HANS-JÜRGEN BUCHER

»Man sieht, was man hört« oder: Multimodales Verstehen als interaktionale Aneignung. Blickaufzeichnungsstudie zur Rezeption von zwei Werbespots

1. Multimodal statt audio-visuell – eine neue Betrachtungsweise für Film und Video

Es ist eine allgemein bekannte Erfahrung, dass man Filme beim zweiten Mal Anschauen anders sieht: Man entdeckt Einzelheiten, die beim ersten Mal nicht aufgefallen sind, erkennt, dass ein Aspekt für eine der folgenden Sequenzen wichtig ist, oder widmet auch der Machart, dem ›Design‹ des Films Aufmerksamkeit. Für eine Theorie der Filmrezeption sind solche Verschiebungen der Sichtweisen und der Aufmerksamkeit eine interessante und weitreichende Beobachtung, die eine ganze Reihe von theoretisch relevanten Fragen aufwirft. Eine Möglichkeit, den Unterschied zwischen dem ersten und dem zweiten Sehen zu beschreiben, ist: Wir wählen jeweils anderes aus, worauf wir unsere Aufmerksamkeit lenken, und sehen zwischen dem Ausgewählten auch verschiedene Zusammenhänge. Da es in beiden Fällen derselbe Film ist, der betrachtet wird, können die Unterschiede zwischen einem ersten und einem zweiten Sehen nicht mit Merkmalen des Films erklärt werden. Die Veränderung kann nur die Folge einer anderen Aneignung des Films sein, für die in erster Linie der Rezipient verantwortlich ist.

Dieser Aneignungsprozess steht im folgenden Beitrag im Mittelpunkt: Es werden Befunde einer Rezeptionsstudie präsentiert, die zeigen sollen,

wie sich Rezipienten ihr Verständnis eines Videofilms aufbauen. Gegenstand der empirischen Studie sind zwei Werbespots der Firma LG, die mit verschiedenen Plots – im Hotel und in der U-Bahn – für ein TV-Handy werben. Die Filme werden dabei als multimodale Kommunikationsformen betrachtet, die sich nicht nur aus Text und Bild zusammensetzen, sondern auch aus Geräuschen, szenischer Sprache, Sprache aus dem Off, Bild- und Textdesign, Animation, Farbgestaltung und Dramaturgie. Die Studie selbst ist so angelegt, dass in verschiedenen Szenarien jeweils unterschiedliche Rezeptionsfaktoren wie Intentionen oder Vorwissen der Rezipienten, die Relevanz einzelner Modi oder der Aufbau des Films überprüfbar werden. Grundgedanke dieser Vorgehensweise ist es, die Medienlogik von Film und Video über deren Rezeption zu erschließen. Dieser methodische Weg basiert auf dem sogenannten ›dialogischen Prinzip‹, wie es bereits in der analytischen Sprachphilosophie formuliert wurde: Eigenschaften, Strukturen, Bedeutungen von Kommunikationsbeiträgen zeigen sich erst in den Anschlusshandlungen der Adressaten, also in der Art ihrer Einbettung in ein Sprachspiel (exemplarisch und stellvertretend: AUSTIN 1986).

Auch in der Systemtheorie ist dieser Gedanke verortet: Die Kontingenz der Kommunikation löst sich erst in ihrer Verarbeitung – also im ›Weitermachen‹ eines Adressaten – auf, weshalb der Sinn eines Kommunikationsbeitrags erst auf der Ebene des dritten Zuges etabliert wird. »Die Mitteilung selbst ist zunächst nur eine Selektionsofferte. Erst die Reaktion schließt die Kommunikation ab, und erst an ihr kann man ablesen, was als Einheit zustande gekommen ist« (LUHMANN 1984: 212; weiterführend SCHNEIDER 1996).

Die Betrachtung von Filmen – audiovisuellen Medien – als multimodale Kommunikationsformen bedeutet gegenüber der herkömmlichen Filmanalyse eine erhebliche Ausweitung des Gegenstandsverständnisses, aber auch eine Erweiterung der theoretischen Perspektiven. »Film, Fernsehen, Video stellen sich dem Betrachter als eine Abfolge von Bildern dar«, heißt es in der Einführung zur *Film- und Fernsehanalyse* von Knut Hickethier (2007: 52). Diese begrenzte Sichtweise blendet aus, dass Filme neben der Abfolge von Bildern auch eine Abfolge von gesprochener Sprache im On und Off, Geräuschen, Musik und eventuell auch eingeblendeten Texten darstellen. Zwar ergänzt Hickethier seine Definition des Films in einem späteren Kapitel um das Auditive – Sound, Musik, Sprache –, in der Analyse aber bleibt das Visuelle gegenüber dem Auditiven vorrangig. Es finden sich zwar auch vereinzelte Hinweise darauf, dass Bild und Sprache den Gesamtsinn des »audio-visuellen Textes« (HICKETHIER 2007: 22f.) kom-

positorisch konstituieren (ebd.: 23, 93, 102).Wie das geschieht, wie »Hörraum und Bildraum« (ebd.: 90) zusammenhängen, wird aber ebenso wenig geklärt wie die spezifischen semiotischen Leistungen der einzelnen Modi.

Eine differenziertere Sicht auf Filme, in der diese nicht auf eine ›Doppelstruktur‹ aus Bild und Ton reduziert werden, liefert die angewandte Medienästhetik von Herbert Zettl (1999), die sich sowohl als Analysemethode als auch als Produktionsanleitung versteht (ebd.: 13: »analysis and synthesis«). Im Sinne eines formalistischen Ansatzes isoliert Zettl fünf Elemente, aus denen Filme konstruiert sind: 1. Licht und Farbe, 2. der zweidimensionale Raum, 3. der dreidimensionale Raum, 4. Zeit und Bewegung und 5. Sound. In den sehr ausführlichen Darstellungen zu diesen Analyse- und Gestaltungsbereichen wird deren funktionaler Beitrag zum Gesamtsinn eines Films ausführlich beschrieben und Kriterien für deren Abstimmung zu einem ästhetischen Ganzen (»aesthetic whole«) vorgeschlagen (vgl. z. B. ebd.: 351-357 für Sound). Allerdings werden diese Beschreibungen aus der Perspektive des Filmemachers formuliert, ohne dass dabei zeichentheoretische Grundlagen geklärt werden.

Auch eine dritte Forschungsrichtung, die kommunikationswissenschaftliche Medienwirkungs- und Mediennutzungsforschung, trägt zu den Fragen der Bedeutungskonstitution von Film und Fernsehangeboten eher wenig bei. Hier ist der Trend vorherrschend, Medienangebote ganz allgemein als Stimuli oder als Inhalte und Informationen zu behandeln und darin das ›tertium comparationis‹, das allen Medienkontakten Gemeinsame, zu sehen (BONFADELLI 2004; JÄCKEL 2005; SCHWEIGER 2007). Dass der Rezeptionsprozess neben den situativen, sozialen und individuellen Faktoren entscheidend davon abhängen kann, in welcher Mediengattung, mit welchen Modi, in welcher Aufmachungs- und Darstellungsform die Inhalte und Informationen präsentiert werden, wird nur in Ausnahmefällen in Betracht gezogen (BROSIUS 1998; BILANDZIC 2004; JENSEN 2002).

Betrachtet man Filme als multimodale Kommunikationsform, so rückt nicht nur die ganze Bandbreite bedeutungstragender Symbole ins Blickfeld, sondern es eröffnet sich auch die Möglichkeit, die klassische Unterscheidung von linearen und nicht linearen Medien zu präzisieren. Filme und Videos werden gemeinhin als lineare Medien bezeichnet, womit zum einen die sequenzielle Aufbaustruktur und zum anderen die schrittweise Rezeption im Zeitverlauf gemeint ist. Unter einer multimodalen Perspektive ist diese Kategorisierung allerdings zu eng: Neben ihrer linearen Struktur wei-

sen Filme und Videos auch eine non-lineare Struktur auf (BUCHER 2010a: Kap. 2,3). Die Kohärenz eines Filmes wird eben nicht nur durch seine zeitliche Sequenzstruktur – seinen Rhythmus (VAN LEEUWEN 2005: 181) – bestimmt, sondern auch durch eine räumliche Anordnung des Gezeigten in der jeweiligen Einstellung oder Szene. »In films the composition of the shots and the arrangements of the set and locations are spatially organized, while the action, the dialogue, the music and the other sounds are organized according to the rhythmic principles« (VAN LEEUWEN 2005: 181; vgl. auch: BALDRY/THIBAULT 2005).

Wie das räumliche und das zeitliche Prinzip beim Verständnis von Bildsequenzen ineinandergreifen, hat Lim am Beispiel von Comics demonstriert: Als Basiseinheiten, auf denen wir ein Bildverständnis aufbauen, bestimmt er sogenannte ›Associated Elements‹ (AE), die in Teilen oder als Teile mit einer größeren Einheit assoziiert sind und die dem Betrachter dabei helfen, Zusammenhänge zwischen aufeinanderfolgenden Abbildungen zu sehen (LIM 2007: 202). Aus diesen AEs lassen sich dann ›Visual Linking Devices‹ (VLDs) ableiten, sogenannte ›Referenz-Ketten‹, wie wir sie auch in Texten finden. Um Kohärenz in einem Film zu erkennen, ist also erstens auszuwählen, welche Teile eines Bildes/einer Filmeinstellung relevant sind – welche also die verbindenden ›Associated Elements‹ sind –, und zweitens, welche Aspekte relevant sind, um die Kohärenz innerhalb einer Film- oder Bildersequenz zu erkennen – die ›Visual Linking Devices‹. Bevor wir in einer Film- oder Bildsequenz Zusammenhänge erkennen, müssen wir dementsprechend erst auswählen, was eigentlich zusammenhängt. Vor der Kohärenz steht die Selektion. Oder: Bevor wir Linearität in einer Filmsequenz oder einer Bildfolge sehen, müssen wir die Non-Linearität der kleineren Einheiten aufgelöst haben.

Da jede Selektionsentscheidung von den vorausgegangenen abhängt, kann man diesen iterativen Prozess mit einem *Modell der Interaktivität* erklären. Wie in einer Interaktion zwischen anwesenden Partnern jeder Kommunikationsbeitrag den Interaktionsstand hinsichtlich des gemeinsamen Wissens, der eingegangenen Festlegungen und der thematischen Progression verändert, so verschiebt auch die fortlaufende Rezeption eines monologischen Kommunikationsangebotes das Wissen, die Erwartungen und die Relevanzkriterien der Rezipienten. Jede Selektion eines relevanten Aspektes aus dem Kommunikationsangebot dient immer auch der Verifizierung, Falsifizierung oder Modifikation des bereits erreichten Verständnisses, erlaubt also dem Rezipienten andere Fragen an das Kom-

munikat zu richten und andere Antworten zu erkennen. Angesichts des Fehlens eines anwesenden Partners kann man auch von einer *kontrafaktischen Interaktion* sprechen (BUCHER 2010b; BUCHER/SCHUMACHER 2011). Diese interaktionale Auffassung der Rezeption liegt auch dem Design der hier vorzustellenden Studie zugrunde.

Wie für andere multimodale Kommunikationsformen in Printmedien, Internet, mobilen Endgeräten oder in der direkten Kommunikation gilt auch für Film und Video, dass Multimodalität eine dreidimensionale Eigenschaft ist: diese Kommunikationsformen sind *multimedial*, indem sie verschiedene Mediengattungen wie Print, Hörfunk, Fernsehen verbinden, sie sind *multikodal,* indem sie gleichzeitig verschiedene semiotische Kodes wie Text, Sprache, Sound, Design, Layout, Farbe, Grafik, Bild bedienen, und sie sind als dritte Eigenschaft *non-linear*, insofern das Arrangement der verschiedenen Kommunikationselemente dem Rezipienten eine Selektionsleistung abverlangt (ausführlicher in BUCHER 2010a).

Die eingangs aufgeführten Unterschiede zwischen einer Erst- und einer Folgebetrachtung lassen sich auf der Basis des explizierten Multimodalitätsbegriffs auf unterschiedliche Selektionsleistungen und unterschiedliche sequenzielle Deutungen bei der Filmrezeption zurückführen. Um diese multimodale Auffassung von Film und Video auch empirisch zu belegen, muss erstens gezeigt werden, dass Selektionsprozesse von Elementen eines Video- oder Filmangebotes stattfinden, und zweitens, dass auf der Basis der als relevant ausgewählten Elemente entsprechende Sinnzusammenhänge zwischen diesen Elementen hergestellt werden. Die theoretische Herausforderung für eine Analyse von Film und Video besteht in der Frage, ob es Regeln oder Muster für solche Selektions- und Deutungsleistungen gibt und ob Rezeption und Angebot in systematischer Weise miteinander zusammenhängen. Die empirische Rezeptionsstudie zu den beiden Werbespots soll diese beiden Fragestellungen klären.

2. Empirische Multimodalitätsforschung: zum Zusammenhang von Theorie und Methode

Der Begriff der Multimodalität kommt in zwei Verwendungsweisen vor. Er wird erstens als *empirischer Begriff* verwendet, um Veränderungen der Medienkommunikation zu beschreiben. In diesem Sinne bildet der Begriff eine kommunikative Praxis ab, die darin besteht, unterschiedliche semiotische

Ressourcen zu kombinieren. Kommunikationsgeschichtlich betrachtet ist die Multimodalisierung ein Metaprozess der Mediengeschichte (BUCHER 2010a), der durch neue Produktionstechniken und vor allem durch die Technik der Digitalisierung eine Radikalisierung erfahren hat (KRESS/VAN LEEUWEN 1996; KRESS 2002; IEDEMA 2003; BATEMAN 2008: insb. 1–9).

Die zweite Verwendungsweise des Begriffs ›Multimodalität‹ ist *kategorialer Art*: Multimodalität ist keine historisch entstandene Erscheinungsform oder Ausprägung der Kommunikation, sondern eine *konstitutive Eigenschaft aller Formen der Kommunikation*, »an inherent feature of all aspects of our lives« (MATTHIESSEN 2007: 1). Aufgrund der »essentially multimodal nature of all human meaning making« (IEDEMA 2003: 39), derzufolge in allen Formen der Kommunikation neben den sprachlichen auch andere semiotische Ressourcen zur Sinnerzeugung eingesetzt werden, heißt es bei Kress und van Leeuwen: »all texts are multimodal« (1998: 186; ebenso: BALDRY/THIBAULT 2005: 19). Diese kategoriale Verwendungsweise des Begriffs ›Multimodalität‹ impliziert einen *Wechsel in der Betrachtungsweise* auf alle Formen der Kommunikation. Das bedeutet, dass jede Kommunikationsanalyse multimodal ausgerichtet sein muss und zeigen sollte, wie sich Sinn und Bedeutung eines Kommunikationsbeitrags aus den unterschiedlichen Modi ergeben. Mit dem Begriff der Multimodalität ist eine analytische Perspektive auf alle Formen der Kommunikation verbunden, die erstmals die umfassende Erschließung aller Sinn- und Bedeutungspotenziale erfassbar macht. Wenn Multimodalität ein konstitutiver Aspekt aller Kommunikation ist, so sind die klassischen Fragen der Kommunikations- und Medienanalyse neu zu stellen. Für eine Klärung dieser Fragen lassen sich zwei zentrale Problemfelder unterscheiden:

Das *Problem der Kompositionalität*, das zwar in der Linguistik in Bezug auf monologische Texte und dialogische Äußerungen bereits behandelt wird, im Hinblick auf multimodale Kommunikationsformen aber auf alle beteiligten Kommunikationsmodi erweitert werden muss. Angesichts der Ko-Okkurenz verschiedener semiotischer Ressourcen und intermodaler sowie »intersemiotischer Relationen« (IEDEMA 2003; LIM 2004) ergibt sich die Frage, welchen Beitrag einzelne Elemente aus verschiedenen Modi zum Gesamtsinn eines Kommunikationsbeitrags leisten und wie diese Leistungen integriert sind. Dabei besteht Übereinstimmung, dass das Ganze – das Kommunikationsangebot – mehr ist als die Summe seiner Teile und dementsprechend der Gesamtsinn nicht additiv, sondern in einem noch zu klärenden Sinne ›multiplikatorisch‹ als intersemiotischer

Prozess zu erklären ist (LEMKE 1998; O'HALLORAN 2008; LIM 2004, 2007). Die Begriffe der ›Intersemiosis‹, der ›semantischen Multiplikation‹ und die Annahme eines ›space of integration‹, in dem der semantische Mehrwert in der Interaktion der verschiedenen Modi generiert werden soll (LIM 2004), sind theoretische Konstrukte, um den kommunikativen Gesamtsinn eines multimodalen Angebotes zu erklären. Die Bearbeitung des Problems der Kompositionalität lässt sich in drei grundlegende Fragen zerlegen:

1. Welche spezifischen kommunikativen Leistungen werden von den einzelnen Modi wie Text, Sprache, Bild, Musik, Ton, Design etc. erbracht?
2. Wie ist das funktionale Zusammenspiel der einzelnen Modi zu einer Gesamtbedeutung – die intersemiotischen Relationen – zu erklären?
3. Welche Rolle spielen bei der Sinnerzeugung holistische und lokale Aspekte des Kommunikationsangebotes?

Das zweite Problemfeld, das *Problem der Rezeption*, ist gewissermaßen das Spiegelbild des ersten. Es ist eine Besonderheit des hier vertretenen Ansatzes, dass diese beiden Probleme im Zusammenhang gesehen werden, was in der bisherigen Multimodalitätsforschung nicht der Fall war. Die linguistischen Ansätze zur Multimodalität haben sich auf die produktanalytischen Aufgaben beschränkt, die psychologischen, lerntheoretischen oder medienwissenschaftlichen Ansätze auf die rezeptionsanalytischen Aspekte der Multimodalität. Die Frage, wie Rezipienten die non-linearen und fragmentierten multimodalen Kommunikationsangebote zu einem kohärenten Verständnis integrieren, kann nicht unabhängig von einer Theorie des entsprechenden Gegenstandsbereichs geklärt werden. Allein die Entscheidung, welches die verstehensrelevanten, bedeutungstragenden Bausteine eines multimodalen Angebotes sind, setzt eine solche Theorie voraus. Auf der anderen Seite ist auch eine Theorie multimodaler Kommunikationsformen auf eine Rezeptionstheorie angewiesen.

Die hier vorgestellte empirische Rezeptionsstudie zu den beiden Filmen ist so angelegt, dass beide der genannten Problembereiche behandelt werden: In den Befunden zur Rezeption der beiden Videos wird sich zeigen, wie Probanden den multimodalen Gesamtsinn dieser Angebote zusammensetzen. Die Rezeptionsdaten können insofern herangezogen werden, um eine Theorie des multimodalen Verstehens zu begründen.

Methodisch basiert die Studie auf einem Mehrmethodenansatz und vermeidet damit den Paradigmen-Konflikt zwischen quantitativen Verfahren einerseits und qualitativ-phänomenologischen Verfahren ande-

rerseits (TASHAKKORI/TEDDLIE 1998). Folgende Verfahren kommen in der Studie zum Einsatz:

1. Ein *Blickaufzeichnungsverfahren*, mit dem die Aufmerksamkeitsverteilung beim Betrachten der Videos erfasst wird. Blickdaten gelten als verlässliche Indikatoren für die kognitive Verarbeitung von visuellen Stimuli, da sie simultan zum Rezeptionsprozess erfasst werden und zum größten Teil nicht willkürlicher Art sind (HYÖNÄ et al. 2003; BENTE 2004; RICHARDSON/SPIVEY 2004; DUCHOWSKI 2007: Part I-III). Gegenüber den Post-hoc-Verfahren der Rezeptionsforschung wie Befragungen, Selbstauskünften oder Wissenstests, die durch Erinnerungsverzerrungen und individuelle Motive der Probanden beeinflusst sein können, sind Blickdaten weitestgehend authentisch. Während die Anwendungsfelder der Blickaufzeichnung relativ weitgestreut sind – von der Mensch-Computer-Kommunikation über die Zeitungsnutzung bis hin zur Cockpit-Ergonomie (HYÖNÄ et al. 2003; HENDERSON/FERREIRA 2004; DUCHOWSKI 2007: Part IV; GOMPEL et al. 2007) –, beschränken sich die methodischen Szenarien auf die Wahrnehmung stehender Bilder und Printmedien, die Orientierung in realen Szenen und auf das Lesen von Texten (HENDERSON/FERREIRA 2004). Blickstudien zu Film und Video liegen nahezu keine vor (DUCHOWSKI 2007: Kap. 19).

Blickdaten können Aufschluss über ganz verschiedene Aspekte der Aufmerksamkeitsverteilung geben:

- über die fixierten Regionen des Blickfeldes und damit das Selektionsmuster (Was wird rezipiert?)
- über den Grad der Aufmerksamkeit und des Interesses für bestimmte Regionen (Wie lange und wie häufig wird etwas wahrgenommen?)
- über die Rezeptionsabfolge oder Scan-Pfade (In welcher Abfolge werden die Elemente wahrgenommen?)
- über die Qualität der Rezeption (Wann wird gescannt und wann gelesen/angeschaut?)

Für die Auswertung der Blickdaten werden in den verschiedenen Einstellungen der beiden Videos sogenannte ›Areas of Interest‹ festgelegt (siehe Abb. 1). Grundlage für diese Festlegungen sind die realen Blickverläufe der Probanden, wie sie die Blickkamera dokumentiert.

2. Bei der *Methode des Lauten Denkens* sind die Probanden aufgefordert, das auszusprechen, was ihnen während der Rezeption ›durch den Kopf geht‹ (vgl. BILANDZIC 2005). Entgegen der Bezeichnung der Methode werden dabei nicht Denkprozesse verbalisiert, die die Rezeption steuern. Vielmehr handelt es sich um eine Form der Spontankommentierung des-

ABBILDUNG 1

Areas of Interest (Aufmerksamkeitsregionen) für den Hotel-Spot

sen, was die Probanden sehen, verstehen oder auch nicht verstehen. Diese Spontankommentierungen liefern in vielen Fällen erst den Kontext zur Interpretation der Blickdaten. Während die Blickdaten anzeigen, was die Probanden anschauen, geben die Äußerungsdaten darüber Auskunft, was sie sehen (HOLSANOVA 2008: Kap. 5; vgl. auch Kap. 5.1 des Beitrags).

3. Mit den *Nacherzählungen* wird ein weiterer Typus von Äußerungsdaten erhoben, der dazu dient, den Grad zu ermitteln, in dem die gezeigten Videos verstanden wurden. Im Unterschied zum Lauten Denken handelt es sich um Post-Hoc-Verbaldaten.

4. Mit einem *Behaltenstest* wird festgestellt, in welchem Ausmaß die Werbebotschaft für das beworbene Handy die Probanden erreicht hat.

Dieser Mehrmethoden-Ansatz sorgt dafür, dass die Rezeptionsdaten möglichst breit gefächert und wechselseitig erhellend sind. Die so erhobenen Daten erlauben auch verschiedene Kombinationen von qualitativ-interpretierenden und quantitativ-zählenden Auswertungsschritten (vgl. TASHAKKORI/TEDDLIE 1998: 44). So lassen sich beispielsweise die Fixationsdauer für eine Area of Interest oder die Häufigkeit von Blickwechseln zwischen verschiedenen AOIs mit den interpretierenden Äußerungen zu diesen AOIs abgleichen.

Eine weitere Maßnahme zur Erhöhung der empirischen Validität und der Datenqualität ist die Aufspaltung der Vorgehensweise in verschiedene *Rezeptionsszenarien*: Durch das Arrangieren unterschiedlicher Nutzungssituationen im Labor wird es möglich, einzelne Rezeptionsfaktoren zu isolieren. Folgende Rezeptionsszenarien wurden differenziert:

- *Szenario 1*: Aufgabenstellung Handy-Kauf: Die Probanden betrachten die Spots mit der Vorgabe, sie als Entscheidungshilfe für einen Handykauf heranzuziehen.
- *Szenario 2*: Aufgabenstellung Spot-Analyse: Die Probanden betrachten die Spots mit der Aufgabe, ihre mediale Machart herauszufinden.
- *Szenario 3*: Keine Aufgabenstellung: Dieses neutrale Szenario dient der Erhebung von Kontrolldaten.
- *Szenario 4*: Keine Aufgabe, zweimaliges Anschauen der Spots: mit diesem Szenario soll der Einfluss des Vorwissens und der Vertrautheit mit dem Stimulus gemessen werden. Da für alle Probanden abgefragt wurde, ob sie die Spots kannten oder nicht, kann auch der Einfluss des erinnerten Wissens ermittelt werden.
- *Szenario 5*: Keine Aufgabe, Spots werden ohne Ton gezeigt: Durch dieses Szenario kann der Einfluss der Modi Sound, szenische Sprache und Sprache aus dem Off auf die Blickdaten und auf das Verständnis der Spots überprüft werden.

Tabelle 1 zeigt die Szenarien sowie die jeweilige Anzahl der Probanden im Überblick. Insgesamt wurden 46 Probanden getestet, von sechs Probanden waren die Blickdaten aus technischen Gründen nicht verwertbar.

Aus einer vorausgegangenen Studie, bei der die Tonaufzeichnung aus technischen Gründen ausfiel, konnten für Szenario 5 ›ohne Ton‹ die Blickdaten von acht weiteren Probanden für den Vergleich mit den Blickdaten aus den Szenarien 1 - 4 ›mit Ton‹ berücksichtigt werden.

Die Daten aus den verschiedenen Szenarien ermöglichen mehrere Vergleichsauswertungen. So kann durch einen Vergleich der Daten aus den

TABELLE 1
Rezeptionsszenarien
(N = 46, techn. Ausfall Blickdaten 6)

Testszenarien	Blickdaten	Äußerungsdaten
Szenario 1: Aufgabenstellung Handy-Kauf	6 Probanden	8 Probanden
Szenario 2: Aufgabenstellung Spot-Analyse	8 Probanden	8 Probanden
Szenario 3: Keine Aufgabenstellung	6 Probanden	7 Probanden
Szenario 4: Keine Aufgabe, zweimaliges Sehen	6 Probanden	7 Probanden
Szenario 5: Keine Aufgabe, Spots ohne Ton	14 Probanden	16 Probanden (Lautes Denken)

Szenarien 1 bis 3 der Einfluss der Aufgabenstellung und damit der Intention der Betrachter auf die Rezeption ermittelt werden. Ein Vergleich der Daten zwischen dem ersten und dem zweiten Betrachten der Videos erlaubt Rückschlüsse auf die Auswirkungen des Vorwissens, was auch durch einen Vergleich zwischen Probanden mit und ohne Kenntnis der Spots ermöglicht wird. Für die Ermittlung des Zusammenspiels verschiedener Modi werden die Daten von Probanden verglichen, die den Spot *mit* bzw. *ohne* Ton gesehen haben. Bei diesen Vergleichen wurden aus Szenario 4 jeweils nur die Daten aus dem ersten Anschauen des entsprechenden Spots berücksichtigt.

Das Design dieser Studie beruht auf einer interaktionalen Theorie der Medienrezeption: Rezeption wird nicht als Wirkungsprozess verstanden, sondern als inter-aktive Aneignung durch die Rezipienten (BUCHER 2008, 2010b; BUCHER/SCHUMACHER 2011). Diese Betrachtungsweise stützt sich nicht nur auf neuere Entwicklungen in der Rezeptionstheorie (THOMPSON 1995; BONFADELLI 2004), sondern schließt auch an interaktionale Auffassungen des Textverstehens (ISER 1980; BAKHTIN 1981, 1986), der Mensch-Computer-Kommunikation (MAYER 1998; KIOUSIS 2002; MCMILLAN 2002; BUCHER 2004) und der psycholinguistischen Diskursforschung (HOLSANOVA 2008) an. Die Blickaufzeichnung und die erhobenen Äußerungsdaten des Lauten Denkens sollen den interaktiven Prozess der Aneignung für die

beiden Videos rekonstruierbar machen. Die Präsentation der Befunde zur Rezeption der beiden Videos orientiert sich dementsprechend an den folgenden Fragen, die an die oben unterschiedenen Grundprobleme der Multimodalitätsforschung – das Kompositionalitätsproblem und das Rezeptionsproblem – anschließen:

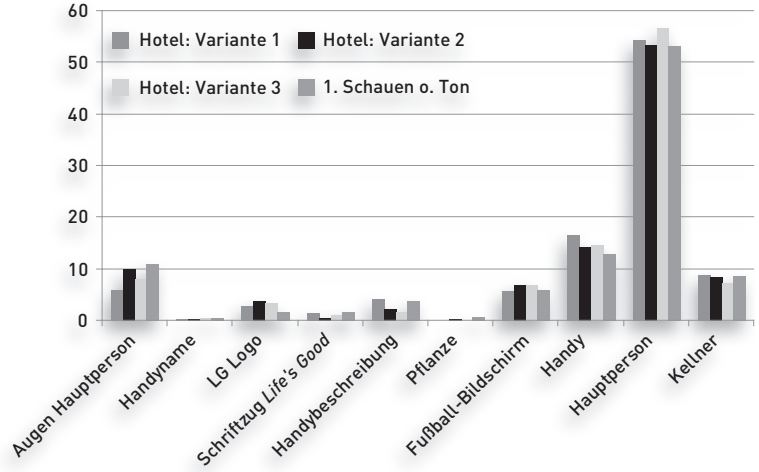
- Welchen Einfluss haben Merkmale der Rezipienten – ihre Motive, Intentionen, ihr Vorwissen – auf die Rezeption der Videos? (Abschnitt 3)
- Welchen Einfluss haben Merkmale der beiden Videos auf die Rezeption? (Abschnitt 4)
- Wie hängt die Bedeutung der Beobachtungsobjekte – ihre semiotische Dimension – mit deren Wahrnehmung zusammen? (Abschnitt 5.1)
- In welcher Weise beeinflusst die Dynamik der Video-Erzählung die Dynamik der Aneignung? (Abschnitt 5.2)
- Wie wird das Zusammenspiel der verschiedenen Modi von den Rezipienten erfasst und welche Funktionen kommen dabei den einzelnen Modi zu? (Abschnitt 6.1)
- Welche Muster sind im Prozess der Aneignung der beiden Videos erkennbar? (Abschnitt 6.2)

3. Die Top-down-Logik des multimodalen Verstehens: Zur Relevanz von Intention und Vorwissen

In seinem klassischen Experiment hat der russische Kognitionspsychologe Alfred L. Yarbus am Beispiel des Bildes *An Unexpected Visitor* von Ilya Repin gezeigt, dass die Blickbewegungen der Betrachter von den jeweiligen Beobachtungsaufgaben – und damit den entsprechenden Beobachtungsintentionen – abhängen (YARBUS 1967: 174, 192f.). Aufgrund seiner Blickbefunde kam Yarbus zu dem Schluss: »All the records [of eye movements, HJB] show conclusively that the character of the eye movements is either completely independent of or only very slightly dependent on the material of the picture and how it was made« (YARBUS 1967: 190). Offensichtlich bedingen die Beobachtungsintentionen ein Rezeptionsmuster, das funktional auf diese abgestimmt ist und die Aufmerksamkeit auf diejenigen Elemente eines Bildes richtet, die »wesentliche und nützliche Information« (»essential and useful information«, ebd.: 175, 182) für den Betrachter enthalten, d. h. ihm die Lösung der Beobachtungsaufgabe ermöglichen (dazu weiterführend: ARNHEIM 2001; HOLSANOVA 2008; HENDERSON/BROCKMOLE et al. 2007).

Um die Einflüsse von Intentionen und Wissen sowie deren Dynamik auf die Rezeption multimodaler Texte überprüfen zu können – die sogenannte »schema hypothesis« (DUCHOWSKI 2007: 219) –, wurden in der hier vorzustellenden Studie – ähnlich wie bei Yarbus – die verschiedenen *Laborszenarien* arrangiert (vgl. Abschnitt 2). Wie Abb. 2 zeigt, führen die Unterschiede in der Aufgabenstellung auch zu Unterschieden bei den Blickdaten.

ABBILDUNG 2
Verteilung der Prozentanteile der Fixationszeit auf die verschiedenen AOIs in vier verschiedenen Szenarien



Variante 1: Kaufentscheidung für ein Handy
Variante 2: Analytische Betrachtung des Spots
Variante 3: Keine Vorgabe
Variante 4: Ohne Ton

Im Falle der Aufgabe, den Spot jeweils unter medienanalytischen Gesichtspunkten zu betrachten, erhalten gegenüber den anderen beiden Szenarien solche Elemente höhere Aufmerksamkeit, die ausschließlich für den Handlungsstrang der Geschichte relevant zu sein scheinen. Im U-Bahn-Spot sind das die Augen des Protagonisten und eine rechts von ihm sitzende Frau dunkler Hautfarbe, im Hotel-Spot ist es neben den Augen

des Protagonisten die Zimmerpflanze, deren Tongranulat der Protagonist isst (ohne es zu bemerken).

Bei der Aufgabe, den Handy-Spot als eine Hilfe für die eigene Kaufentscheidung anzuschauen, wird das Handy selbst während des ganzen Spots bedeutend länger fixiert als in den übrigen beiden Szenarien. In der abschließenden Produktpräsentation des Spots erhalten die Handybeschreibung und das Handy-Logo eine signifikant höhere Aufmerksamkeit. Insgesamt hat die Aufgabenstellung zur Folge, dass verstärkt diejenigen Aspekte des Videos betrachtet werden, die für die Lösung der jeweiligen Beobachtungsaufgabe relevant sind, also beispielsweise die Produktmerkmale für eine Kaufentscheidung.

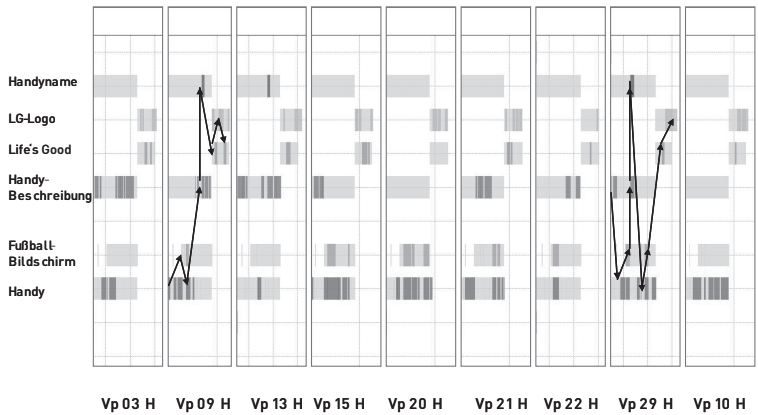
Während sich die Blickdaten der 20 Probanden, die keinen der Spots kannten, von den Blickdaten der zehn Probanden, die beide (7) oder zumindest einen der beiden Spots (3) kannten, nur geringfügig unterscheiden, zeigt ein Vergleich der Blickdaten des ersten und zweiten Anschauens für beide Spots erhebliche Unterschiede – aber auch bemerkenswerte Übereinstimmungen. Beim ersten Betrachten ist die Fixationszeit für das Handy und die auf seinem Bildschirm gezeigte Fußballsequenz 3 (Hotel) bzw. 1,5 Mal (U-Bahn) länger. Beim zweiten Anschauen verschiebt sich die Aufmerksamkeit auf die Produktbeschreibungen. So wird beim ersten Betrachten in der abschließenden Produktpräsentation hauptsächlich das Handy-Gerät fixiert, gewissermaßen in Fortsetzung der erzählten Geschichte, während die Probanden beim zweiten Anschauen im Hotelspot viermal und im U-Bahn-Spot dreimal länger den Text der Handy-Beschreibung lesen. Die Rezeption der Spots ist beim ersten Anschauen insgesamt deutlich stärker storyorientiert, beim zweiten Anschauen dagegen produktorientiert. Einer der Probanden beschreibt diese Aufmerksamkeitsverschiebung von der Story zum beworbenen Produkt in einer Selbstkommentierung folgendermaßen:

> *»Ich hab beim zweiten Durchlauf, als dann das Produkt eingeblendet wurde, nicht mehr so stark aufs Produkt an sich geguckt, sondern unten auch versucht, den Produktnamen mir irgendwie einzuprägen [lacht], also zumindest wahrzunehmen. Also am Anfang habe ich nur das Handy an sich wahrgenommen und mir dieses Produkt angeguckt und danach wollt ich auch wissen, also was ist das für'n Typus von Handy.«*

Mit den Zeitangaben im letzten Satz (*am Anfang*, *und danach*) wird explizit auf den Wissenszuwachs durch das erste Anschauen Bezug genommen, der beim zweiten Betrachten dann auch andere Relevanzkriterien zur Folge hat: Wenn die Art des Produktes bekannt ist, wird in einem zweiten Schritt der Typ dieses Produktes ermittelt. Werden die Spots also direkt

hintereinander ein zweites Mal angeschaut, führt der Lernprozess zu einer Verschiebung der Aufmerksamkeit vom beworbenen Gegenstand zur Gegenstandsbeschreibung. Für den abschließenden Teil der Spots, in dem das Produkt präsentiert wird, hat das zur Folge, dass der Fixationsverlauf deutlich de-linearisiert ist (vgl. Abb. 3). Die Probanden nutzen offensichtlich wechselweise die verschiedenen als AOIs markierten Elemente, um sich so kumulativ ein Verständnis des Produktteils des Spots aufzubauen. Der Prozess einer interaktiven Aneignung wird hier in der Zick-Zack-Linie des Aufmerksamkeitsverlaufs deutlich sichtbar.

ABBILDUNG 3
Sequenz-Charts verschiedener Probanden für den Produktteil des Spots



Erklärungsbedürftig sind allerdings auch die Übereinstimmungen in den Blickdaten, sowohl zwischen Probanden mit und ohne Vorkenntnissen der Spots als auch zwischen dem ersten und dem zweiten Anschauen: Offensichtlich ist die Logik der Spots so zwingend, dass sie vor allem im Story-Teil die Selektionsmöglichkeiten für die Rezipienten in hohem Maße kontrolliert. Insgesamt ist festzuhalten, dass Vorwissen und Intentionen den Blickverlauf weniger deutlich beeinflussen, als es die Befunde von Yarbus erwarten lassen. Vor allem im narrativen Teil der beiden Spots zeigt sich ein hohes Maß an Übereinstimmung zwischen den Blickmustern der drei Szenarien. Man muss daraus den Schluss ziehen,

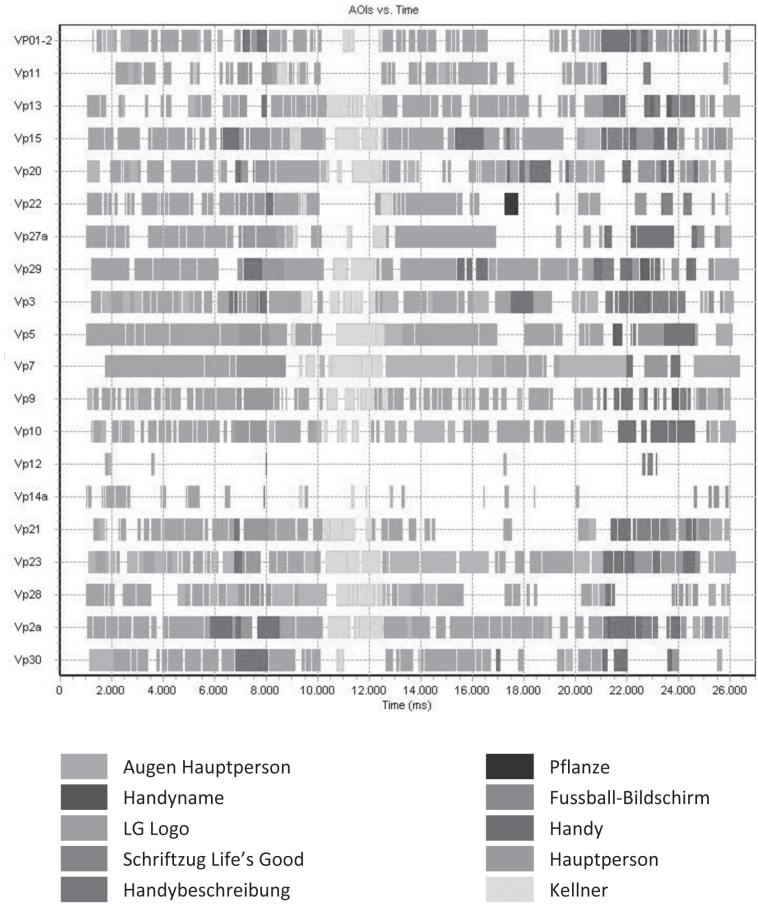
dass im Falle der beiden Filme auch Faktoren des Stimulus selbst die Aufmerksamkeitsverteilung beeinflussen. Der visuelle Reiz der Videos – ihre Salienz – ist neben ihrer Semantik – der Bedeutung der verschiedenen AOIs – ein weiterer Faktor, der für eine Erklärung des Rezeptionsprozesses herangezogen werden muss.

4. Bottom-up-Logik des Verstehens: Die Macht des Visuellen

Das Gegenstück zur ›idealistischen‹ Auffassung der Rezeption, wie sie der ›cognitive control hypothesis‹ zugrunde liegt, ist die von der ›visual salience hypothesis‹ vertretene ›realistische‹ Konzeption der Wahrnehmung. Ihre Vertreter gehen davon aus, dass es die Stimulusfaktoren sind, die die Aufmerksamkeit bestimmen, ohne dass deren Bedeutung dabei erkannt werden muss. Nach dem von Itti, Koch und Niebur vorgeschlagenen Modell einer salienzbasierten Aufmerksamkeit (ITTI et al. 1998) erstellt ein Rezipient im Prozess der Wahrnehmung ausschließlich mithilfe von Stimulusfaktoren wie Farbe, Intensität, Kontrast und Anordnung durch entsprechende Selektion eine Salienz-Landkarte des betrachteten Objektes. Was an einer gezeigten Szene oder einer gezeigten Abbildung informativ und relevant ist, wird diesem Modell zufolge ausschließlich ›bottom-up‹ durch die optische Beschaffenheit des visuellen Stimulus bestimmt. Allerdings lassen sich die nach dem Modell prognostisch errechneten Informationswerte von Abbildungen in vielen Fällen nicht mit den Blickdaten realer Probanden zur Deckung bringen (vgl. HENDERSON/FERREIRA 2004: 22 - 26; HENDERSON/BROCKMOLE et al. 2007: insb. 551, 557). Auch die in Abschnitt 3 vorgestellten Belege für eine Top-down- und schemabasierte Erklärung der Wahrnehmung sprechen gegen eine uneingeschränkte Gültigkeit des Salienzmodells (vgl. auch HENDERSON/BROCKMOLE et al. 2007) und gegen die ihm zugrunde liegende Annahme einer Wahrnehmung, unabhängig vom Bedeutungsgehalt eines fixierten Objektes. Dennoch finden sich in den Blickdaten zu den beiden LG-Werbespots Hinweise, die eine Bottom-up-Erklärung der Wahrnehmung für die beiden genannten Szenarien nahelegen. So zeigt Abb. 4, dass sich die Rezeptionsmuster für den narrativen Teil des Hotel-Spots und für die Produktpräsentation bei allen Probanden gleichförmig unterscheiden: Die Blickbewegungen sind bei der Produktpräsentation deutlich delinearisiert und inhomogener, mit häufigen Wechseln zwischen den Aufmerksamkeits-

bereichen. Im Erzählteil dagegen folgen die Fixationen über alle Probanden hinweg deutlich ausgeprägt der filmischen Umsetzung der Geschichte.

ABBILDUNG 4
Sequenzchart von 20 Probanden für den Hotelspot



Die Begrenzung auf 20 Probanden ist darstellungsbedingt. Für die Auswertung wurden alle Probanden berücksichtigt.

Diese Verteilung der beiden Muster findet sich – bis auf einige zeitliche Unterschiede, auf die ich noch eingehen werde – auch bei der Probanden-

gruppe, die die Spots ohne Ton rezipiert hat. Daraus lässt sich schließen, dass es die jeweilige visuelle Umsetzung in den beiden Teilen ist, die ein unterschiedliches Aneignungsmuster bewirkt: So hat die narrative Struktur der Visualisierung in der Erzählsequenz eine deutlich stärker fokussierte Aufmerksamkeitsverteilung zur Folge als die repräsentationale Struktur der Visualisierung im Produktteil.

Die einheitliche Dynamik der Blickbewegungen ist im Erzählteil des U-Bahn-Spots deutlich schwächer ausgeprägt, was auf eine andere filmische Gestaltung und eine andere Struktur des Plots zurückgeführt werden kann: Bis auf die Eröffnungssequenz werden Einstellungen gewählt, die den Protagonisten im Kontext mit anderen – ebenfalls sitzenden – Personen oder im Kontext des gesamten U-Bahn-Waggons zeigen. Da außerdem in der Story keine neuen Personen eingeführt werden und keine Person in Bewegung gezeigt wird – wie im Hotelspot der Kellner –, eröffnet der U-Bahn-Spot dem Rezipienten offensichtlich mehr Selektionsmöglichkeiten für die Aufmerksamkeitsverteilung und damit eine stärker interaktive Aneignung.

Diese unterschiedliche Konzentration der Wahrnehmung drückt sich auch quantitativ aus: So entfallen im Hotelspot rund 60 Prozent der Fixationszeit auf den Protagonisten und seine Augen, während das im U-Bahn-Spot nur 50 Prozent sind. Knapp 17 Prozent der Fixationszeit entfallen dagegen auf die beiden Frauen, die links und rechts vom Protagonisten des U-Bahn-Spots sitzen, während der Hotelkellner mit gut 8 Prozent deutlich weniger Aufmerksamkeit erhält.

Die Annahme, dass »the earliest fixations will be determined by the visual properties of the objects« (HENDERSON/FERREIRA 2004: 34), wird, wie Abb. 4 zeigt, bestätigt durch die Blickdaten im Falle neu eingeführter Gegenstände und Personen. So korreliert die Erstfixation des Kellners bei allen Probanden signifikant mit seinem dynamischen Eintritt ins Blickfeld. Im U-Bahn-Spot dagegen wird die rechts vom Protagonisten ›bewegungslos‹ sitzende Frau überhaupt nur von einem Drittel der Probanden wahrgenommen, während die Erstfixation der Frau links vom Protagonisten zeitlich gestreut stattfindet. Ein Steuerungsmittel für die Aufmerksamkeitsverteilung ist offensichtlich auch der Bildschnitt: So verschiebt sich die Aufmerksamkeit in beiden Spots in dem Moment auf das Handy und den Handybildschirm, in dem der entsprechende Umschnitt erfolgt, was der bisherigen Szene einen neuen Rahmen gibt. Dass Bewegungen und Veränderungen im Gesichtsfeld einen Fixationswechsel und eine Ver-

schiebung der Aufmerksamkeit auslösen, entspricht auch Befunden aus einer Studie zur Rezeption von wissenschaftlichen Präsentationen, bei denen begleitend Projektionen – z. B. Powerpoint – eingesetzt werden (vgl. BUCHER/KRIEG/NIEMANN 2010). Insgesamt ist die zeitliche und räumliche Synchronisation von Filmdynamik und Fixationsdynamik signifikant, allerdings im Hotelspot deutlich enger, was auf die genannten strukturellen Unterschiede in der Storyentwicklung einerseits und der filmischen Gestaltung andererseits zurückzuführen ist.

Für eine Erklärung des multimodalen Verstehens lassen sich aus diesen Daten folgende Schlüsse ziehen: In der Anfangsphase der Rezeption einer Szene und im Falle von Veränderungen im Gesichtsfeld der Rezipienten tragen auch Salienzfaktoren zur Erklärung von Wahrnehmungsmustern bei. Allerdings sind die Salienzfaktoren, die für stehende Abbildungen vorgeschlagen wurden, wie Farbe, Kontrast, Anordnung oder Linienführung für bewegte Bilder um die filmischen Gestaltungsmittel wie Schnitt, Zoom, Einstellung, Kamerafahrten, aber auch um dynamische Elemente im Betrachtungsobjekt selbst – hier z. B. der auftauchende Kellner – zu erweitern. Filmische Gestaltungsmittel sind in der Lage, die Aufmerksamkeit der Betrachter zu lenken, während unvermittelt eingeführte Gegenstände und Personen eine neue Szene schaffen, in der auch neu festzulegen ist, was optisch informativ sein kann.

5. Die Logik des modalen Sinns: Vom Sehen zum Verstehen

5.1 *›Sehen als‹: Die semantische Dimension des Visuellen*

Blickdaten öffnen zwar ein Fenster zur Aufmerksamkeitsverteilung, nicht aber zum Prozess des Verstehens. Zu wissen, wohin jemand schaut, bedeutet nicht, auch zu wissen, was er sieht. Auch die Gründe für das Anschauen eines bestimmten Bereichs im Wahrnehmungsraum werden mit den Blickdaten nicht geliefert. Dass jemand wiederholt und entsprechend lang einen bestimmten Gegenstand betrachtet, könnte sowohl durch starkes Interesse als auch durch Verstehensprobleme motiviert sein. Ein Großteil der Geschichte der Blickaufzeichnungsforschung lässt sich als Versuch verstehen, diese Kluft zwischen dem ›Wo(hin)‹, dem ›Warum‹ und dem ›Was‹ der Wahrnehmung zu schließen (DUCHOWSKI 2007: 14; HENDERSON/BROCKMOLE et

al. 2007; HOLSANOVA 2008: 81ff.). Eine der Möglichkeiten, diese Lücke zu schließen, besteht darin, neben den *Blickdaten* auch *Äußerungsdaten* des Lauten Denkens zu erheben. Damit wird auch die Untersuchungsmethode selbst multimodal (vgl. HOLSANOVA 2008: 94ff.) und eröffnet einen bedeutungsreicheren und informationshaltigeren Zugang zum multimodalen Verstehen. »By using ›two windows to the mind‹, we obtain more than twice as much information about cognition, since vision and spoken language interact with each other« (ebd.: 94).

Für die Rezeptionsanalyse zu den beiden Werbespots wurden zwei Typen sprachlicher Äußerungen erhoben: Erstens die simultan zur Rezeption geäußerten *Spontankommentierungen* des Lauten Denkens und zweitens die nachträglich formulierten *Wiedergaben* des jeweils gesehenen Videos. Während die Spontankommentierungen darüber Aufschluss geben, wie die Probanden einzelne Elemente, Sequenzen und gestalterische Aspekte der Videos verstanden haben, liefern die Wiedergaben Hinweise auf den Gesamtsinn der beiden Videos. Die Auswertung zeigt, dass der Zusammenhang zwischen den Blickdaten und den Äußerungsdaten mehr-mehrdeutig ist – also in beide Richtungen offen: Einerseits korrelieren Äußerungsdaten und Blickdaten, sodass die Äußerungen für die Deutung der Blickdaten genutzt werden können. Andererseits ist der Zusammenhang zwischen dem Betrachteten und dem Verstandenen auch willkürlich, da sich Unterschiede in der Sichtweise und im Verständnis der Spots nicht in den Blickdaten widerspiegeln müssen. Als Basis für die Auswertung der Blickdaten und der Äußerungsdaten im Hinblick auf das Verstehen der in den beiden Spots jeweils erzählten Geschichte lassen sich folgende Verstehenskriterien formulieren: Die Probanden sollten

- erkennen, wo die Geschichte spielt,
- die Handlungen des Protagonisten verstehen,
- die subjektive Realität des Protagonisten erkennen,
- den Übergang in die objektive szenische Realität erkennen,
- die Funktion des Handys als Story-Element verstehen,
- die Funktion des Spots als Werbung für ein LG-Handy erkennen,
- verstehen, für welches Produkt geworben wird.

Den wichtigsten Anhaltspunkt für die Lösung dieser Teilaufgaben erwarten sich die Probanden offensichtlich von den jeweiligen Protagonisten selbst. Auf sie entfällt über alle Probanden hinweg die weitaus längste Betrachtungszeit mit über 50 Prozent der Gesamtfixationszeit (siehe Abb. 2). Im Szenario ohne Ton liegen die Werte sogar über 60 Prozent. Auch die

Augen der Protagonisten werden länger betrachtet, wenn der Spot ohne Ton angeschaut wird. Die Relevanz dieser beiden AOIs für das Verständnis des Spots steigt offensichtlich durch die Abwesenheit der gesprochenen Sprache und des Tons. *Als was* die Probanden den Protagonisten allerdings sehen, ist entsprechend ihren Äußerungen beim Lauten Denken sehr unterschiedlich. Für den Protagonisten des Hotelspots finden sich folgende Bezeichnungen: *er*, *ein Mensch*, *ein Geschäftsmann*, *ein gut gekleideter Herr*, *der junge Mann*, *so 'ne Art Manager* oder aber Beschreibungen wie *er sieht am Anfang ein bisschen wütend aus*. Im U-Bahn-Spot ist die Varianz der Kennzeichnungen für den Protagonisten einheitlicher. Hier dominieren Bezugnahmen mit dem Pronomen *er* oder unspezifische Kennzeichnungen wie *ein junger Mann*.

Relevant sind diese Kennzeichnungsunterschiede insofern, als sie nicht nur die Offenheit einer Personenidentifizierung bei gleichem optischen Stimulus demonstrieren, sondern auch jeweils für unterschiedliche Erzählungen stehen: Mit den Kennzeichnungen der Akteure geht der Erzählende bestimmte Festlegungen ein, die mit jeweils unterschiedlichen Darstellungsaufgaben verifiziert werden müssen (vgl. FRITZ 1982: Kap. 6.2; BUCHER 1991: 51 - 52). Wird der Protagonist als ›Geschäftsmann‹ oder ›Manager‹ bezeichnet, so kann die erzählte Geschichte eine andere sein, als wenn er als ›junger Mann‹ agiert oder ›am Anfang ein bisschen wütend aussieht‹. Kennzeichnungen der Akteure sind folgenreiche Trigger für die Art der zu erzählenden Geschichte.

Noch stärker variieren die Beschreibungen der zentralen Handlung der Protagonisten, um die sich der ganze Spot dreht: die Nutzung des Handys. Sie reichen von unspezifischen Handlungsbeschreibungen (*hat das Handy in der Hand*) über Formen des Schauens ohne Objektangabe (*beobachtet etwas*), mit adverbialer Angabe (*starrt gebannt*), mit Objektangabe (*schaut auf sein* LG-*Handy*) oder mit propositionaler Angabe (*schaut sich auf seinem Handy ein Fußballspiel an*) bis zu reflexiven Beschreibungen (*Man denkt zuerst, dass er Fernsehen schaut, dann stellt sich heraus, nee, es ist das Handy*). In den Nacherzählungen der Probanden ›ohne Ton‹ kommen hauptsächlich unspezifische Handlungsbeschreibungen vor, mit denen sich der eigentliche Witz der Geschichte, der den Slogan *Fernsehen wie zu Hause* vorbereitet, gerade nicht ausdrücken lässt. In den Nacherzählungen der Probanden, die den Spot mit Ton gesehen haben, kommen im Grunde nur zwei Beschreibungsvarianten vor: *schaut Fernsehen* oder *schaut sich ein Fußballspiel (auf seinem Handy) an*. Das sind offensichtlich die Beschreibungen, die in

die Logik der Geschichte passen und in denen die o. a. Verstehenskriterien 1-7 erfüllt sind. Die Tatsache, dass die meisten der Probanden, die die Spots mit Ton gesehen haben, diesen zentralen Aspekt verstanden haben, deutet bereits auf die Relevanz der auditiven Informationen hin, die im Abschnitt 6 ausführlicher behandelt wird.

Eine weitere zentrale Episode des Hotelspots ist das Auftreten des Kellners, der in allen experimentellen Szenarien fast die gleiche Aufmerksamkeit erhält, mit bis zu 10 Prozent der gesamten Fixationszeit. Auch wenn alle Probanden diese Sequenz mit einer hohen Intensität betrachtet haben, fallen ihre Beschreibungen doch sehr unterschiedlich aus. In der folgenden Auflistung sind die Beschreibungen der Gruppe, die das Video ohne Ton gesehen hat, in vier Varianten eingeteilt, die sich auch in der jeweils eingenommenen Perspektive unterscheiden:

1. Perspektive auf den Kellner: Handlungsbeschreibungen des Kellners:
   Dann kommt halt der Kellner, schaut im Prinzip aus dem Hintergrund, was er grade eigentlich isst [der Protagonist knabbert Granulat aus dem Blumentopf, HJB].
2. Perspektive auf den Kellner: Beschreibung einer interaktiven Handlung des Kellners:
   Der Ober geht vorbei und schaut ihn [den Protagonisten, HJB] *an.*
3. Perspektive auf den Protagonisten: Beschreibung der interaktiven Handlung des Protagonisten:
   …raunt da irgendeine Zwischenbemerkung zum Kellner, die scheinbar mit dem Fußballspiel zu tun hat.
4. Perspektive auf beide Akteure: Beschreibung der interaktiven Handlungen beider Personen:
   Dann Ober kommt vorbei, er sagt irgendetwas, gibt ihm irgend ne Anweisung wahrscheinlich. …bestellt sich wohl n' Bier beim Kellner. Der versteht ihn nicht so richtig.

Die Unterschiede in den Beschreibungen sowohl des Protagonisten als auch der Kellner-Episode machen deutlich, dass Wahrnehmung immer auch eine semiotische Dimension aufweist. Sinn und Bedeutung der Objekte und Regionen im Blickfeld haben ebenso Einfluss auf die Aufmerksamkeitsdynamik wie Intentionen und die Salienzfaktoren, wenn auch jeweils in anderen Rezeptionsphasen (HENDERSON/FERREIRA 2004: 30-36). Die folgenden dynamischen Blickdaten zeigen, in welcher Weise die Semiotik des Filmbildes die Aufmerksamkeit lenken kann (Abb. 5 Blickverlauf).

5.2 *Der implizite Rezipient: Perspektivenübernahme als Strategie der Sinnerschließung*

Es ist ein zentraler Befund von Blickuntersuchungen in realen Aufgabenszenarien wie z. B. ›ein Sandwich bereiten‹, ›Tee kochen‹, oder ›einen Ball fangen‹, dass der Blick antizipatorisch eingesetzt wird: Fixationen gehen den entsprechenden Bewegungen immer voraus (HAYHOE et al. 2007: 644). Die Aufgabe im Falle der beiden Videos besteht darin, in einem begrenzten Zeitrahmen die beiden Spots als filmische Erzählung und als Werbung für ein Handy zu verstehen. Dementsprechend finden sich auch in den Blickdaten zu den beiden Videos solche antizipatorischen Strategien der Aufmerksamkeit, die sich auf Erfahrungen und Vorwissen zurückführen lassen, die in Lernprozessen zur Rezeption von Werbespots und zur Struktur von Erzählungen erworben wurden.

ABBILDUNG 5
Scanpfad im Hotelspot: Der Blick des Probanden folgt dem Blick des Protagonisten in Richtung Handy

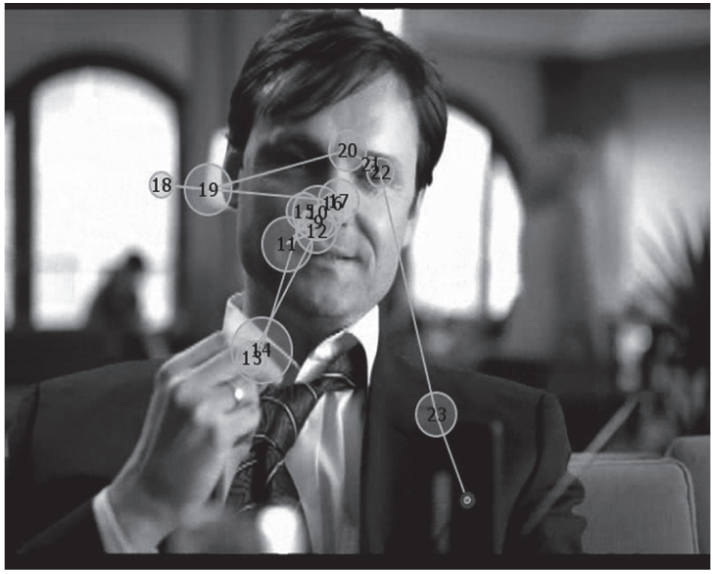


ABBILDUNG 6

Scanpfad im U-Bahn-Spot: Proband folgt dem Blickwechsel zwischen den beiden Akteuren im Spot

Die Verschiebung der Blickmarkierungen ist durch die Kamerabewegung und die damit verbundene relative Positionsveränderung der Probanden bedingt.

In diesen Scan-Pfaden ist eine Aneignungsstrategie erkennbar, die darauf abzielt, das Erzählte aus der Perspektive des Protagonisten mittels rezeptiver Rollenübernahme zu verstehen. Die Blickverläufe, wie sie in Abb. 5 und 6 dokumentiert sind, lassen sich durch eine Salienztheorie nicht erklären. So springt der Blick des Rezipienten im Falle des Hotelspots (Abb. 5) in einen Bereich des Filmbildes, der optisch unstrukturiert und unauffällig ist. Auch die Blickbewegungen der Probanden, die das Video ohne Ton anschauen, haben an dieser Stelle des Films dieselbe antizipatorische Tendenz: Die Hälfte von ihnen fixiert das Handy bereits, während das Zoom aufgezogen wird und es nur von hinten zu sehen ist. Das heißt: Sie antizipieren den Blick des Protagonisten und fixieren das, wovon sie annehmen, dass auch er es fixiert. Die Rollenübernahme ist im Falle der Rezeption ohne Ton eine noch stärker in Anspruch genommene Erschließungsstrategie.

Im Falle des Blickverlaufs in der U-Bahn (Abb. 6) ist es mithilfe einer Salienztheorie nicht möglich, die wechselweise wiederholte Fixierung der Augen

der beiden Protagonisten zu erklären, da dort ja keine neue Information zu erwarten ist. Die Erklärung für die Blickverläufe in den beiden Abbildungen liefert die Logik der Erzählung selbst: Die Probanden fixieren diejenigen Stellen, die aus der erzählten Geschichte heraus relevant sind. Das sind diejenigen Stellen, die auch der Protagonist durch sein Blickverhalten als relevant markiert: im einen Fall die Augen der Frau, der er gleich um den Hals fallen wird, im andern das Handy, auf dessen Display er starrt. Die Aufmerksamkeit folgt also kommunikativen – oder hier: narrativen – Kriterien.

Die Blickdaten zeigen, wie der Rezipient zum ›impliziten‹ Rezipienten im Sinne Isers wird (vgl. ISER 1994: 66ff.). Er greift die Signale des Videos auf, um diejenige Perspektive einzunehmen, die durch Machart und Struktur des Videos vorgegeben wird. Anhand der Blickdaten lassen sich vier Darstellungsstrategien rekonstruieren, mit denen der Rezipient zu einer Perspektivenübernahme veranlasst wird:

Erstens signalisiert der sogenannte *Eyeline-Match* (vgl. z. B. MIKOS 2003: 219; ZETTL 1999: 192f.), also der Blick des Protagonisten auf etwas außerhalb des Filmbildes, dem Rezipienten, dass dort etwas zu sehen ist, was für den Protagonisten relevant ist. Der Rezipient antizipiert die Blickrichtung und errechnet so das vom Protagonisten fixierte Objekt. Im U-Bahn-Spot kann er sich sogar an der Blickrichtung von drei Personen orientieren und deren Schnittpunkt errechnen (Abb. 7).

Zweitens: Der *Standpunkt der Kamera*, von dem aus in der Eingangssequenz des Hotelspots das Zoom aufgezogen wird, entspricht genau der Position des Handys, das der Protagonist in der Hand hält und auf das er schaut. Der Blickwinkel der Kamera auf den Protagonisten liefert insofern auch den visuellen Hinweis, was der Protagonist anschaut. Der Kameraschwenk im U-Bahn-Spot, der das Handy ins Bild führt, erfüllt dieselbe aufmerksamkeitssteuernde Funktion, nämlich: »to clarify and reinforce the point of view of the people appearing on-camera, at what and in what direction they are looking« (vgl. ZETTL 1999: 200).

Drittens wird die Perspektivenübernahme der Rezipienten durch den *Point-of-View-Schnitt* gesteuert: Wenn mit Ende des Zooms auf den Handybildschirm, auf den der Protagonist schaut, umgeschnitten wird, zeigt sich dem Rezipienten genau das Blickfeld des Protagonisten. Die Tatsache, dass dabei jeweils seine Finger zu sehen sind, die das Handy halten, sorgt für eine visuelle Kohärenz zwischen den beiden Einstellungen. Als »Visual Linking Device« (LIM 2007: 202) zeigen sie, dass es das Handy des Protagonisten ist, das zu sehen ist, und verifizieren nachträglich den antizipa-

ABBILDUNG 7
Scanpfad im U-Bahn-Spot: Der Proband folgt den Blicken der drei Akteure und errechnet aus ihrem Blickwinkel den Ort des Gegenstandes, der für sie relevant ist: das Handy

torischen Blickverlauf. Die ›subjektive Kamera‹ der zweiten Einstellung veranlasst den Rezipienten, vom Beobachter zum Teilnehmer der Szene zu werden (vgl. ZETTL 1999: 192), um ihn in die »mentalen und emotionalen Prozesse des blickenden Akteurs ein[zu]beziehen« (MIKOS 2003: 219).

Ein *viertes* Gestaltungsmittel, auf das in Kapitel 6 noch genauer eingegangen wird, ist die *szenische Vertonung* dieser Sequenz. Die Soundcollage aus Reporter-Kommentar und Stadion-O-Ton, wie sie für eine Fußball-Live-Übertragung typisch ist, gibt dem Rezipienten einen inhaltlichen (propositionalen) Hinweis auf das, *was* der Protagonist betrachtet. In den beiden Filmen ist nicht anzunehmen, dass der Ton der Fußballübertragung aus einem realen Stadion in der Nachbarschaft oder aus einem Radiogerät kommt. Diese Deutung wird durch die Position der Protagonisten im Raum – Hotel-Lobby und U-Bahn-Waggon – sowie deren spezifischen Blick, den wir als ›Fernsehblick‹ erkennen können, ausgeschlossen. Unterstützt wird diese Deutung durch die Gestik und Mimik der Protagonisten, die als typische Fernseh-Rezeptionssignale engagierten Zuschauens identifizierbar

sind. Aufgrund dieses dominanten Fernseh-Framings der Szene bemerkt keiner der Probanden, dass die Vertonung der Fußballszene tatsächlich ein Radio-Kommentar ist, gesprochen von einem bekannten Radiomoderator.

Die Blickdaten zeigen durch ihre quantitative Ausprägung, ihre zeitliche Dynamik und die entsprechenden Scan-Pfade, wie sich die Rezipienten entlang der beschriebenen Gestaltungsstrategien die Perspektive des Protagonisten schrittweise auf der Basis der verfügbaren Indizien erarbeiten. Mit dem Konzept des ›impliziten Lesers‹ als ›transzendentalem Modell‹ für die Erklärung der Textrezeption hat Iser gezeigt, dass solche Übernahmen einer Perspektive auf das im Text – oder im Film – Dargestellte konstitutiv sind für jeden Rezeptionsprozess (ISER 1994: 66). Die Funktion des Textes bestehe darin, dass er »ein bestimmtes Rollenangebot für seine möglichen Empfänger parat [hält]« (ebd.: 61), aus dem sich die relevanten Perspektiven auf den Text ableiten lassen. Die Rekonstruktion einer Perspektive erfolgt mithilfe der Fragen: Welche Person an welchem Ort des gezeigten Raumes kann das, was gezeigt wird, sehen, die Töne, Geräusche oder das Gesagte hören oder über das entsprechende Wissen verfügen? Die Übernahme der Perspektive des Protagonisten durch den Zuschauer ist Teil der intendierten Werbewirkung der beiden Spots: Der Rezipient soll nicht nur sehen, wie eine andere Person das LG-Handy zum Fernsehschauen nutzt, sondern er soll die Erfahrung des Fernsehschauens auf einem Handy mit dem Anschauen des Films selbst machen. Der Spot hat damit auch eine *Simulationsfunktion*, die es dem potenziellen Käufer erlaubt, die Nutzung des beworbenen Produktes virtuell auszuprobieren. Das funktioniert im Falle medialer Produkte besonders gut, nicht aber im Falle von Getränken, Nahrungsmitteln, Putzmitteln etc. Die Blickdaten mit ihrer starken Ausrichtung am Protagonisten und speziell an seinen Augen machen deutlich, dass insbesondere der Hotel-Spot es hervorragend zu leisten vermag, diese Simulation einer virtuellen Fernsehsituation durch die Perspektivenübernahme hervorzurufen.

6. Die Logik des intermodalen Sinns: Der Ton macht die Geschichte

6.1 *Man sieht, was man hört*: *auditive Aufmerksamkeitssteuerung*

Es gilt als eine der zentralen Fragen für die Analyse multimodaler Kommunikationsformen, in welcher Weise die jeweils kombinierten Modi zur

Erzeugung des Gesamtsinns zusammenspielen. Die verschiedenen Multimodalitätstheorien stimmen darin überein, dass der Gesamtsinn eines Kommunikationsangebotes mehr ist als die summierte Bedeutung seiner modalen Elemente und dementsprechend der Gesamtsinn nicht additiv, sondern ›multiplikatorisch‹ als intersemiotischer Prozess zu erklären ist (LEMKE 1998; O'HALLORAN 1999; LIM 2004, 2007). Eine experimentelle Möglichkeit, das Problem der Kompositionalität zu bearbeiten, besteht darin, die Leistung eines Modus indirekt sichtbar zu machen, indem er ›abgeschaltet‹ wird. Durch einen Vergleich mit den Rezeptionsbefunden aus den vollständigen Kommunikationsangeboten lässt sich dann der kommunikative Beitrag rekonstruieren, den der entsprechende Modus zur Herstellung des Gesamtsinns im Normalfall leistet.

In der vorliegenden Studie wurden dafür zwei Gruppen von 16 und 8 Probanden die beiden Videos *ohne Ton* gezeigt. Die Probanden waren dabei angehalten, laut zu denken, also das, was sie sahen, simultan zu formulieren. Neben den Blickdaten sind somit für die Auswertung auch interpretierende Äußerungsdaten verfügbar.[1] Durch das Abschalten des Tons werden verschiedene Modi ausgeblendet: der Sound, die gesprochene Sprache der Protagonisten, die gesprochene Sprache aus dem Off. Beim *Vergleich der Rezeptionsdaten* dieser Probandengruppe mit denjenigen, die beim Zeigen der Videos mit Ton entstanden sind, zeigen sich deutliche Unterschiede, die Rückschlüsse erlauben, wie die Modi zum kompositionellen Gesamtsinn der beiden Spots beitragen.

Das kommunikative Zusammenwirken von szenischen Sounds und szenischer Sprache ›im Film‹ einerseits und dem Bewegtbild andererseits wird sehr deutlich in der Kellner-Episode erkennbar. Die *Blickdaten* der Probanden ›ohne Ton‹ unterscheiden sich in dieser Sequenz erheblich von den Blickdaten der Probanden ›mit Ton‹:[2]

1. Die Fixationen auf den Kellner erfolgen bei der Gruppe ›ohne Ton‹ bedeutend später: im Durchschnitt 160 ms, was in etwa eine Verzögerung von zwei informationsaufnehmenden Fixationen bedeutet. Geht man vom statistischen Mittelwert, also dem Wert in der Mitte

1 Aus technischen Gründen sind die Äußerungsdaten nur für die Gruppe der 16 Probanden verfügbar.

2 Bei den folgenden Daten wurde auch die Erhebung miteinbezogen, für die nur Blickdaten vorliegen, aber keine Äußerungsdaten. Dementsprechend beträgt die Anzahl der Probanden mit Ton n = 26, die der Probanden ohne Ton n = 24.

der beiden Extremwerte, aus, wird der Kellner bei der Gruppe ›ohne Ton‹ sogar 643 ms später angeschaut.

2. Die durchschnittliche Dauer *einer* Fixation beim Betrachten des Kellners ist bei der Gruppe ohne Ton länger, im Schnitt um 104 ms je Fixation, mit einer maximalen Differenz von 320 ms.
3. Probanden ›ohne Ton‹ fixieren den Kellner durchschnittlich um 216 ms länger.
4. Der Blickwechsel vom Kellner zur nächsten AOI, dem Protagonisten oder seinem Handy, erfolgt im Schnitt bei den Probanden ›ohne Ton‹ 145 ms später.

Eine erste Erklärung für diese Ausprägungen der Blickverläufe liefert Abb. 8: Sie zeigt für einen der Probanden ›mit Ton‹, dass es der Klingelton ist, der den Fixationswechsel vom Kellner in Richtung des Protagonisten auslöst, und zwar bereits vor dem filmischen Umschnitt auf ihn. Mit einer optischen Auffälligkeit, also mittels des Salienz-Modells, wäre diese Verschiebung der Aufmerksamkeit nicht zu erklären, da die Region, in die der Blick wandert, visuell uninformativ ist. Der Proband hört den Klingelton und sucht nach dem Gegenstand, der diesen Ton ausgelöst hat. Aus dem bisherigen Verlauf des Spots kann er schließen, dass das Klingeln vom Handy des Protagonisten stammt, was die Richtung seines Blickwechsels erklärt. Durch den Klingelton ist auch indiziert, von welcher Art der zu suchende Gegenstand ist, dass es sich nämlich um ein Handy handelt und eben um keinen anderen Gegenstand oder um keine andere Person. Den Probanden, denen der Ton nicht zur Verfügung steht, fehlen diese Indikatoren für eine Verschiebung des Fokus der Geschichte und damit auch ihrer Relevanz-Kriterien. Der Kellner bleibt für sie so lange relevant, bis ihn ein neuer Gegenstand durch visuelle Präsenz ablöst, was erst nach dem Umschnitt auf den Protagonisten der Fall ist.

Es handelt sich bei diesem Übergang um eine der Schlüsselstellen zum Verständnis des gesamten Spots: Mit dem Klingeln des Handys wird der Realitätswechsel aus der *subjektiven Welt des Protagonisten* in die *objektive, auktoriale Welt* der Hotel-Lobby eingeleitet. Nur das Zusammenspiel von Bild und Ton macht diesen Realitätswechsel erkennbar. Während die Zuschauer bis zu diesem Zeitpunkt die erzählte Handlung aus der Perspektive des Protagonisten verfolgt haben, werden sie an dieser Stelle durch das Klingeln des Handys und die entsprechende Reaktion des Protagonisten vom teilnehmenden Beobachter zum externen Beobachter.

Mit dem Einsetzen der Stimme aus dem Off kommt eine dritte Realitätsebene ins Spiel: die *Realität der Werbung*. Dieser zweite Rollenwechsel

ABBILDUNG 8
Scanpfad im Hotel-Spot: Der Klingelton löst den Blickwechsel aus

des impliziten Rezipienten in einen beworbenen Konsumenten ist die Voraussetzung dafür, dass die Stimme des Sprechers aus dem Off, die sich ab diesem Zeitpunkt mit dem szenischen O-Ton mischt, nicht als Bruch erlebt wird, sondern in den Plot des Spots eingebaut werden kann. Es sind die auditiven Modi des Videos, die an dieser Stelle die Realitätsebenen sichtbar machen und in Verbindung setzen: die subjektive Welt des Protagonisten, repräsentiert durch seine Äußerungen am Telefon (*Hallo, Ja, Ja*) und die kollektive Konsumenten-Welt der Rezipienten, repräsentiert durch die gleichzeitig formulierte und ›auf Lücke‹ geschnittene Produktwerbung durch den Off-Sprecher. Ohne diese Collage von szenischer Sprache und Sprache aus dem Off ist dieses Vexierbild zweier Realitätsebenen nicht zu sehen. Die Bilder alleine können das in dieser Sequenz nicht leisten.

Für die meisten Rezipienten, die den Spot ohne Ton sehen, bleibt die erzählte Welt die einzige Realitätsebene. Wie aus den Nacherzählungen ersichtlich wird, verstehen jedoch auch einige Rezipienten ohne Ton den Übergang aus der subjektiven Realität des Protagonisten in die szenische Realität der Hotel-Lobby, der mit der Entgegennahme des Anrufs eintritt. Sie können sich dabei auf Mimik und nicht sprachliche Handlungen des Protagonisten

stützen: seine Überraschungsmimik und den Abbruch seiner Esshandlung, als er bemerkt, dass er nicht Nüsse, sondern Tongranulat aus dem Blumentopf isst. Beides drückt aus, dass er jetzt der Realität der Hotel-Lobby gewahr wird. Da bei der Rezeption ohne Ton die Werbebotschaft aus dem Off allerdings nicht gehört werden kann, geht für diese Rezipienten auch nach dem Zeitpunkt des Klingelns die Geschichte ausschließlich auf fiktionaler Ebene weiter, bis die Produktwerbung sie durch einen harten Schnitt beendet. Da sie die Werbebotschaft *Fernsehen wie zu Hause* aus dem Off nicht hören können, entgeht ihnen auch das auditive ›Linking Device‹, das den narrativen Teil mit der Produktbeschreibung verkoppelt. Die Schwierigkeit, den Abschluss des narrativen Teils und den Übergang zur Produktwerbung ohne Ton zu verstehen, formuliert einer der Probanden sogar explizit:

> *»Also bis zum dem Plot, an dem er mit dem Kellner redet, fand ich's ganz plausibel, aber ab dann konnt' ich ohne Ton nicht mehr so nachvollziehen, warum er jetzt ans Handy geht, und warum der Werbespot dann auch schon vorbei war.«*

6.2 *Die Ökonomie der Aufmerksamkeit*: *ausblenden und übersehen*

Die auditiven Modi sind aber nicht nur aufmerksamkeitslenkend, sondern auch aufmerksamkeitsablenkend. Die Episode, in der der Protagonist in seiner geistigen Abwesenheit das Ton-Granulat aus dem Blumentopf als Nuss-Ersatz isst, wird von keinem der 30 Probanden, die den Spot mit Ton gesehen haben, in der Nacherzählung erwähnt. Dagegen findet sich diese Episode in mehr als 70 Prozent der Nacherzählungen der Probanden ohne Ton (12 von 17) entweder vollständig oder wenigstens zum Teil (*isst irgendetwas*). Wer weniger hört, sieht offensichtlich mehr.

Auch dieser Rezeptionsunterschied ist auf die modale Leistung der Vertonung des Spots zurückzuführen: Wie die oben zitierten Ausschnitte aus den Wiedergaben der Probanden ohne Ton zeigen, erkennt diese Probandengruppe die Granulatepisode vor allem im Schlussabschnitt des narrativen Teils, als der Protagonist selbst seinen Irrtum bemerkt. Verantwortlich dafür ist die ausschließliche Konzentration dieser Gruppe auf die erzählte Realität selbst. Die Probanden, die den Spot mit Ton hören, müssen an dieser Stelle die Verschränkung von zwei Realitätsebenen bewerkstelligen: die der erzählten Realität, die in den Telefonäußerungen des Protagonisten manifest wird, und die der Werberealität, die mit dem

Off-Kommentar eingeführt wird. Da die Äußerungen in den beiden Realitätsebenen ›auf Lücke‹ geschnitten sind und sich so kontinuierlich vermischen, entsteht für die Rezipienten eine komplexe Verstehenssituation, die offensichtlich zur Ausblendung eines Teils der erzählten Realität führt. Man kann diesen Befund als Phänomen der *kognitiven Ökonomie* interpretieren: Die Begrenztheit der kognitiven Kapazitäten hat zur Folge, dass die Aufmerksamkeit sowohl für bildliche als auch für sprachliche Stimuli scheinwerferartig organisiert wird und nur das fokussiert, was im Moment als relevant erscheint (NEUMANN 1992; NEUMANN 1996; LANG 2000; BUCHER/SCHUMACHER 2006; HOLSANOVA 2008: 83 - 84). Die Esshandlung des Protagonisten ist eine Nebenepisode und nur begrenzt relevant für die erzählte Geschichte, da sie keinen neuen Aspekt für ihr Verständnis liefert: Die Probanden mit Ton haben bereits verstanden, dass der Witz der Geschichte in dem Wechsel zwischen subjektiver und auktorialer Realität liegt. Sie brauchen die zusätzliche Information aus der geistesabwesenden Esshandlung zu diesem Zeitpunkt nicht mehr, um die subjektive Perspektive des Protagonisten aufzubauen. Für die Probanden ohne Ton ist sie dagegen ein relevanter Indikator.

Während die Behaltensleistungen für das beworbene Produkt für das Betrachten mit und ohne Ton nahezu gleich positiv ausfallen, wird der Realitätswechsel im narrativen Teil – wie bereits erwähnt – von 76 Prozent (13 von 17) der Probanden ohne Ton nicht erkannt, wobei diejenigen, die ihn bemerkt haben, den Spot bereits kannten. In der Gruppe der Probanden mit Ton sind die Verhältnisse fast genau umgekehrt: 64 Prozent (14 von 22) erkennen den Realitätswechsel und nur 36 Prozent erkennen ihn nicht.

Vergleicht man die *Wiedergabeleistungen* der Probanden mit und ohne Ton, so bestätigen sich die bisherigen Befunde. Probanden, die den Spot mit Ton angeschaut haben, geben eine *Nacherzählung*, während die Probanden ohne Ton eine *Inhaltsangabe* machen. Der Unterschied zwischen diesen beiden Textsorten besteht darin, dass die Nacherzählung den Witz der Geschichte, ihre Pointe enthält, während in der Inhaltsangabe aufgelistet wird, was man gesehen hat. Das zeigen die beiden folgenden Beispiele:

Nacherzählung (Proband mit Ton):

> *»Ja, äh, der Mensch, der ist so begeistert von dem Fernsehangebot und seinem Handy, dass er vergisst, dass er nicht daheim is, wo Schatz das Bier bringt, sondern in nem Restaurant, wo der Kellner hintendran rumspaziert, und erst als der Anruf kommt, wird er quasi wieder zurückgeholt in die Realität un merkt, dass es ja nur aufm Handy un nich wie da im Fernsehen. Also das* TV*-Handy bietet dasselbe Fernseherlebnis*

wie ein Heimat-Fernseher [lacht kurz].«

Inhaltsangabe (Proband ohne Ton):

»Also man sieht n gut gekleideten Mann, würde sagen Manager oder so, in ner Hotel-Lobby. Ähm, also ne Großaufnahme seines Gesichts, anfangs weiß man nicht, was er da genau tut, doch dann wird raus gezoomt, man merkt, dass eigentlich äh, ja, er sein Handy betrachtet voller Spannung und Fußball schaut. Dann kommt n Kellner vorbei, bei dem er sich wohl noch n Bier bestellt, der hat n leeres Bierglas auf seinem Tablett stehen und ja, durch die ganze Spannung, die er jetzt hatte, also, ähm, hat er gar nicht gemerkt, dass er, anstatt Popcorn wie zu Haus zu essen, einfach so die Blumentopferde gegessen hat.«

Die Besonderheit der Nacherzählung besteht darin, dass die Geschichte aus der Perspektive des Protagonisten erzählt wird, als personale Erzählung (*dass er vergisst, wo Schatz das Bier bringt, wird er zurückgeholt in die Realität*), während die Inhaltsangabe aus einer Beobachterperspektive erfolgt: Der Proband beschreibt, was er sieht, aber nicht, wie er die Geschichte versteht (*man sieht 'ne Großaufnahme seines Gesichts*). Die unzutreffende Deutung, dass er sich beim Kellner ein Bier bestellt, erfolgt offensichtlich auf der Grundlage eines Schemas, das der Proband für die Situation ›Warten in einer Hotel-Lobby mit Ausschank‹ zur Verfügung hat. Während in der Erzählung Zusammenhänge zwischen den Episoden hergestellt werden (*erst als der Anruf kommt*) und Schlussfolgerungen gezogen werden (*Also das TV-Handy…*), hat die Inhaltsangabe eine dominant additive Struktur (*man sieht – dann kommt – und ja*).

Bemerkenswert ist allerdings, dass die Beobachtungen in der Inhaltsangabe bedeutend detaillierter sind: das Aussehen des Protagonisten, das Tablett des Kellners, die filmischen Gestaltungsmittel (›Großaufnahme‹, ›rausgezoomt‹). Signifikant unterschiedlich sind auch die Abschlussteile der beiden Wiedergaben. Der Proband mit Ton formuliert in eigenen Worten und mit witzelnder Distanz die Werbebotschaft, die mit dem Spot verbreitet werden soll (*das TV-Handy bietet dasselbe Fernseherlebnis wie ein Heimat-Fernseher*). Der Proband ohne Ton bleibt im Wiedergabemodus und formuliert einen Schluss der erzählten Geschichte. Dass er dafür die Granulat-Episode heranzieht, ist kein Zufall: Diese Handlung des Protagonisten liefert beim Betrachten des Spots ohne Ton den einzigen Anhaltspunkt für dessen Verwechslung der Realitätsebenen. Für den Probanden mit Ton ist die Diskrepanz zwischen der subjektiven Realität des Protagonisten und der szenischen Realität der erzählten Geschichte schon aus der geistesabwesenden Anrede des Kellners als ›Schatz‹ erkennbar geworden. Er kann

dementsprechend die Armbewegung, mit der nach dem Granulat im Blumentopf gegriffen wird, als nicht relevant ausblenden.

Die Analyse der Blick- und Äußerungsdaten zu den beiden Spots zeigt deutlich die Spezifik multimodaler Kommunikationsangebote: Die einzelnen Modi kontextualisieren sich gegenseitig. Ohne den Ton – Sound, Geräusche, szenische Sprache – sehen die Probanden etwas anderes als mit Ton. Der Ton seinerseits braucht aber die Visualisierung als Kontextualisierung. Der Witz der Äußerung *Schatz bringst du mir noch'n Bier?* funktioniert nur auf der Grundlage der visuellen Darstellung der Kellner-Szene. Wie die Wiedergaben der Probanden ohne Ton zeigen, gehen sie davon aus, dass der Protagonist entweder etwas beim Kellner bestellt oder einen Kommentar zum Fußballspiel an ihn richtet. Sie nehmen dabei Bezug auf Interaktionsschemata, die für die gezeigte Situation erwartbar sind.

Aus diesen Gründen sind die auditiven Modi des Films nicht akustisches Beiwerk oder Begleitung, sondern dienen dem Vollzug von Erzählhandlung mit jeweils spezifischen Funktionen. Der szenische Ton liefert – analog zur szenischen Beschreibung in einer Reportage – die Identifizierung des Schauplatzes, die szenische Sprache dient der Redewiedergabe, die Sprache aus dem Off der Kommentierung des Erzählten. Über diese Funktionen hinaus ist der Ton ein überaus wirksames kommunikatives Mittel, um den Rezipienten in eine bestimmte Perspektive zum Erzählten zu bringen. Ton, Geräusche und Gesprochenes werfen in Video und Film immer die Frage auf: Wer kann das von welchem Standpunkt aus hören? Es ist diese Frage, die die Blickbewegungen und damit die Aufmerksamkeit und die Selektionsleistung des Rezipienten steuert. Das Auditive liefert deshalb auch Kriterien dafür, was visuell relevant ist. Die Wahrnehmung als Äußerungen aus dem Off – wie im Falle der Werbebotschaft der Spots – hat deshalb keinen Einfluss auf die Blickbewegung, weil die Rezipienten davon ausgehen, dass nur sie diese Äußerungen hören, nicht aber die Protagonisten in der gezeigten Szene. Der Witz des Videospots ist für die Probanden ohne Ton deshalb so schwer zugänglich, weil ihnen die Perspektivierungshilfe von Sound und gesprochener Sprache nicht zur Verfügung steht. Ohne Perspektivenwechsel ist der Wechsel der Realitätsebene kaum erkennbar.

In ihrem Überblick zu den bisherigen Erkenntnissen der Blickaufzeichnungsforschung stellen Henderson und Ferreira fest: »The influence of linguistic inputs as another possible source of top-down contextual information is an important issue that has not been systematically studied yet«

(HENDERSON/FERREIRA 2004: 29). Auch wenn dieser Zusammenhang in mancher Hinsicht noch Rätsel aufgibt, so kann aus den vorliegenden Befunden immerhin abgeleitet werden, dass erstens die gesprochene Sprache in multimodalen Erzählungen zum Aufbau der Erzählperspektive beitragen kann und dass sie zweitens auch dazu führen kann, dass Aspekte der visuellen Modi ausgeblendet werden. Genauso wichtig für die Perspektivierungsleistung ist aber der Sound eines Films oder Videos: Man kann zwar über etwas hinwegsehen, viel schwieriger ist es allerdings, über etwas hinwegzuhören, da man Schallwellen nicht durch Verschließen der Ohren ausblenden kann. Die auditive Wahrnehmung läuft im Normalfall automatisch mit (vgl. GIBSON 1973: Kap. v).

7. Schlussfolgerungen: Plädoyer für eine handlungstheoretische Multimodalitätstheorie

Die vorgestellten Rezeptionsbefunde geben Aufschluss darüber, wie die beiden Grundprobleme einer Theorie der Multimodalität – das Kompositionalitätsproblem und das Problem des multimodalen Verstehens (das Rezeptionsproblem) – gelöst werden können. Sie tragen darüber hinaus auch dazu bei, die Reichweite anderer Multimodalitätstheorien einzuschätzen. Man kann die bisherigen Theorien zur Multimodalität in zwei Traditionen einteilen: erstens die *systemfunktionale Diskursanalyse*, die auf die funktionale Grammatik von Halliday zurückgreift und dessen für die Sprache entwickelte ›Metafunktionen‹ (›modes of meaning‹), die repräsentationale, personale und textuelle Funktion, auf alle Kommunikationsmodi überträgt (IEDEMA 2003; LIM 2004; MATTHIESSEN 2007; BATEMAN 2008; O'HALLORAN 2008; JEWITT 2009); zweitens die *soziale Semiotik*, die auf einer kritischen Zeichentheorie basiert (KRESS/VAN LEEUWEN 1996; KRESS/VAN LEEUWEN 1998; KRESS/VAN LEEUWEN 2001; VAN LEEUWEN 2005; KRESS 2009). Letztere betrachtet Multimodalität auf der Basis einer allgemeinen Semiotik, die den Zeichengebrauch (›sign-making‹) ganz umfassend auch auf Architektur, das Design einer Ausstellung, Kinderzeichnungen, didaktische Arrangements oder die Ausstattung eines Kinderzimmers ausweitet (ausführlich zu diesen Theorieansätzen: BUCHER 2010a, 2010b). Gemeinsam ist beiden Theorietraditionen, dass sie Multimodalität zeichenorientiert zu klären versuchen.

Die Alternative dazu besteht darin, nicht am Zeichen, sondern an der Kommunikation selbst anzusetzen. Multimodal ist ein Kommunikati-

onsangebot nicht deshalb, weil es Zeichen unterschiedlichen Typs kombiniert, sondern weil die kommunikativen Handlungen – z. B. des Erzählens oder des Werbens – mit unterschiedlichen modalen Ressourcen vollzogen werden. Der Gegenstand einer handlungstheoretischen Multimodalitätsauffassung ist die Verwendung unterschiedlicher Typen von Zeichen als Mittel der Kommunikation. Den Klingelton des Handys in den beiden Werbespots zu verstehen bedeutet deshalb nicht, ein Zeichen zu entschlüsseln, das für etwas Bestimmtes steht – z. B. einen Telefonanruf –, sondern zu verstehen, was der Erzähler an dieser Stelle seiner Erzählung mit dem Klingelzeichen zu verstehen geben will. Zu wissen, dass das Klingeln eines Handys bedeutet, dass jemand anruft und den Handybesitzer sprechen möchte, ist zwar Voraussetzung für das Verständnis dieser Episode, aber nicht hinreichend. Man kann den Sinn des Handyklingelns in diesem Kontext des Spots nicht aus der Kenntnis dieser Zeichenbedeutung ableiten. Ableitbar ist der Sinn aus dem Kontext der Erzählung, der sich hier aus der bisherigen Geschichte und dem Zusammenspiel der auditiven und visuellen Modi in der entsprechenden Szene ergibt. Weder das Klingeln des Handys noch das, was im Bild gezeigt wird, ergibt für sich betrachtet im vorliegenden Kontext den gemeinten Sinn.

Eine Analyse der semiotischen Potenziale einzelner Modi im Rahmen einer Theorie des kommunikativen Handelns eröffnet die Möglichkeit, die Parallelität von räumlicher und zeitlicher Logik multimodaler Kommunikationsformen funktional aufzulösen. Kommunikative Handlungen können einerseits durch einen *Und-dann-Zusammenhang* verbunden sein: In einer Erzählung wird eine Szene akustisch eingeführt *und dann wird* in einem Bild die Szene gezeigt. Andererseits können Handlungen auch gleichzeitig ausgeführt werden: Im Bild wird eine Person eingeführt *und gleichzeitig* wird mittels des Tons gezeigt, was diese Person tut, und mit der Kameraführung *gleichzeitig* angezeigt, dass diese Person die Hauptperson ist. Multimodale Formen der Kommunikation – seien es Filme, wissenschaftliche Vorträge oder Online-Angebote – zeichnen sich gerade dadurch aus, dass *Und-dann-Zusammenhänge* und *Und-gleichzeitig-Zusammenhänge* miteinander kombiniert werden. Die Idee, dass der multimodale Sinn *multiplikatorisch* erzeugt wird, das Ganze also mehr ist als die Summe seiner Teile, kann durch einen weiteren Zusammenhangstyp operationalisiert werden, der in der Handlungstheorie gut eingeführt ist: den sogenannten *Indem-Zusammenhang*. Man kann eine komplexe Handlung – ›higher-level-actions‹ (NORRIS 2009: 81) – vollziehen, *indem* man andere, weniger komplexe Handlungen – ›lower-level-actions‹

(ebd.) – vollzieht. So kann man erzählen, wie einmal ein Mann sein Handy für sein Fernsehgerät zu Hause gehalten hat, indem man zeigt, dass er sich in der Hotel-Lobby wie zu Hause benimmt, indem man ihn den Kellner als seine Frau mit ›Schatz‹ ansprechen lässt.

Das, was in metaphorischer Weise als *Fusion*, *Verbindung* oder *Multiplikation* der modalen Elemente bezeichnet wird, besteht darin, dass die verwendeten Elemente aus gesprochener Sprache, Ton, Bild und Filmgestaltung in ein übergeordnetes Handlungsmuster – eine Erzählung, eine Handywerbung – eingebettet sind. Die Rekonstruktion dieses übergeordneten Handlungsmusters ist nur möglich, wenn alle modalen Elemente berücksichtigt werden. Die kommunikative Verbindung zwischen dem übergeordneten Handlungsmuster und den untergeordneten kann über die *Indem-Relation*, ihre räumlichen und zeitlichen Zusammenhänge können durch *Und-dann-* sowie *Und-gleichzeitig-Relationen* beschrieben werden.

Für das zweite Problem, das Rezeptionsproblem, ist der Befund aufschlussreich, dass multimodales Verstehen sowohl angebots- als auch rezipientengesteuert ist, d. h. als Zusammenspiel von Stimulusmerkmalen, von Intentionen und Motiven der Rezipienten, der Bedeutungen einzelner modaler Elemente und des übergeordneten Kommunikationszusammenhangs zu modellieren ist. Wie die Blickdaten und die Äußerungsdaten der Studie zeigen, kann dieses Zusammenspiel theoretisch befriedigend als interaktiver Aneignungsprozess beschrieben werden, mit einer räumlichen Dimension, auf der die Selektionsleistungen der Rezipienten angesiedelt sind, und einer zeitlichen Dimension, in der die Kohärenz gebildet wird (vgl. BUCHER 2007: 58-67). Multimodales Verstehen ist *reziprok*, insofern die einzelnen Elemente nicht isoliert, sondern im Zusammenhang anderer Elemente gedeutet werden; und es ist *rekursiv*, insofern die Deutungen permanent weiterbearbeitet und modifiziert werden, bis ein befriedigendes Verständnis erzielt ist. Diese beiden Strukturmerkmale des multimodalen Verstehens spiegeln sich in den Blickaufzeichnungsdaten: Einerseits wird dieselbe AOI im Verlauf des Aneignungsprozesses mehrfach fixiert, und andererseits findet sich in den Blickdaten das Muster einer wechselseitigen Fixation von zwei oder drei AOIs. Auch in den Äußerungsdaten der Probanden sind explizite Beschreibungen des schrittweisen Aufbaus eines Verständnisses enthalten. Man kann diesen Erschließungsprozess im Sinne einer unterstellten Als-ob-Interaktion auffassen: Mit jeder Deutung begegnet der Rezipient dem Angebot in anderer Weise, nimmt andere Anregungen aus dem Angebot auf und erweitert dadurch sein Verständnis. Inter-aktiv ist dieser Prozess insofern, als der Rezi-

pient mithilfe des medialen Angebotes schrittweise die typischen Probleme des multimodalen Verstehens zu lösen versucht: die Auswahl der relevanten bedeutungstragenden Elemente und die Rekonstruktion der zwischen ihnen bestehenden Handlungszusammenhänge.

Literatur

ARNHEIM, R.: *Anschauliches Sehen. Zur Einheit von Bild und Begriff* [engl. Orig. 1969]. Köln [DuMont] 2001

AUSTIN, J. L.: Ein Plädoyer für Entschuldigungen. In: Ders.: *Gesammelte philosophische Aufsätze*. Stuttgart [Reclam] 1986, S. 229–268

BAKHTIN, M. M.: *The Dialogic Imagination*. Austin [University of Texas Press] 1981

BAKHTIN, M. M.: *Speech Genres and Other Late Essays*. Austin [University of Texas Press] 1986

BALDRY, A.; P. J. THIBAULT: *Multimodal Transcription and Text Analysis. A Multimedia Toolkit and Coursebook with Associated Online-Course*. London, Oakville [Equinox] 2005

BATEMAN, J.: *Multimodality and Genre. A Foundation for the Systematic Analysis of Multimodal Documents*. London [Palgrave Macmillan] 2008

BENTE, G.: Erfassung und Analyse des Blickverhaltens. In: MANGOLD, CH.; P. VORDERER; G. BENTE (Hrsg.): *Lehrbuch der Medienpsychologie*. Göttingen, Bern, Toronto, Seattle [Hogrefe] 2004, S. 297–324

BILANDZIC, H.: *Synchrone Programmauswahl. Der Einfluss formaler und inhaltlicher Merkmale der Fernsehbotschaft auf die Fernsehnutzung*. München [Reinhard Fischer Verlag] 2004

BILANDZIC, H.: Lautes Denken. In: MIKOS, L.; C. WEGENER (Hrsg.): *Qualitative Medienforschung. Ein Handbuch*. Konstanz [UVK] 2005, S. 362–370

BONFADELLI, H.: *Medienwirkungsforschung I. Grundlagen und theoretische Perspektiven*. Konstanz [UVK] 2004

BROSIUS, H.-B.: Visualisierung von Fernsehnachrichten. Text-Bild-Beziehungen und ihre Bedeutung für die Informationsleistung. In: KAMPS, K.; M. MECKEL (Hrsg.): *Fernsehnachrichten. Prozesse, Strukturen, Funktionen*. Opladen, Wiesbaden [Westdeutscher Verlag] 1998, S. 213–224

BUCHER, H.-J.: Pressekritik und Informationspolitik. Zur Theorie und Praxis einer linguistischen Medienkritik. In: BUCHER, H.-J.; E. STRASSNER (Hrsg.): *Mediensprache – Medienkommunikation – Medienkritik*. Tübingen [Narr] 1991, S. 3–109

BUCHER, H.-J.: Online-Interaktivität – ein hybrider Begriff für eine hybride Kommunikationsform. Begriffliche Klärungen und empirische Rezeptionsbefunde. In: BIEBER, CH.; C. LEGGEWIE (Hrsg.): *Interaktivität. Ein transdisziplinärer Schlüsselbegriff*. Frankfurt/M. [Campus] 2004, S. 132–167

BUCHER, H.-J.: Textdesign und Multimodalität. Zur Semantik und Pragmatik medialer Gestaltungsformen. In: ROTH, K. S.; J. SPITZMÜLLER (Hrsg.): *Textdesign und Textwirkung in der massenmedialen Kommunikation*. Konstanz [UVK] 2007, S. 49–76

BUCHER, H.-J.: Vergleichende Rezeptionsforschung. Theorien – Methoden – Befunde. In: MELISCHEK, G.; J. SEETHALER; J. WILKE (Hrsg.): *Medien- und Kommunikationsforschung im Vergleich. Grundlagen, Gegenstandsbereiche, Verfahrensweisen*. Wiesbaden [vs Verlag für Sozialwissenschaften] 2008, S. 309–340

BUCHER, H.-J.: Multimodalität – eine Universalie des Medienwandels. Problemstellungen und Theorien der Multimodalitätsforschung. In: BUCHER, H.-J.; TH. GLONING; K. LEHNEN (Hrsg.): *Neue Medien – neue Formate. Ausdifferenzierung und Konvergenz in der Medienkommunikation*. Frankfurt/M. [Campus] 2010, S. 41–79

BUCHER, H.-J.: Multimodales Verstehen oder Rezeption als Interaktion. Theoretische und empirische Grundlagen einer systematischen Analyse der Multimodalität. In: DIEKMANNSHENKE, H.; M. KLEMM; H. STÖCKL (Hrsg.): *Bildlinguistik. Theorien – Methoden – Fallbeispiele*. Berlin [Erich Schmidt] 2011, S. 123–156

BUCHER, H.-J.; M. KRIEG; PH. NIEMANN: Die wissenschaftliche Präsentation als multimodale Kommunikationsform. Empirische Befunde zu Rezeption und Verständlichkeit von Powerpoint-Präsentationen. In: BUCHER, H.-J.; TH. GLONING; K. LEHNEN (Hrsg.): *Neue Medien – neue Formate. Ausdifferenzierung und Konvergenz in der Medienkommunikation*. Frankfurt/M. [Campus] 2010, S. 381–412

BUCHER, H.-J.; P. SCHUMACHER: The Relevance of Attention for Selecting News Content. An Eye-Tracking Study on Attention Patterns in the Reception of Print- and Online Media. In: *Communications. The European Journal of Communications Research*, 31/3, 2006, S. 347–368

BUCHER, H.-J.; P. SCHUMACHER (Hrsg.): *Interaktionale Rezeptionsforschung. Theorie und Methode der Blickaufzeichnung in der Medienforschung*. Wiesbaden [vs Verlag für Sozialwissenschaften] 2011

DUCHOWSKI, A. T.: *Eye Tracking Methodology. Theory and Practice* [2. Aufl.]. London [Springer] 2007

FRITZ, G.: *Kohärenz. Grundfragen der linguistischen Kommunikationsanalyse*. Tübingen [Narr] 1982

GIBSON, J. J.: *Die Sinne und der Prozess der Wahrnehmung*. Bern u. a. [Huber] 1973

GOMPEL, R. P. G. VAN; M. H. FISCHER; W. S. MURRAY; R. L. HILL (Hrsg.): *Eye Movements. A Window on Mind and Brain*. Amsterdam, Boston, Heidelberg u. a. [Elsevier] 2007

HAYHOE, M. M.; J. DROLL; N. MENNIE: Learning Where to Look. In: GOMPEL, R. P. G. VAN; M. H. FISCHER; W. S. MURRAY; R. L. HILL (Hrsg.): *Eye Movements. A Window on Mind and Brain*. Amsterdam, Boston, Heidelberg u. a. [Elsevier] 2007, S. 641–659

HENDERSON, J. M.; F. FERREIRA: Scene Perception for Psycholinguists. In: HENDERSON, J. M.; F. FERREIRA (Hrsg.): *The Interface of Language, Vision, and Action. Eye Movements and the Visual World*. New York [Psychology Press] 2004, S. 1–58

HENDERSON, J. M.; J. R. BROCKMOLE; M. S. CASTELHANO; M. MACK: Visual Saliency Does Not Account for Eye Movements During Visual Search in Real-World Scenes. In: GOMPEL, R. P. G. VAN; M. H. FISCHER; W. S. MURRAY; R. L. HILL (Hrsg.): *Eye Movements. A Window on Mind and Brain*. Amsterdam, Boston, Heidelberg u. a. [Elsevier] 2007, S. 537–562

HICKETHIER, K.: *Film- und Fernsehanalyse*. Stuttgart, Weimar [J. B. Metzler] 2007

HOLSANOVA, J.: *Discourse, Vision, and Cognition*. Amsterdam; Philadelphia [Benjamins] 2008

HYÖNÄ, J.; R. RADACH; H. DEUBEL (Hrsg.): *The Mind's Eye. Cognitive and Applied Aspects of Eye Movement Research*. Amsterdam, Boston, Heidelberg u. a. [Elsevier] 2003

IEDEMA, R.: Multimodality, Resemiotization. Extending the Analysis of Discourse as Multi-Semiotic Practice. In: *Visual Communication*, 2/1, 2003, S. 29–57

ISER, W.: Interaction between Text and Reader. In: SULEIMAN, S. R.; I. CROSMAN (Hrsg.): *The Reader in the Text. Essays on Audience and Interpretation*. Princeton (NJ) [Princeton University Press] 1980, S. 106–119

ISER, W.: *Der Akt des Lesens* [4. Aufl.]. München [Fink] 1994

ITTI, L.; CH. KOCH; E. NIEBUR: A Model of Saliency-Based Visual Attention for Rapid Scene Analysis. In: *IEEE Trans. Patt. Anal. Mach. Intell.*, 20/11, 1998, S. 1254–1259

JÄCKEL, M.: *Medienwirkungen. Ein Studienbuch zur Einführung*. Wiesbaden [VS Verlag für Sozialwissenschaften] 2005

JENSEN, K. B. (Hrsg.): *A Handbook of Media and Communication Research. Qualitative and Quantitative Methodologies*. London, New York [Routledge] 2002

JEWITT, C.: Different Approaches to Multimodality. In: JEWITT, C. (Hrsg.): *The Routledge Handbook of Multimodal Analysis*. London, New York [Routledge] 2009, S. 28–39

KIOUSIS, S.: Interactivity. A Concept Explication. In: *New Media & Society,* 4/3, 2002, S. 355–383

KRESS, G.: The Multimodal Landscape of Communication. In: *Medien-Journal*, 26/4, 2002, S. 4–18

KRESS, G.: What is Mode? In: JEWITT, C. (Hrsg.): *The Routledge Handbook of Multimodal Analysis*. London, New York [Routledge] 2009, S. 54–67

KRESS, G.; T. VAN LEEUWEN: *Reading Images. The Grammar of Visual Design*. New York [Routledge] 1996

KRESS, G.; T. VAN LEEUWEN: The Critical Analysis of Newspaper Layout. In: BELL, A.; P. GARRETT (Hrsg.): *Approaches to Media Discourse*. Oxford [Blackwell] 1998, S. 186–219

KRESS, G.; T. VAN LEEUWEN: *Multimodal Discourse. The Modes and Media of Contemporary Communication*. London [Arnold] 2001

LANG, A.: The Limited Capacity Model of Mediated Message Processing. In: *Journal of Communication*, 50/1, 2000, S. 46–70

LEMKE, J. L.: Multiplying Meaning. Visual and Verbal Semiotics in Scientific Text. In: MARTIN, J. R.; R. VEEL (Hrsg.): *Reading Science. Critical and Functional Perspectives on Discourses of Science*. London [Routledge] 1998, S. 87–113

LIM, F. V.: Developing an Integrative Multi-Semiotic Model. In: O'HALLORAN, K. L. (Hrsg.): *Multimodal Discourse Analysis. Systemic Functional Perspectives*. London, New York [Continuum] 2004, S. 220–246

LIM, F. V.: The Visual Semantics Stratum. Making Meaning in Sequential Images. In: ROYCE, T. D.; W. L. BOWCHER (Hrsg.): *New Directions in the Analysis of Multimodal Discourse*. New Jersey, London [Lawrence Erlbaum] 2007, S. 195–214

LUHMANN, N.: *Soziale Systeme. Grundriß einer allgemeinen Theorie*. Frankfurt/M. [Suhrkamp] 1984

MATTHIESSEN, CH. M. I. M.: The Multimodal Page. A Systematic Functional Exploration. In: ROYCE, T. D.; W. L. BOWCHER (Hrsg.): *New Directions in the Analysis of Multimodal Discourse*. New Jersey, London [Lawrence Erlbaum] 2007, S. 1–62

MAYER, P. A.: Computer-Mediated Interactivity. A Social Semiotic Perspective. In: *Convergence*, 4/3, 1998, S. 40–58

MCMILLAN, S. J.: A Four-Part Model of Cyber-Interactivity. Some Cyber-Places are More Interactive than Others. In: *New Media & Society*, 4/2, 2002, S. 271–291

MIKOS, L.: *Film- und Fernsehanalyse*. Konstanz [UVK] 2003

NEUMANN, O.: Theorien der Aufmerksamkeit. Von Metaphern zu Mechanismen. In: *Psychologische Rundschau*, 43, 1992, S. 83–101

NEUMANN, O.: Theorien der Aufmerksamkeit. In: NEUMANN, O.; A. F. SANDERS (Hrsg.): *Enzyklopädie der Psychologie. Themenbereich C: Theorie und Forschung. Serie II: Kognition. Band 2: Aufmerksamkeit*. Göttingen, Bern, Toronto; Seattle [Hogrefe] 1996, S. 559–643

NORRIS, S.: Modal Density and Modal Configurations. In: JEWITT, C. (Hrsg.): *The Routledge Handbook of Multimodal Analysis*. London, New York [Routledge] 2009, S. 78–90

O'HALLORAN, K. L.: Interdependence, Interaction and Metaphor in Multisemiotic Texts. In: *Social Semiotics*, 9/3, 1999, S. 317–354

O'HALLORAN, K. L.: Systemic Functional-Multimodal Discourse Analysis (SF-MDA). Constructing Ideational Meaning Using Language and Visual Imagery. In: *Visual Communication*, 7/4, 2008, S. 443–475

RICHARDSON, D. C.; M. J. SPIVEY: Eye Tracking. Characteristics and Methods. In: WNEK, G. E.; G. L. BOWLIN (Hrsg.): *Encyclopedia of Biomaterials and Biomedical Engineering*. New York, Oxford [Marcel Dekker] 2004, S. 568–572

SCHNEIDER, W. L.: Die Komplementarität von Sprechakttheorie und systemtheoretischer Kommunikationstheorie. Ein hermeneutischer Beitrag zur Methodologie von Theorievergleichen. In: *Zeitschrift für Soziologie*, 25/4, 1996, S. 263–277

SCHWEIGER, W.: *Theorien der Mediennutzung. Eine Einführung*. Wiesbaden [VS Verlag für Sozialwissenschaften] 2007

TASHAKKORI, A.; CH. TEDDLIE: *Mixed Methodology. Combining Qualitative and Quantitative Approaches*. Thousand Oaks [SAGE] 1998

THOMPSON, J. B.: *The Media and Modernity. A Social Theory of the Media*. Cambridge [Polity Press] 1995

VAN LEEUWEN, T.: Multimodality, Genre and Design. In: NORRIS, S.; R. JONES (Hrsg.): *Discourse in Action. Introducing Mediated Discourse Analysis*. London [Routledge] 2005, S. 73–94

YARBUS, A. L.: *Eye Movements and Vision*. New York [Plenum Press] 1967

ZETTL, H.: *Sight, Sound, Motion. Applied Media Aesthetics*. Belmont u. a. [Wadsworth Publishing Company] 1999