seinen Formen, seiner kulturellen, soziologischen und psychologischen Bedeutung sowie seiner Ästhetik. Kapitel 3 behandelte die menschliche Wahrnehmung, von der Reizverarbeitung der Sinnesorgane, bis zu neuronalen Netzen und kognitiver Verarbeitung. Kapitel 4 beschrieb multisensorische Prozesse zwischen akustischer und visueller Wahrnehmung. Kapitel 5 beschäftigte sich mit Synästhesie und den möglichen Rückschlüssen auf allgemeine Wahrnehmungsprozesse. In Kapitel 6 wurden ausgewählte Thesen des Medientheoretikers M. Chion analysiert und mit den Ergebnissen aus Kapitel 4 verglichen. Kapitel 7 hat den praktischen Teil dieser Diplomarbeit vorgestellt und erläutert, welche wissenschaftlichen Erkenntnisse in den Gestaltungsprozess eingeflossen sind.

### *Ergebnisse*

Die theoretischen Untersuchungen brachten eine Reihe an Erkenntnissen über die Funktion, den Nutzen und die Ästhetik von Medien, über die menschliche Wahrnehmung im Allgemeinen, sowie der akustischen und visuellen im Speziellen, insbesondere über Interaktionen zwischen diesen Modalitäten. Es stellte sich heraus, dass es Zusammenhänge zwischen Klangeigenschaften und Bildeigenschaften in dem Sinne gibt, als dass es zwischen den Sinnen Hören und Sehen z.B. Beeinflussungen in der qualitativen Beurteilung der jeweiligen Reize gibt. Das heißt, die Anwesenheit eines Tons kann die qualitative Beurteilung der Helligkeit eines Lichtes beeinflussen. Die Umsetzung einiger dieser Erkenntnisse im praktischen Teil dieser Arbeit in einen Visual Music Clip zeigen, dass sich aus Funktionsprinzipien der menschlichen Wahrnehmung mediale Ton-Bild-Relationen ableiten lassen. Ein Transfer wissenschaftlicher Ergebnisse in Gestaltungsprinzipien für audiovisuelle Medien scheint demnach möglich. Die Frage nach einer Art perzeptuellen Ästhetik kann zwar prinzipiell bejaht werden, bedarf allerdings einer differenzierten und kritischen Betrachtung. Zunächst werden jedoch die Ergebnisse der einzelnen Kapitel genauer erläutert.

Die Betrachtungen von Multimedia in Kapitel 2 ergaben, dass Medien technische Mittel der Kommunikation zwischen Menschen sind, die die Reichweite menschlicher Wahrnehmung erweitern. Sie übernehmen in einer Gesellschaft entscheidende soziale, politische und kulturelle Funktionen und haben eine psychologische Dimension. Da Wahrnehmung als Sinnestätigkeit unter Einbeziehung von Ästhetik beschrieben werden kann, ist sie wichtiger Bestandteil der Kommunikation und damit auch von Medien. Multimedia, das simultane Ansprechen mehrerer Sinne, erfuhr durch die Entwicklung des Kinos und des Fernsehens eine globale Verbreitung und machte audiovisuelle zu Leitmedien. In der ästhetischen Entwicklung audiovisueller Medien nehmen das Musikvideo, sowie die Kunstform Visual Music eine besondere Rolle ein. Beide Formen stellen enge Bezüge zwischen Musik und bewegtem Bild her und beeinflussen andere Genres und Medien.

Die Ergebnisse aus Kapitel 3 betreffen die Funktionsweise, Fähigkeiten und Einschränkungen der Sinnesorgane Auge und Ohr, neuronale Verarbeitung der

Reize sowie Integrations- und Funktionsprinzipien des menschlichen Gehirns bei der Wahrnehmung. Demnach reagieren die verschiedenen Sinnesorgane jeweils auf andere physikalische Reizdimensionen, das Ohr auf Schall und das Auge auf Licht. Diese haben verschiedene Ursachen und vermitteln daher unterschiedliche Informationen. Auge und Ohr sind den Lebensbedingungen des Menschen angepasst und bilden einen Ausschnitt der physikalischen Realität ab. Demzufolge kommt es bereits auf der Reizebene zu einer Bewertung. Hinzu kommt, dass Auge und Ohr über das von ihnen abgebildete Spektrum nicht gleichmäßig empfindlich sind und darüber hinaus unterschiedlich genaue Informationen über Raum und Zeit liefern. Die Sinnesorgane setzen physikalische Reizungen in Nervensignale um, die zum Gehirn weitergeleitet und dort verarbeitet werden.

Neuronale Prozesse sind in einer Netzarchitektur organisiert, in denen Informationen in Form von Aktivierungsmustern der Netz-Knotenpunkte (Neuronen) repräsentiert und weitergeleitet werden. Die eingehenden Sinnesreizungen werden unter Einbeziehung von Vorwissen und Annahmen zu konsistentem Erleben konstruiert. Funktionsprinzipien dieser neuronalen Netze sind Spezialisierung, Ordnung, Parallelverarbeitung sowie laterale und vertikale Integration. Wahrnehmung geschieht auf verschiedenen hierarchischen Stufen, der sensorischen, der Stufe der Wahrnehmungsorganisation, sowie der Stufe Identifikation und Wiedererkennung.

In Kapitel 4 ergaben Untersuchungen von kognitionswissenschaftlichen Studien eine Vielzahl von Interaktionen zwischen akustischer und visueller Wahrnehmung.

Bei Prozessen der Aufmerksamkeit spielen kreuzmodale Prozesse eine wichtige Rolle. Reizungen einer Modalität können zu einer erhöhten Sensitivität anderer Sinne in einem bestimmten Punkt führen, sowohl offen, bewusst, als auch verdeckt, unbewusst. Besonders interessant ist, dass die Reizung einer Modalität zu erhöhter Aktivität von Hirnarealen anderer Modalitäten führen kann.

Es gibt Modulationen zwischen akustischer und visueller Wahrnehmung. Das heißt, Reize in einer Modalität können die Wahrnehmung einer anderen Modalität qualitativ verändern. Einflüsse akustischer Reize auf das Visuelle zeigen sich auf folgende Arten: Sie können die zeitliche Auflösung des Sehens

erhöhen, die empfundene Intensität eines visuellen Reizes steigern, sowie die Identifikation visueller Zielreize verbessern. In mehrdeutigen Situationen mit bewegten Objekten können sie eine Veränderung der Interpretation des Gesehenen hervorrufen. Bewegte akustische Reize sind in der Lage Bewegungswahrnehmungen von statischen oder fluktuierend bewegten (ohne eindeutige Richtung) visuellen Stimuli hervorzurufen. Darüber hinaus ist es möglich, dass Schallereignisse die Struktur des Visuellen auf perzeptueller Ebene verändern, wie der „*illusory flash effect*“ zeigt. Einflüsse visueller Reize auf das Hören zeigen sich beim Bauchredner- und McGurk-Effekt. Der Bauchredner-Effekt bewirkt, dass Schallquellen falsch lokalisiert werden, da sie räumlich an einen passenden visuellen Reiz anhaften. Dieses Phänomen macht die Immersion des Kinos und Fernsehens überhaupt erst möglich, da die Schallquelle bei der Wiedergabe dieser Medien örtlich nicht deckend mit der Bildfläche sein kann. Der McGurk-Effekt zeigt, dass das Lippenlesen das akustische Sprachverständnis modulieren kann.

Studien zur multisensorischen Bewegungswahrnehmung zeigen, dass es von der Situation und der Beschaffenheit der Stimuli abhängt, ob die Bewegung in einer Modalität durch eine andere beeinflusst oder gestört wird. Falls akustische und visuelle Reize simultan stattfinden, kann es selbst bei entgegengesetzter Bewegungsrichtung zu einer Anhaftung des Akustischen an das Visuelle kommen. Durch die Untersuchung von Interaktionen in schnellen Klassifizierungen lassen sich Beziehungen zwischen Reiz-Dimensionen feststellen. Diese können sich in Form von gegenseitiger Hemmung oder als Kongruenzen äußern. Als kongruent können solche Dimensionen bezeichnet werden, die in einer bestimmten Weise passend zueinander verlaufen. Dies ist multimodal zwischen Tonhöhe und Position, Tonhöhe und Farbe (schwarz/weiß), Tonhöhe und Helligkeit, Lautstärke und Helligkeit sowie Tonhöhe und Form (rund/eckig) der Fall. Dabei stellt sich heraus, dass viele Synästhesien Ähnlichkeiten mit den gefundenen Kongruenzen aufweisen.

Ergebnisse aus Kapitel 5 sind, dass Synästhesien statistisch selten auftretende Phänomene sind, die sich als tatsächliches sensorisches Erleben äußern und nicht etwa einer übersteigerten Phantasie oder Assoziationen zuzuschreiben sind. An manchen Graphem-Farb-Synästheten ließ sich nachweisen, dass das

Übersprechen zwischen den Sinnen auf perzeptuellen Ebenen abläuft. Aus einem Netzwerk an bekannten Fakten, wie z.B. dass Synästhesien gehäuft unter Künstlern und Schriftstellern sowie in Familien auftreten, lässt sich die These entwickeln, dass übermäßige Kreuzverbindungen zwischen sensorischen Hirnarealen, verursacht durch eine genetische Mutation, der Auslöser sein könnten. Wenn Synästhesien durch Kreuzverbindungen sensorischer Areale ausgelöst werden, könnten Kreativität und Metaphern Kreuzverbindungen zwischen konzeptuellen Arealen entspringen. An Sprache lassen sich derartige Verbindungen vor allem zwischen Konzepten und Perzepten ablesen: „Süßes Mädchen“ oder „heller Ton“ können als synästhetische Metaphern oder kreuzmodale Assoziationen bezeichnet werden.

Die Analyse einiger Thesen des Medientheoretikers Michel Chion in Kapitel 6 ergaben Parallelen zu kognitionswissenschaftlichen Erkenntnissen, insbesondere im Bezug auf zeitliche Aspekte der Filmgestaltung, wie schnelle Bewegungsabläufe, oder die Zeitlichkeit einer Szene. Von besonderem Interesse im Hinblick auf den praktischen Teil dieser Arbeit sind die drei Hör-Modi, die Chion unterscheidet. Seine Idee vom reduzierten Hören, welches die genuinen Klangeigenschaften eines Schallereignisses in den Blick nimmt, wird dort mit den in Kapitel 5 erarbeiteten synästhetischen Metaphern in Verbindung gebracht. Denn Chion stellt fest, dass es zur Beschreibung eben jener Klangeigenschaften, neben der Tonhöhe und Lautstärke, kaum sprachliche Mittel gibt, um diese zu charakterisieren. Hier können synästhetische Metaphern, wie z.B. „rauer Klang“ helfen.

Kapitel 7 erläutert die Entstehung der praktischen Arbeit. In dem Visual Music Clip wurden die Ergebnisse folgender wissenschaftlicher Themen aufgegriffen: Kongruenz-Effekte in schnellen Klassifizierungen, kreuzmodale Assoziationen (synästhetische Metaphern), der Ventriloquist-Effekt und „*Reduced listening*“. Die Kongruenz-Beziehungen zwischen Tonhöhe und Position, Tonhöhe und Größe, sowie Lautstärke und Helligkeit sind im Clip durch die ‚Tonale Ebene‘ umgesetzt. Das Ergebnis sind horizontale Balken simultan zu den Tönen der Musik: Höheren Tönen sind höhere und schmalere Balken zugeordnet, tieferen Tönen tiefere und breitere Balken. Lauteren Tönen sind hellere, leiseren Tönen

dunklere Balken zugeordnet. Dadurch bekommen die schnell wechselnden Töne eine visuelle Entsprechung, die als passend empfunden werden kann. Bei intensivem Betrachten kann der Eindruck entstehen, dass Töne und Balken eine Einheit bilden und keine voneinander getrennten Entitäten sind.

Kreuzmodale Assoziationen bilden das Gestaltungsprinzip der Rhythmus-Ebene im Clip. Jedes Geräusch dieser Ebene erhält im Bild eine animierte, geometrische Form. Die Methode des *Reduced listening* wurde dafür genutzt, die originären Klangeigenschaften dieser Geräusche umzusetzen und nicht etwa die visuellen Eigenschaften einer möglichen Schallquelle (z.B. visuelles Symbol für eine Rassel). Wie auf der tonalen Ebene, so verbinden sich hier ebenfalls Ton und Bild zu einer Einheit. Bei der Rezeption lassen sich auf beiden Ebenen (tonaler und Rhythmus-Ebene) zwar jeder Zeit durch Bewusstmachen Ton und Bild voneinander separiert wahrnehmen, dennoch besteht eine intensive und deutlich erkennbare Verbindung zwischen den Elementen. Diese wird insgesamt zusammengehalten durch eine genaue Synchronität zwischen Ton und Bild. Synchronität ist ein wichtiger Faktor für die Anhaftung des Akustischen an das Visuelle beim Ventriloquist-Effekt.

Damit wurden wissenschaftliche Ergebnisse in Gestaltungsprinzipien für audiovisuelle Medien transformiert und hierdurch Interaktionen zwischen den Modalitäten der menschlichen Wahrnehmung in mediale Ton-Bild-Relationen umwandelt.

Dies kann gleichermaßen als Ästhetik dieses Clips betrachtet werden, eine abstrakte Ästhetik, die Tönen Helligkeit und Position verleiht, sowie Geräuschen Form und Bewegung gibt, und dabei Prinzipien der Reizverarbeitung und Wahrnehmungsorganisation folgt.

### ***Kritisch Betrachtung und Bewertung der Ergebnisse***

Die Ergebnisse weisen eine Reihe von Interaktionen zwischen akustischer und visueller Wahrnehmung aus. Anhand von Analyse und Vergleich mit Thesen des Medientheoretikers Michel Chion auf der einen Seite, sowie der praktischen Umsetzung auf der anderen Seite konnte gezeigt werden, dass diese Interaktionen in der Wahrnehmung als Beziehungen zwischen medialem Ton und Bild interpretiert werden können. Es lassen sich durch Kongruenz-Beziehungen die Klangeigenschaften Tonhöhe und Lautstärke mit den Bildeigenschaften

Helligkeit, geometrische Form, Farbe (schwarz/weiß) und Position in Beziehung setzen. Durch synästhetische Metaphern lassen sich Verbindungen zwischen akustischen Attributen wie Definition, Reinheit, Volumen und visuellen Eigenschaften wie Kontrast, Animation, Textur und Farbe (Helligkeit, Sättigung, sehr bedingt Farbton) sowie geometrische Form herstellen. Akustische Entsprechungen für den Farbton mit universeller Gültigkeit konnten nicht gefunden werden. Kongruenz-Beziehungen gibt es lediglich zwischen Temperatur (warm-kalt) und Farbe (Anm.: Der technische Begriff ‚Farbtemperatur‘ spiegelt genau diesen Zusammenhang wider). Untersuchungen an Synästheten, bei denen Farbwahrnehmung durch akustische oder andere Reize ausgelöst werden zeigen, dass es eine große individuelle Streuung gerade auch in Hinblick auf die Farbtöne gibt. Das heißt, ein bestimmter Ton, Name oder Buchstabe löst bei einem Synästheten die Farbe Blau aus, bei einem weiteren Grün und bei dem nächsten Gelb oder Rot. Es konnte bisher keine Studie Tendenzen feststellen, die Zuordnung von Farbtönen scheint demzufolge willkürlich zu sein.

Direkte visuelle Entsprechungen zu musikalischen Parametern wie Melodik, Harmonik, Rhythmik und Dynamik konnten auf diese Art ebenfalls nicht gefunden werden. Es ist jedoch durchaus möglich, entsprechende Zusammenhänge aus den gewonnenen Ergebnissen indirekt abzuleiten, wie im praktischen Teil geschehen. Hier wurden Beziehungen zwischen Farbtönen und Harmonik hergestellt. Grundlage dessen ist, dass die emotionale Wirkung von Harmonik -Dur wird als hell und fröhlich, Moll als dunkel und traurig beschrieben- mit der Temperatur-Farben Kongruenz -rot/gelb/orange sind kongruent zu warm, grün/blau sind kongruent zu kalt- in Verbindung gebracht wurde. Es wurde dann warm (gelb/rot) mit fröhlich und hell (Dur), sowie kalt (grün/blau) mit traurig und dunkel (Moll) assoziiert. Der so hergestellte Zusammenhang zwischen musikalischer Harmonik und Farb-Harmonik wurde allerdings nicht analog umgesetzt, sondern floss in die Feinabstimmung der Farbtöne ein.

Ferner könnte die Melodik über die Tonhöhe in ein mittelbares Verhältnis zu Position und zu Größe gestellt werden. Melodik kann steigend oder fallend sein, entsprechend kann Position steigend oder fallend sein, sowie Größe zu- oder abnehmen. Dies wurde im Clip nicht umgesetzt, da das Musikstück keine Melodik

enthält. Auf ähnliche Weise ließe sich Dynamik (Lautstärke-Differenz) mit visuellem Kontrast (Helligkeits-Differenz) verknüpfen.

Zur visuellen Umsetzung von Rhythmik wären zwei Ansätze vorstellbar: Das bewegte Bild lässt sich auf zwei Arten „rhythmisieren“, einmal örtlich und einmal zeitlich. Örtlich „Rhythmisierung“ würde sich über die Flächigkeit oder Strukturiertheit des Bildes ausdrücken, technisch gesehen über die Ortsfrequenz. Zeitliche „Rhythmisierung“ erfährt das Bild über Bild-Schnitte, bzw. die zeitliche Varianz von Strukturen innerhalb des Bildes. Zu dieser Überlegung gibt es bisher jedoch keine kognitionswissenschaftlichen Untersuchungen.

Da die Gestaltung dieses Visual Music Clips anhand der dargestellten Prinzipien und Interaktionen der menschlichen Wahrnehmung realisiert wurde, kann gesagt werden, dass ihm eine Art perzeptueller Ästhetik innewohnt. Tatsächlich ist es so, dass der Clip semantische, narrative und linguistische Wahrnehmungskonzepte nicht direkt anspricht. Allerdings lassen sich diese Ebenen der Wahrnehmung nicht isolieren oder gar ausschalten. Es ist demnach unvermeidlich, dass derartige konzeptuelle Strukturen dennoch aktiviert werden. Dies zeigen auch subjektive Einschätzungen von Menschen, die diesen Clip sehen. Sinngemäß zitiert sagte jemand z.B.: „Bei dieser einen Form, da sehe ich immer eine Jalousie.“ Dies lässt sich darauf zurückführen, dass Synthese auf Basis von Erfahrung und Vorwissen ein zentrales Element von Wahrnehmungen ist. Der Mensch ist darauf spezialisiert, den Objekten in seiner Umgebung einen Sinn zu geben, was zwingend erforderlich für sein Überleben ist. Trotzdem zeigt der Clip, dass Ästhetik zu einem Teil auch perzeptueller Natur ist und auf Prinzipien der Reizverarbeitung beruht.

### *Ausblick*

Ein weiterer Schritt zur Bearbeitung des Themas dieser Diplomarbeit wäre eine empirische Untersuchung zur Rezeption des entstandenen Visual Music Clips. Es könnte darin z.B. nach der emotionalen Einstellung gegenüber dem Clip gefragt werden, oder auch bei einer Gruppe Testpersonen eine „Nur-Ton-Vorführung“, bei einer weiteren eine „Nur-Bild-Vorführung“ und bei einer dritten der gesamte Clip präsentiert werden. Diese Gruppen könnte man anschließend z.B. in

Hinblick auf das Erkennen und Erinnern von akustischen oder visuellen Elementen vergleichen. Prägen sich beispielsweise visuelle Elemente besser ein, wenn sie zusammen mit kongruentem Ton präsentiert werden?

In dieser Diplomarbeit wurden vorwiegend Interaktionen auf frühen, perzeptuellen Stufen der Verarbeitung betrachtet. Viele spannende Fragen ergeben sich, wenn wissenschaftliche Studien im Hinblick auf audiovisuelle Medien analysiert würden, die höhere Stufen der Verarbeitung mit einbeziehen, wie z.B. die Objekterkennung, oder auch die emotionale Wirkung von Perzepten. Eine mögliche Fragestellung könnte lauten, welche Interaktionen zwischen den Sinnen sich bei komplexeren Stimuli ergeben.

Eine weitere Frage betrifft die neuronale Farbkodierung. Warum ist synästhetisches Farberleben individuell so unterschiedlich? Warum lassen sich Kongruenz- oder Interferenz-Effekte bei dieser Dimension nur zum Temperaturempfinden feststellen und gibt es einen Zusammenhang zur individuellen Streuung bei Synästheten?

Bisher wurden durch schnelle Klassifikationen Interferenz- und Kongruenz-Effekte zwischen simplen (akustischen und visuellen) Reizdimensionen, wie Tonhöhe, Lautstärke, Helligkeit und Position untersucht. Es wäre denkbar, derartige Untersuchungen auch für komplexere Reizdimensionen durchzuführen. Diese könnten musikalische, psychoakustische und visuell gestalterische Dimensionen sein, wie z.B. die „Klang-Definition“, „Klang-Farbe“, „Rauigkeit“ und „Reinheit“ von akustischen Reizen oder „Textur“, „Kontrast“, „Animation“, „Struktur“ und „Farbigkeit“ von visuellen Reizen. eine mögliche Fragestellung wäre, ob es beispielsweise Kongruenz-Effekte zwischen der „Rauigkeit“ eines Klangs und der „Textur“ eines Bildes gibt.

Die Wahrnehmung audiovisueller Medien ist von kultureller, sozialer und politischer Bedeutung und hat gestalterische, ästhetische und psychologische Dimensionen. Moderne Hirnforschung verändert zunehmend die Sicht auf den Menschen, sie liefert Erkenntnisse über menschliches Denken, Handeln, Fühlen und über die Wahrnehmung. Ihre Ergebnisse können daher aufschlussreich für

---

Medienschaffende und -Theoretiker sein, ebenso können Mediengestalter viele interessante Fragen für die Kognitionsforschung aufwerfen.

## 9 LITERATURVERZEICHNIS

- Altenmüller, Eckart (2005): *Musik - die Sprache der Gefühle? Neurobiologische Grundlagen emotionaler Musikwahrnehmung*. In: Schnell, Ralf [Hg.] (2005): *Wahrnehmung - Kognition - Ästhetik*. Bielefeld: Transcript, 139-155
- Baron-Cohen, S. et al. (1993): *Coloured speech perception: Is synaesthesia what happens when modularity breaks down?* In: *Perception*, 22, 419-426.
- Bruce, Vicki/ Green, Patrick R. /Georgeson, Marl A. (2003): *Visual perception: physiology, psychology and ecology*. 4. ed., Hove [u.a.]: Psychology Press
- Carlson, Neil A. (2004): *Physiologische Psychologie*. 8. aktualisierte Auflage, München: Pearson Studium
- Chion, Michel (1994): *Audio-Vision*. New York: Columbia University Press
- Cytowic, Richard E. (1995): *Synesthesia: Phenomenology and Neuropsychology*. In: *Psyche* 2(10), Juli 1995
- Daniels, Dieter (2004): *Sound & Vision in Avantgarde & Mainstream*. Verfügbar unter: <[http://www.medienkunstnetz.de/themen/bild-ton-relationen/sound\\_vision/1/](http://www.medienkunstnetz.de/themen/bild-ton-relationen/sound_vision/1/)>. Abgerufen am 5.3.2009
- Day, Sean A. (2009): *Types of synesthesia*. Internetquelle: <<http://home.comcast.net/~sean.day/html/types.htm>>. Abgerufen am 21.2.2009
- Eska, Georg (1997): *Schall und Klang*, Basel: Birkenhäuser
- Fahle, Manfred (2005): *Ästhetik als Teilaspekt bei der Synthese menschlicher Wahrnehmung*. In: Schnell, Ralf [Hg.] (2005): *Wahrnehmung - Kognition - Ästhetik*. Bielefeld: Transcript, 61-109
- Fischer, Agnes /Hubert, Ines (2005): *Visuelle Musik - Von der Avantgarde über das Musikvideo zum VJing*. Verfügbar unter: <<http://server4.medienkomm.uni-halle.de/filmsound/kap1-4.htm>>. Abgerufen am 5.3.2009
- Goldstein, Eugen Bruce (2008): *Wahrnehmungspsychologie*. 7. Auflage, Berlin [u.a.]: Spektrum, Akademischer Verlag
- Golenhofen, Klaus (2006): *Basislehrbuch Physiologie*. 4., überarbeitete und aktualisierte Auflage, München [u.a.]: Elsevier, Urban & Fischer
- Haase, Günther (1973): *Physik für Mediziner - Physikalische Grundlagen*, Frankfurt am Main: Akademische Verlagsgesellschaft
- Hartmann, Frank (2008): *Multimedia*. Wien: Facultas.wuv
- Haverkamp, Michael (2003): *Visualisierung auditiver Wahrnehmung*. Vortragsskript von der Jahrestagung der Deutschen Gesellschaft für Akustik 2003, Verfügbar unter:

<[http://www.michaelhaverkamp.mynetcologne.de/visualisierung\\_text.pdf](http://www.michaelhaverkamp.mynetcologne.de/visualisierung_text.pdf)>. Abgerufen am 16.2.2008

- Heyrman, Hugo (2005): *Art and Synesthesia: in search of the synesthetic experience*, Lecture presented at the First International Conference on Art and Synesthesia, Juli 2005, Almeria, Spanien, Verfügbar unter: <<http://www.doctorhugo.org/synaesthesia/art/index.html>>. Abgerufen am 28.1.2008
- Itten, Johannes (2001): *Kunst der Farbe*, Gekürzte Studienausgabe, 27. Auflage, Ravensburg, Ravensburger Buchverlag
- Jerrentrup, Andreas (2008): *Physik für Mediziner*. 18. Auflage, Stuttgart [u.a.]: Thieme
- Keazor, Henry /Wübbena, Thorsten (2007): *Video thrills the radio star*. 2., überarbeitete Auflage, Bielefeld: Transcript
- Marks, Lawrence E. (2004): *Cross-Modal Interactions in Speeded Classifications*. In: Calvert, G.; Spence, C.; Stein, B. E. (Eds.): *The Handbook of Multisensory Processes*, Massachusetts: MIT press, 85-105
- Massaro, Dominic W. (2004): *From Multisensory Integration to Talking Heads and Language Learning*. In: Calvert, G.; Spence, C.; Stein, B. E. (Eds.): *The Handbook of Multisensory Processes*, Massachusetts: MIT press, 153-175
- Meschede, Dieter (2006): *Gerthsen Physik*. 23. überarb. Auflage, Berlin [u.a.]: Springer
- Moritz, William (1987): *Der Traum von der Farbmusik*. In: Bódy, Veruschka [Hg.]: *Clip, Klapp, Bum*. Köln: Dumont
- Moritz, William (1993): *Oskar Fischinger: Leben und Werk*. In: Gehr, Herbert [Hg.]: Deutsches Filmmuseum in Frankfurt am Main: *Optische Poesie*. Frankfurt am Main, 1993, 7-90
- Pinel, John P.J. /Pauli, P. [Hg.] (2007): *Biopsychologie*, 6. Aktualisierte Auflage 2007, München, Pearson Studium
- Ramachandran, V. S. /Hubbard, E. M. (2001a): *Psychophysical investigations into the neural basis of synaesthesia*. In: Proceedings of the Royal Society, London, B, 2001, 268: 979-983
- Ramachandran, V. S. /Hubbard, E. M. (2001b): *Synaesthesia – A Window Into Perception, Thought and Language*. In: Journal of Consciousness Studies, 8, No. 12, pp. 3-34
- Ramachandran, V. S. /Hubbard, E. M. (2003): *Hearing Colors, Tasting Shapes*. In: Scientific American, Vol. 288(5), 53-59, May 2003
- Ramachandran, V. S. /Hubbard, E. M. /Butcher, P. A. (2004): *Synesthesia, Cross-Activation, and Foundations of Neuroepistemology*. In: Calvert, G.; Spence, C.; Stein, B. E. [Eds.] (2004): *The Handbook of Multisensory Processes*, Massachusetts: MIT press, 867-883

- Reeves, Adam (1994): *Zeitliche Auflösung der visuellen Wahrnehmung*. In: Prinz, Wolfgang [Hg.] (1994): *Enzyklopädie der Psychologie, Themenbereich C: Theorie und Forschung, Serie II: Kognition, Band 1: Wahrnehmung*. Göttingen [u.a.], Hogrefe, Verlag für Psychologie, 1994
- Roth, Gerhard (2005): *Wahrnehmung: Abbildung oder Konstruktion?* In: Schnell, Ralf [Hg.] (2005): *Wahrnehmung - Kognition - Ästhetik*. Bielefeld: Transcript, 15-33
- Schneider, Werner X. (1994): *Neuronale Netzwerke und visuelle Informationsverarbeitung*. In: Prinz, Wolfgang [Hg.] (1994): *Enzyklopädie der Psychologie, Themenbereich C: Theorie und Forschung, Serie II: Kognition, Band 1: Wahrnehmung*. Göttingen [u.a.], Hogrefe, Verlag für Psychologie, 1994
- Schnell, Ralf [Hg.] (2005): *Wahrnehmung - Kognition - Ästhetik*. Bielefeld: Transcript
- Schnell, Ralf (2000): *Medienästhetik*. Stuttgart [u.a.]: Metzler
- Shams, Ladan /Kamitani, YukiYasu. /Shimojo, Shinsuke (2004): *Modulations of Visual Perception by Sound*. In: Calvert, G.; Spence, C.; Stein, B. E. (Eds.): *The Handbook of Multisensory Processes*, Massachusetts: MIT press, 27-33
- Soto-Farako, Salvador /Kinstone, Alan (2004): *Multisensory Integration of Dynamic Information*. In: Calvert, G.; Spence, C.; Stein, B. E. (Eds.): *The Handbook of Multisensory Processes*, Massachusetts: MIT press, 49-67
- Spence, Charles /McDonald, John (2004): *The Cross-Modal Consequences of Exogenous Spatial Orienting of Attention*. In: Calvert, G.; Spence, C.; Stein, B. E. (Eds.): *The Handbook of Multisensory Processes*, Massachusetts: MIT press, 3-25
- Voland, Eckardt (2005): *Das ‚Handicap-Prinzip‘ und die biologische Evolution der ästhetischen Urteilskraft*. In: Schnell, Ralf [Hg.] (2005): *Wahrnehmung - Kognition - Ästhetik*. Bielefeld: Transcript, 35-60
- Warren, Richard M. (1982): *Auditory Perception - a new synthesis*. New York: Pergamon Press
- Webers, Johannes (1989): *Tonstudioteknik*, 5. Neu bearbeitete Auflage, München: Franzis-Verlag
- Wilke, Jürgen (2000): *Grundzüge der Medien- und Kommunikationsgeschichte*. Köln [u.a.]: Böhlau
- Woods, Timothy M. /Recanzone, Gregg H. (2004): *Cross-Modal Interactions Evidenced by the Ventriloquism Effect in Humans and Monkeys*. In: Calvert, G.; Spence, C.; Stein, B. E. (Eds.): *The Handbook of Multisensory Processes*, Massachusetts: MIT press, 35-48
- Zenner, Hans-Peter (1994): *Hören Physiologie, Biochemie, Zell- und Neurobiologie*. Stuttgart [u.a.]: Thieme

Zimbardo, Philip G./Gerrig, Richard J. (2008): *Psychologie*. 18., aktualisierte Auflage, München [u.a.]: Pearson Studium

Weitere Internetquellen:

<http://www.dienststelle.de>

<http://www.medienkunstnetz.de>

<http://www.visual-music-award.de>

## 10 ABBILDUNGSVERZEICHNIS

Abbildung 3.1:	Spektralfarben.....	22
Abbildung 3.2:	Strahlungsbewertung des menschlichen Auges.....	23
Abbildung 3.3:	Schematischer Querschnitt durch das menschliche Auge.....	24
Abbildung 3.4:	Empfindlichkeitskurven der Zapfen und Stäbchen.....	25
Abbildung 3.5:	Dichte der Stäbchen und Zapfen auf der Netzhaut.....	26
Abbildung 3.6:	Schematische Darstellung des Ohrs.....	29
Abbildung 3.7:	Isophone.....	30
Abbildung 3.8:	Schematischer medialer Schnitt durch das menschliche Gehirn.....	34
Abbildung 3.9:	Seitliche Ansicht des menschlichen Gehirns.....	35
Abbildung 3.10:	Fläche der Fovea im visuellen Feld und im visuellen Cortex...	37
Abbildung 3.11:	Hermann-Gitter.....	39
Abbildung 3.12:	Simultankontrast.....	40
Abbildung 3.13:	Größenkonstanz. ....	41
Abbildung 4.1:	Display, mit verschieden interpretierbarem visuellem Stimulus.....	48
Abbildung 4.2:	„Illusory flash effect“ .....	49
Abbildung 5.1:	Display für Versuche zum Pop-Out Effekt. ....	75
Abbildung 5.2a:	Testmatrix für perceptual grouping.....	75
Abbildung 5.2b:	Testmatrix aus Abbildung 5.1a, wie sie Synästheten sehen....	75
Abbildung 5.3:	Versuchsaufbau zu „crowding“ .....	76
Abbildung 5.4:	Nachlassen synästhetischer Farben bei größer werdender Auslenkung.....	76
Abbildung 5.5:	Testreiz zur Messung von Top-Down Verarbeitung bei Synästhesien.....	79

---

Abbildung 5.6:	Mehrdeutiger Stimulus zur Demonstration von Top-Down Verarbeitung bei Synästhesien (THE CAT).....	79
Abbildung 5.7:	Mehrdeutiger Stimulus zur Demonstration von Top-Down-Verarbeitung bei Synästhesien(,römische Vier').....	79
Abbildung 5.8:	Visuelle Stimuli ,Kiki' und ,Bouba' .....	84
Abbildung 7.1:	Höheneinteilung der tonalen Ebene. ....	107
Abbildung 7.2a:	Höheneinteilung der tonalen Ebene, 2te Oktave. ....	108
Abbildung 7.2b:	Höheneinteilung der tonalen Ebene, 3te Oktave. ....	108
Abbildung 7.3a:	Farbtöne für a' und cis' .....	111
Abbildung 7.3b:	Farbtöne für a' und c' .....	111
Abbildung 7.4a:	Farbtöne für A-Dur-Akkord; a', e' und cis' .....	113
Abbildung 7.4b:	Farbtöne für F-Dur-Akkord; a', f' und c' .....	113
Abbildung 7.5a-h:	Verschiedene visuelle Umsetzungen der Rhythm-Clicks....	114
Abbildung 7.6a-c:	visuelle Umsetzung des Rhythm-Beats. ....	116

## 11 DANKSAGUNG

Danken möchte ich

Prof. Dr. Karin Welkert-Schmitt, für eine außerordentlich intensive Auseinandersetzung mit meinem Thema und eine wirklich gute Betreuung.

Prof. Dr. Heike Sperling für Inspiration, Motivation, intensive Gespräche, richtungsweisende Vorschläge und dafür, dass sie mit als erste an meine Ideen geglaubt hat.

Besonderen Dank möchte ich an Fabian Schulz richten, für die Komposition der Musik zu dem Visual Music Clip dieser Arbeit. Die Zusammenarbeit hat mir sehr viel Spaß gemacht und ohne seine Musik wäre der Clip nicht halb so gut.

Constantin Rothkopf, für wichtige Orientierungshilfen in Sachen Kognitionsforschung.

Meinen Eltern für die tolle Unterstützung während der letzten, sehr intensiven Wochen.

Christian Handel, Kathrin Nolzen, Gitte Wolffson und Christian Egners fürs Korrekturlesen.

Mein ganz persönlicher Dank geht an Anke. Nicht zuletzt ohne sie wäre diese Arbeit so nicht möglich gewesen.

## 12 EIDESSTATTLICHE ERKLÄRUNG

Hiermit versichere ich, dass die vorliegende Diplomarbeit von mir selbst angefertigt wurde, dafür nur die angegebenen Hilfsmittel benutzt, sowie alle Stellen, die dem Wortlaut oder dem Sinn nach anderen Werken entnommen sind, durch Angabe der Quellen als Entlehnung kenntlich gemacht wurden.

Düsseldorf, den 11.3.2009

\_\_\_\_\_  
Christoph Eggener