

intermedial. Stimme und Schrift, Bild und Ton. – Berlin/New York: de Gruyter, 2010. S. 447-467. (Jahrbuch des Instituts für Deutsche Sprache 2009), <https://doi.org/10.1515/9783110223613.447>

Angela Keppler

Die wechselseitige Modifikation von Bildern und Texten in Fernsehen und Film

Abstract

Es scheint eine Selbstverständlichkeit zu sein, dass Bild und Text sowie Bild und Klang einander in den audiovisuellen Prozessen filmischer Produkte wechselseitig konturieren. Dieser geläufige Befund jedoch bedarf weiterer Klärung. Vor allem die wissenschaftliche Analyse von Sendungen des Fernsehens operiert nicht selten weiterhin mit irreführenden Hierarchien zumal zwischen Text und Bild, so als würde der Bildverlauf eine weitgehend autonome Bedeutung des im sprachlichen Text Gesagten lediglich modifizieren oder transformieren. Demgegenüber vertritt der Beitrag die These, dass Bild und Text in filmischen Erzeugnissen jeglicher Art *gleichursprünglich* an der Konstitution der in ihnen präsentierten Gehalte beteiligt sind. Es wird gezeigt, dass dieser basale audiovisuelle Zusammenhang auch dort besteht, wo eines dieser Elemente – etwa der mündliche Dialog oder ein verbaler Kommentar – den Verlauf eines Films oder einer Filmsequenz zu dominieren scheint, und folglich, dass er alle filmischen Genres einschließlich der Sendetypen des Fernsehens gleichermaßen prägt. Eine Interpretation ausgewählter Beispiele stützt das Plädoyer für ein nichthierarchisches Verständnis der Bedeutungsdimensionen in den Produktionen von Film und Fernsehen.

Innerhalb der Film- und Fernsehwissenschaft ist es längst zu einer Selbstverständlichkeit geworden, dass Bild und Text sowie Bild und Ton einander in den audiovisuellen Prozessen filmischer Produkte wechselseitig konturieren. Diesem Befund möchte ich nachgehen und ihm dabei einige Aspekte abgewinnen, die vielleicht noch nicht hinreichend selbstverständlich sind. Ich werde die These vertreten, dass Bild und Ton in filmischen Erzeugnissen jeglicher Art *gleichursprünglich* an der Konstitution der in ihnen präsentierten Gehalte beteiligt sind – und zwar gerade auch dort, wo eines dieser Elemente – etwa der mündliche Dialog oder ein verbaler Kommentar – in bestimmten Formaten oder Sequenzen sowohl des Spielfilms als auch des Fernsehens zu dominieren scheint.

Filmische Produkte stellen eine Einheit von akustischer und visueller Bewegung dar. Bei dem Verhältnis von visueller und akustischer Dimension handelt es sich hier keineswegs um ein additives, sondern durchweg um ein integrales Verhältnis: der Klang fügt dem Bild nicht nur etwas hinzu, der Klang *verwandelt* das Bildgeschehen, das seinerseits das modifiziert, was akustisch vernehmbar ist. Gesprochene Sprache, Musik und andere Geräusche bilden eine integrale Dimension des filmischen Bewegungsbildes. Wie Schnitt, Montage, Kameraführung, Lichtregie, Farbgebung und dergleichen sorgen

Sprache und Musik für den spezifischen Rhythmus eines ganzen Films. Zugleich sind sie wesentliche Medien der in Filmen verkörperten Gehalte: sie transportieren Stimmungen, vermitteln Informationen, geben Rätsel auf oder führen in die Irre, beruhigen oder beunruhigen, wiegeln auf oder wiegeln ab, stellen klar oder lassen im Unklaren. Zusammen bilden Bild und Klang die in einem filmischen Verlauf geschaffenen und von ihm gezeigten Situationen. Die Anerkennung und Interpretation dieses Zusammenhangs von visueller und akustischer Bewegung ist daher die erste Aufgabe einer Wissenschaft des filmischen Bildes, gleich ob sie sich seinen Realisierungen im Kino, im Fernsehen oder in anderen Kontexten der Produktion und Rezeption widmet.¹

Mein Beitrag hat drei Teile. Ich werde zunächst mit einigen grundsätzlichen Bemerkungen etwas genauer an die Komplexität der Verhältnisse zwischen Bild und Text sowie Bild und Ton in filmischen Produkten erinnern. Durch eine Interpretation von vier kurzen Beispielen aus Kino und Fernsehen werde ich anschließend für ein strikt nichthierarchisches Verständnis der Bedeutungsdimensionen in den Produktionen beider Medien argumentieren. Zum Schluss werde ich andeuten, wie sich auf dieser Basis die mediale Differenz zwischen Kino und Fernsehen begreifen lässt.

I.

Als bildliches Medium gehört der Film – nach der Terminologie von Nelson Goodman (1995, Kap. IV) – zu den syntaktisch und semantisch *dichten* Zeichen, die im Unterschied zur gesprochenen und geschriebenen Sprache keine eindeutige Trennung der bedeutungsbildenden Komponenten und der Bezugsgegenstände dieser Zeichen erlaubt. Aus diesem Grund ist alle Aussicht auf eine wie immer geartete *Grammatik* der filmischen Rede vergebens. Nichtsdestotrotz ist das filmische Bild durch Einstellung, Schnitt, Montage, Kameraführung, Farb- und Lichtregie ein vielfach artikuliertes Bild, und es ist für die Filmanalyse entscheidend zu verfolgen, was in diesen Prozessen geschieht: wie die Objekte, Szenen, Personen, Verläufe durch die filmische Aufnahme dargeboten werden, welche Zeit zu ihrer Wahrnehmung bleibt, welche Ansichten jeweils präsentiert und montiert werden usw. Trotz vieler Typisierungen in filmischen Verfahren finden sich hier aber keine *abgegrenzten* Elemente wie Buchstaben, Worte oder Sätze der Sprache, da sich das Bild des Films – als Spezies des *Bildes* – nicht aus distinkten Zeichen und Gruppen von Zeichen zusammensetzt, deren Anordnung einen buchstäblichen *Text* der Kommunikation ergeben würde. Es bietet sich vielmehr in der Fülle und Bewegung seiner Elemente dar, an der

¹ Zu den bildtheoretischen Voraussetzungen und methodischen Konsequenzen dieses Verständnisses vgl. Keppler (2005).

geübte Zuschauer signifikante Dinge und Ereignisse, Figuren und Handlungen, Zustände und Verläufe erkennen. Außerdem wäre mit einer Beschränkung auf diese Aspekte wiederum nur von der visuellen Sprache des Films die Rede, nicht aber von der tatsächlichen Artikuliertheit der Filme, die wir in Kino und Fernsehen zu sehen bekommen. Ebenso unangemessen wäre es freilich, Filme als eine Kombination mehrerer „Sprachen“ – z.B. der bildlichen, der musikalischen und der verbalen – zu verstehen. Was wir jeweils sehen und hören, ist eine audiovisuelle Komposition, die es gerade in der Integration ihrer Elemente zu verstehen gilt. Daher ist die Rede von einer *Sprache* des Films allein in metaphorischer Bedeutung sinnvoll. Metaphorisch verstanden, lenkt sie die Aufmerksamkeit auf die *gesamte Dramaturgie* von Filmen und hebt die Interdependenz ihrer Formelemente hervor. Wie diese Integration aber auch jeweils geleistet wird, ob synchron oder asynchron, harmonisch oder dissonant, kontrastreich oder kontrastarm, alles dies sind *Fülle der Einheit* von Klang und Bild im filmischen Prozess. Was durch Filme gleich welcher Art kommuniziert wird, ergibt sich stets aus dieser ihrer gesamten Dramaturgie: Keine der beteiligten Dimensionen, weder das bildliche, noch das musikalische, noch das sprachliche oder das übrige akustische Geschehen dürfen bei der Interpretation filmischer Produkte vernachlässigt werden, so sehr eine analytische Trennung häufig sinnvoll ist. Denn keine ist allein für den Gehalt von Filmen verantwortlich.²

Unter der Überschrift eines Zusammenhangs von *Text und Bild* kann deshalb vieles angesprochen sein. In einem eher semiotischen Kontext werden manchmal auch Bilder und Bildverläufe als *Texte* klassifiziert, womit es Verhältnisse zwischen *Arten* von Texten wären, die hier angesprochen sind. Wenn dagegen von *Bild und Text* die Rede ist, kann einerseits das Verhältnis von Bild und *Schrift* und andererseits, jedenfalls sobald man auch filmische Bilder im Auge hat, dasjenige von Bild und *verbaler Rede* gemeint sein; drittens schließlich kann die Wendung als ein Stellvertreter für den allgemeinen Zusammenhang von Bild und *Ton* aufgefasst werden. Welche Terminologie man aber auch im Einzelnen bevorzugen mag: Sobald wir über filmische Produkte sprechen, ist mit einer potentiellen Interaktion aller dieser Elemente zu rechnen. Denn da sie in vielen konventionalisierten Formen – wenn auch, wie gesagt, auf sehr unterschiedliche Weise – oft alle versammelt *sind*, trägt von Fall zu Fall gerade die *Abwesenheit* zum Beispiel der

² Obwohl dies – wie gesagt – in der Forschung über Film und Fernsehen ein weithin akzeptierter Grundsatz ist, vermögen nicht alle methodischen Vorgehensweisen ihm tatsächlich zu entsprechen. Ein Beispiel wäre die in Deutschland durch Gerd Albrecht (1964) bekannt gemachte quantitative Inhaltsanalyse; ein aktuelles Beispiel wäre Grimm (2006). Auch die ethnomethodologische Konversationsanalyse z.B. von Nachrichteninterviews im Fernsehen schenkt der visuellen Präsentation allzu wenig Beachtung (vgl. Heritage/Clayman/Zimmermann 1988); Gleiches gilt für Untersuchungen aus dem Umkreis der sozialwissenschaftlichen Hermeneutik, z.B. Reichertz (2000). Zur Kritik an diesen Ansätzen siehe Keppler (2006, S. 96–104).

Musik, der Stimme, anderen Arten des Tons und sogar des Bildes zur spezifischen Dramaturgie filmischer Verläufe bei.³ Wir müssen also, um es ein wenig paradox zu sagen, im Blick auf Film wie Fernsehen manchmal selbst dort von Stimme *und* Schrift, Bild *und* Ton sprechen, wo diese gar nicht alle zusammen kommen.

In diesen allgemeinen Bemerkungen liegt bereits eine deutliche Warnung vor einer im Ansatz hierarchischen Auffassung der für filmische Verläufe konstitutiven medialen Dimensionen.⁴ Dies gilt meines Erachtens auch für die visuellen Relationen von Bild und Schrift, insbesondere aber für die audiovisuellen von Bild (einschließlich der Schrift) und Ton. Da ich anhand meiner Beispiele vorwiegend die Letzteren kommentieren werde, sei ein weiterer Hinweis vorausgeschickt. Zu den vergleichsweise wenigen Filmtheoretikern, die die auditive Dimension von Filmen wirklich ernst genommen haben, zählt Gilles Deleuze. Das 9. Kapitel des zweiten Bandes seiner Filmtheorie, das mit „Die Bestandteile des Bildes“ überschrieben ist, handelt deshalb ausführlich von der Rolle des Tons im filmischen Geschehen. Unter anderem heißt es dort:

Es ist oftmals erwähnt worden, daß es nur eine Tonspur, aber zumindest drei Gruppen von *Tönen* gibt, nämlich Reden, Geräusche und Musik. Möglicherweise müssen wir aber eine größere Gruppe von akustischen Elementen unterscheiden: Geräusche (die einen Gegenstand isolieren und sich voneinander abheben); Töne (die Bezüge markieren und untereinander in gegenseitigem Bezug stehen); Lautbildungen (die Bezüge zerlegen und bei denen es sich um Schreie, aber genauso gut um regelrechten *Jargon* handeln kann, wie in den Slapstik-Tonfilmen von Chaplin oder Jerry Lewis); schließlich Reden und Musik. (Deleuze 1991, S. 300)⁵

So sehr man diese Liste noch verlängern könnte, so sehr mag sich hier zugleich die Frage stellen, ob nicht diese Komplexitäten der Klangregie vor allem auf den Kinofilm zutreffen (dem Deleuzes Überlegungen ausschließlich gewidmet sind), aber gerade nicht auf die spezifische Ästhetik des Fernsehens, die doch alles in allem – bildlich wie klanglich und erst recht klangbildlich – mit schlichteren Mitteln operiert.

³ In Alejandro González Iñárritus Beitrag zu dem Episodenfilm *11'09''01* (Frankreich 2002) z.B. bleibt das Bild über weite Strecken schwarz; ähnliche Effekte gibt es in Abbas Kiarostamis *Five* (Iran 2003).

⁴ Als „hierarchisch“ bezeichne ich dabei die Auffassung, in bestimmten Formaten zumal des Fernsehens stelle das Wort gegenüber dem Bild den primären Bedeutungsträger der filmischen Kommunikation dar; denkbar ist aber auch der umgekehrte Fall. Ein klassisches Beispiel einer solchen Vorrangsthese stellt Hegels in seinen *Vorlesungen über die Ästhetik* vertretene Annahme dar, in der Poesie hätten die durch das Wort evozierten Vorstellungen ein grundsätzlich größeres Gewicht als der Klang. In der Kommunikationswissenschaft finden sich solche Hierarchisierungen in hohem Maß bei der Anwendung quantitativer wie qualitativer Methoden der Inhaltsanalyse; vgl. Merten (1995), Bonfadelli (2002) sowie kritisch Keppler (2006, S. 98 ff.).

⁵ Vgl. ebd.: „Die Stimme ist nicht von den Geräuschen zu trennen, die sie oftmals unvernnehmbar macht.“

Eine solche Annahme erschiene mir jedoch höchst irreführend. Im Kino-film *und* im Fernsehen, so möchte ich darlegen, haben wir es durchweg mit einer dichten *Interaktion* zwischen visuellem und akustischem Geschehen zu tun. Hier wie dort werden der Raum und die Zeit des Bildgeschehens durch den Einsatz eines reichen Spektrums akustischer Mittel gebildet; hier wie dort greift die bildliche Dramaturgie permanent in die Gestaltung des auf der akustischen Ebene Vermittelten ein. Der kommunikative Gehalt, sei es einzelner Sequenzen, sei es von Sendebeiträgen und Sendungen sowie dokumentarischen oder fiktionalen Filmen im Ganzen, ergibt sich stets aus der Verzahnung von visuellen und akustischen Strukturen. Deshalb spreche ich von einer grundsätzlichen und konstitutiven, für das filmische Medium in allen seinen Spielarten charakteristischen *Modifikation* des Bildes durch den Ton und des Tons durch das Bild. Dass dies so ist und wie es sich vollzieht, möchte ich nun an jeweils zwei Beispielen aus dem Bereich des Spielfilms und des Fernsehens verdeutlichen. Der Vergleich der kommentierten Filmausschnitte soll einerseits die Spannweite der Kombination von Bild und Ton in beiden Medien exemplarisch verdeutlichen. Zum andern kann die Interpretation des zweiten, dritten und vierten Beispiels demonstrieren, wie wenig der verbale Diskurs allein für die Realisierung filmischer Darbietungen verantwortlich ist.⁶

II.

In meinem ersten Beispiel werden keinerlei Worte gewechselt, dennoch spielen Bild, Schrift und Klang auf eine intensive Weise ineinander. Es handelt sich um eine Sequenz aus Michelangelo Antonionis Film *Zabriskie Point* aus dem Jahr 1970. Zwei Studenten befinden sich auf einer Fahrt durch Los Angeles zu dem von Protesten und Polizeieinsatz geprägten Campus ihrer Universität. In den knapp zehn Minuten davor war eine längere Passage mit revolutionärem und scheinrevolutionärem Gerede unter Studenten zu sehen. In ihr hat sich der Held des Films zu der Bereitschaft bekannt, in der Revolte sein Leben aufs Spiel zu setzen, was er im Verlauf der Handlung, allerdings in einer eher existentiellen Manier – und mit am Ende tödlichem Ausgang – auch tun wird. Danach werden kurz die beiden anderen Hauptfiguren eingeführt, ein kapitalistischer Bauunternehmer und die Studentin, die der Held auf seiner Flucht ins Death Valley treffen wird. Es folgt ein abrupter Übergang zu einer Sequenz, in der der Held und sein Zimmergenosse durch Los Angeles zum Campus fahren, deren erste 150 Sekunden eine hochimaginative bildliche und klangliche Komposition entwerfen.

⁶ Die Interpretation der letzteren wird durch Transkripte unterstützt, nicht jedoch die des ersten Ausschnitts, da selbst ein detailliertes Protokoll dieser hochkomplexen Sequenz nicht gerecht werden kann. Zum Verfahren der Transkription und dessen Bedeutung für die Analyse filmischer Produkte vgl. Keppler (2006, Kap. 3: Eine Methode der Fernsehanalyse).

Aber auch sprachlicher Text ist in dieser Passage massiv präsent. Er tritt dem Betrachter auf zahlreichen Reklametafeln und -aufschriften entgegen, die zusammen mit den Bildern im Stadtbild sowie den vorbeiziehenden Gewerbeanlagen und Fahrzeugen für den Eindruck einer surreal-chaotischen Zivilisationswüste sorgen. Für eine Fülle von Signalen sorgt hier allein schon die Schrift. Eine entscheidende Rolle aber spielt zudem die akustische Regie. Zunächst wird die Fahrt von ganz *realistischen* Geräuschen begleitet, bis nach 23 Sekunden, direkt nach dem Geräusch der quietschenden Reifen des um eine Kurve biegenden Pick-Ups schrille und harte elektronische Musikklänge einsetzen, die sich in der Folge nahezu ununterscheidbar mit dem Verkehrslärm der Stadt mischen (die Originalmusik zu diesem Film stammt von der Gruppe *Pink Floyd*); zu hören sind Motorgeräusche, das Rattern eines Zuges, Huplaute und Sirenenklänge. Mit diesem Klangdschungel korrespondiert ein vorübergehendes Intransparentwerden des Bildverlaufs, als ein vorbeifahrender Güterzug so ins Bild gesetzt wird, dass sich alle fixierbaren Gestalten in einer Art filmischem Action-Painting vorübergehend auflösen, wiederum scharf kontrastiert von der Palmenallee, die in Richtung des Campus führt. Von dem Einsetzen der Musik bis zum Schnitt in die Allee vergeht gerade mal eine gute Minute. Sie führt in und durch einen verstörenden, ebenso zersplitterten wie enerzierenden Geschehensraum, in dem jene strukturelle Gewalt schon am Werk ist, von der die Erzählung handeln wird. Dieser filmische Bewegungsraum ist gleichermaßen durch Bild und Klang synthetisch und synergetisch hergestellt. Er bildet nach dem aufgeputschten Durcheinanderreden der Studenten das weit kürzere zweite Präludium des Films, das den erzählten Ereignissen wortlos vorausgreift. Zusammen erzeugen Bild und Klang eine Atmosphäre gesellschaftlicher Hochspannung, die sich am Schluss des Films in einer von der weiblichen Heldin imaginierten, über insgesamt fünf Minuten hinweg zunächst in Originalgeschwindigkeit und realistischer Akustik, nach anderthalb Minuten dann – von einer psychedelischen Rockmusik unterlegt – in Superzeitlupe mehrfach wiederholten Explosion einer Hotelanlage entlädt.

Mein zweites Beispiel ist eine Szene aus Martin Scorseses Film *Goodfellas* aus dem Jahr 1990. Nach einer guten Viertelstunde befinden wir uns am Ende der Vorgeschichte der Erzählung, in der die Betrachter ausführlich in das Milieu dieses Films eingeführt worden sind. Die Sequenz wird eingeleitet durch einen in Aufsicht vollzogenen Schwenk über einen Tisch in einem italienischen, von der Mafia frequentierten Restaurant. Der Zuschauer ist durch eine vorangegangene Kamerafahrt bereits in die halbseidene Atmosphäre dieses Lokals eingeführt worden. Dazu war, wie auch am Anfang der Passage, die ich kommentieren möchte, die Stimme der von Ray Liotta gespielten Hauptfigur Henry Hill aus dem Off zu hören, die über den ganzen Film hinweg immer wieder die Ereignisse aus einer sentimental

auf seine Mafia-Zeit rückblickenden Perspektive kommentiert. Hier wird *mit Worten* der Entwicklung der Handlung vorausgegriffen, freilich ohne etwas über ihren Verlauf zu verraten. Darin liegt eine Distanzierung und zugleich Verklärung, die aber – wiederum über den ganzen Film hinweg – immer wieder aufgehoben wird, um von den Stationen einer am Ende desaströsen Mafia-Karriere von innen heraus zu erzählen. So auch in diesem Ausschnitt. Nachdem die Stimme des retrospektiven Erzählers verklungen ist, erhebt sich zunächst der junge Held und kurz darauf sein älterer Kumpan Jimmy Conway von der Tafel, um sich mit einem neu hinzugekommenen, etwas vierschrötigen Gast an der Bar zu treffen. Daraufhin kommt es für eine knappe Minute zu einem kurzen konspirativen Dialog, der sich akustisch und bildlich in einem Binnenraum des zuvor etablierten filmischen Raums abspielt.

Der vor diesem Dreiergespräch zu hörenden Stimme aus dem Off kommt an dieser Stelle vor allem die dramaturgische Funktion zu, in Ergänzung und durchaus auch Abhebung von den sichtbaren Szenen das Selbstbild der Akteure zu konturieren. Sie kommentiert keinen bestimmten Vorgang, der sich auf der Leinwand abspielt, sondern gibt vielmehr das Lebensgefühl dieser verschworenen Gemeinschaft (und vor allem des Helden) wieder, getragen vom Halblicht des von Rot- und Schwarzönen geprägten Ambientes. Die anscheinend im Restaurant erklingende Musik und die vernehmbaren Gesprächsfetzen halten die Szene für ein erneutes Eintauchen der Handlung in ihren imaginativen Raum bereit. Dass die erklingende Musik nicht eindeutig innerhalb der Szene verortet ist, wird an ihrem plötzlichen Anschwellen deutlich, als der von Robert de Niro gespielte Jimmy Conway aufsteht und sich zu der abseits in Gang gekommenen Besprechung zwischen Henry Hill und dem Neuankömmling gesellt, bei der es um die Planung eines größeren Gelddiebstahls geht – eine akustische Zäsur, die einen kurzen Zwischenakt einleitet, bevor es an der größeren Tafelrunde zu lebhaften Auseinandersetzungen kommt.

GoodFellas (R: Martin Scorsese, USA 1990), Ausschnitt

Nr. Zeit	Bild	Ton
01 '08	A, S ^{lo} : Jimmy Conway (JC) geht durch ein Restaurant zur Bar, N, S ⁱ : JC reicht lächelnd Robert McMahon (MM) am BR ^{re} die Hand, S ^{lo} : JC wendet sich Henry Hill (HH) zu, HH blickt kurz nach li, dann wieder zu JC, kurze Zeigegeste mit der Hand am BR ^u	Mu: ((Streichermusik)) (JC): hey, () (MM): () (good to see you) (): ((lachen)) (): () (yeah) Mu: ((Streicher werden leiser)) G: ((HG: Gläserklirren)) HH: <<p> te'=tell= him what=you>

Nr. Zeit	Bild	Ton
02 '02	N: BR ^{li} HH von re hinten, BH ^{li} JC im Profil von re, BH ^{re} MM, JC stößt am BR ^u mit einem Schnapsglas mit MM an	Mu: ((HG: Streichermusik)) HH: <<p> were telling me (my friend)> G: ((Gläserklirren)) MM: too good=to be true;
03 '02	N: BH ^{re} JC im Profil von li, BH ^{li} HH nach li gelehnt, blickt JC an, JC stößt mit seinem Glas am BR ^u mit HH an, HH grinst, JC blickt kurz nach re, setzt sein Glas an und legt den Kopf in den Nacken, HH nippt an seinem Glas und schaut dabei JC an	Mu: ((HG: Streichermusik)) HH: haha: G: ((Gläserklirren))
04 '05	N: BM JC im Profil von re, setzt sein Glas ab, blickt kurz zu MM, dann nach u, BH ^{re} MM vornübergebeugt, den Kopf zu JC gewandt, richtet sich auf, breitet seine Arme aus und lehnt sich wieder nach vorn, JC trinkt einen Schluck aus einem Glas, BR ^{li} HH vornübergebeugt, setzt das Glas ab, senkt seinen Oberkörper Richtung Tisch und blickt JC und MM an	Mu: ((HG: Streichermusik)) MM: <<p> big score coming from Air France; (.) (anyway) G: ((Klopfen)) MM: Ba:gs of money like this comin=in; (-) 'kay,
05 '06	N: BH ^{li} HH kauend, blickt nach re o zu JC, BR ^{re} MM von hinten li mit starken Kopfgesten, JC starrt geradeaus, zieht an Zigarette, beugt sich etwas herunter, blickt kurz zu MM und nickt leicht	Mu: ((HG: Streichermusik)) MM: <<p> (from touris') an American (serviceman will) change their money (according to) French money send=them back? here>
06 '09	N: BH ^{li} HH von hinten re, die Hand nahe am Kinn, BM JC im Profil, BH ^{re} MM blickt nach re u, alle vornübergebeugt, HH und JC MM zugewandt, MM grinst, JC nickt, HH richtet sich kurz auf, deutet mit Salzgebäck in seiner Hand auf JC, JC dreht den Kopf nach li	Mu: ((HG: Streichermusik, weibliche Gesangsstimme)) JC: <<p> sch(hh) calm=down;> MM: <<pp> o: kay (-) it's beautiful (buddy) (> JC: hmm MM: it's totally (.) totally untraceable (-) okay, HH: <<p> (the only) problem>
07 '05	N: BH ^{li} HH, vornübergebeugt, den Kopf schräg haltend, schaut JC an, seine li Hand am BR ^{u,li} , JC in BM im Profil von li und MM am BR ^{re} von hinten li blicken HH an, HH mit Handgesten, schüttelt leicht den Kopf, nickt schwach, JC wendet den Kopf zu MM	Mu: ((HG: Streichermusik, weibliche Gesangsstimme)) HH: <<p> is gettin=a key (-) but i got somethin' (all work'd out)> (MM): yea=yea=yea ((lacht)) HH: [<<p> me and Frenchy> (MM): [yea=yea ((lacht)) HH: <<p> (this fuckin') guy (he's a) [citizen (so what soever)> MM: [((lacht)) (yeah I guess nice) piece=o'=work (-) now?

Nr. Zeit	Bild	Ton
08 '06	N: BH ⁱⁱ Hinterkopf und re Wange von HH, die Hand am Mund, BM JC im Profil von re, BHre MM den beiden zugewandt, alle stark vornübergebeugt, HH und JC blicken MM an, MM zieht die Augenbrauen hoch, blinzelt und nickt, JC nickt, den Kopf leicht geneigt, hebt kurz die Hand, wendet sich zu HH	Mu: ((HG: Streichermusik, weibliche Gesangsstimme)) MM: <<p> if I'm right (-) there could be like häelf=a mil comin' (-)> (): hm=hm MM: <<p> o:hl cash> (): hm=hm HH: <<p> an=he said>
09 '03	N: BH ⁱⁱ HH, BM JC im Profil von li; BE ^{re,o} MM von li hinten, alle vorgebeugt, HH zu JC und MM gebeugt, JC wendet den Kopf zu HH, dann zu MM; VG aufsteigender Zigarettenrauch	Mu: ((HG: Streichermusik, weibliche Gesangsstimme)) HH: <<p> the best time is probably over the weekend (-) so maybe sadurday=> MM: [oh yes
10 '04	N: BH ⁱⁱ HH von hinten, zu MM gewandt; BM JC im Profil von re, blickt MM an; BH ^{re} MM, blickt JC an, Geste mit offener Handfläche, schüttelt schnell den Kopf, richtet sich leicht auf und beugt sich wieder runter, schüttelt erneut den Kopf, zieht die Augenbrauen hoch und lächelt breit; VG aufsteigender Zigarettenrauch	Mu: ((HG: Streichermusik, weibliche Gesangsstimme)) HH: night, MM: <<p> (we=got= a) Jewish holiday on monday (-) hey won't find=out till tuesday (1) > beautiful

Das durchweg in Nahaufnahme präsentierte – hier zu zwei Dritteln protokollierte – Gespräch zwischen den drei Männern an der Bar wird alternierend von rechts und links aufgenommen (Einstellungen 02–10), dem jeweiligen der beiden Sprecher zugewandt, die dem in ihrer Mitte stehenden Jimmy den Coup schmackhaft machen. Die Drei stecken ihre Köpfe zusammen; sie hecken sichtbar etwas aus, wie es selbst dann verständlich wäre, wenn man ihre Worte nicht verstehen könnte. Die Nahsicht der Bildführung lässt sie eine Enklave innerhalb des übrigen Raums bilden. Diese Abkapselung wird zugleich durch den Flüsterton des Gesprächs bewirkt, dem gegenüber alle Hintergrundgeräusche nun fast ganz zurückgenommen werden. Obwohl es sich nur um einen kurzen Dialog handelt, arbeiten auch hier Ton und Bild Hand in Hand; zusammen erzeugen sie eine insuläre Episode innerhalb des zwielichtigen Glamours der abendlichen Geselligkeit.

Mein nächstes Beispiel präsentiert eine Gesprächssituation ganz anderer Art. Es handelt sich um einen Ausschnitt aus der politischen Talkshow *Anne Will*, ausgestrahlt in der ARD am 23. 9. 2007. Das Thema der Sendung lautete: „Deutschland vor dem Anschlag? Das Kalkül mit der Angst“. Geladen – und in einem Halbkreis um die Moderatorin gruppiert – waren die Politikerin Renate Künast (RK) und der Politiker (und Innenminister) Wolfgang Schäuble (WS) auf der linken, sowie der Politiker (und ehemalige Innenminister) Gerhart Baum (GB) und der israelische Diplomat Avi Pri-

mor (AP) auf der rechten Seite. Im Verlauf der Sendung haben die politischen Kontrahenten Baum und Schäuble ein etwas überraschendes Männerbündnis gegen Renate Künast geschlossen, der sie mit überlegenem Lächeln und spitzen Bemerkungen den Wind aus den Segeln zu nehmen versuchen. In der Sequenz, auf die es mir hier ankommt, setzt sich ein schon länger andauernder Kampf um das Rederecht fort.

Talkshow: *Anne Will* (ARD), 23.9.2007

Thema: „Deutschland vor dem Anschlag? Das Kalkül mit der Angst“, Ausschnitt

<u>Nr.</u> <u>Zeit</u>	Bild	Ton
01 '8	D (US ¹): Renate Künast (RK) blickt nach rechts, gestikuliert Z ^h G:RK RK blickt nach links oben, gestikuliert leicht mit der rechten Hand, schüttelt leicht den Kopf	RK: [an der Stelle. WS: [ach Gott Frau Künast; RK: <<f> ja:> [so is=es. das ist das gleiche Bild. WS: [Sie red=n unter Ihr=m Niveau: (.) Publ.: ((Lachen)) WS: ((amüsiert)) sie reden wirklich [unter ihr=m Niveau: GB: [so einfach is=es nich. RK: [ihr Zwischenruf ist auch unter=m Niveau: (.) Publ.: ((Klatschen)) WS: <<pp> (.)> RK: und der Punkt, (.) [Ne: sie sind (-) WS: [De:s=isch wirklich
02 '4	T: Studiopublikum, mehrere Personen klatschen in die Hände, einzelnes Kopfschütteln, F ^h , S ^{re,1} , Personen stellen das Klatschen sukzessive ein	RK: [warte=ma) (-) WS: [traurig AW: Was [haben Sie Herr Baum gegen GB: [also äh (.) ja ich=möcht=an Herrn AW: [Online Durchsuchungen? RK [wir müssen an de:r Stelle diese Vorschläge auch zu Ende=denken und
03 '4	N: RK blickt nach rechts, wiederholte Auf- und Abbewegung der rechten Hand	RK: bring=sie=sie doch ma' .h bring=sie doch [ma Di:nge ein für die sie WS: [gucken Sie=mal; GB: [ja jetzt (.)
04 '6	N (US ¹): Gerhard Baum (GB) nach links vorne gelehnt, blickt nach links und deutet mit ausgestrecktem Zeigefinger der rechten Hand nach links,	RK: [Mehrheiten her:stell=n könn=n WS: [(jetzt lassen sie mich mal) zwei Sätze GB: [(.) (.) ja...hh WS: sagen.

Nr. Zeit	Bild	Ton
	GB nimmt die Hand zurück und lehnt sich nach rechts, GB lehnt sich zurück und zieht die rechte Schulter hoch	RK: statt immer nur das Absolutistische zu fordern mit dem sie [die Frei,heit (einreißen)]. WS: [es foddet doch; es foddet doch GB: [also wir wir sprechen ni' (.) wir RK: [das ist unser Sicherheitsproblem. GB: [sprechen (.) wir sprechen,
05 '3	N (AS ¹): Wolfgang Schäuble (WS) mit leicht nach vorn gesenktem Kopf, blickt nach links unten, WS blickt auf und wendet den Kopf nach rechts	RK: [auch bei (online). WS: [<<f> Es fordert doch niemand GB: [(WS: [Absolutistisches.> .h AW: [wir machen GB: [(also wir wir spreche wir sprechen) AW: [on:line Durchsuchungen; WS: [der=s
06 '3	G (US ¹): Anne Will (AW) wendet Kopf und Oberkörper sehr schnell nach rechts AW nickt mehrmals kurz, lächelt	AW: [für oder wider; okay? (.) GB: [(wir) besprechen aber ein se:hr AW: [des machen wir jetzt mal. hh GB: [schwieriges Problem. WS: [der sach' der Sachverhalt GB: [<<p> (wir sprechen gleich darüber)> WS: [der Sachverhalt war doch
07 '16	HT: Studiogäste, F ^{re} , S ^{li} , Z ^v : HN: WS mit nach vorne geneigtem Kopf, blickt zu RK, WS wendet den Kopf zu AW, AW blickt auf WS, WS Geste mit der linken Hand, blickt nach rechts, nach links, schüttelt leicht den Kopf; Z ^v :N: WS blickt nach rechts, legt die linke Hand auf seine Brust	WS: ganz einfach; (.) .hh die Sicherheitsbehörden; <<p> übrigens auch die Juschtiz die Bundesanwalt> schafft hat gesagt weil Kommunikation nicht nur üb=r Tele:fon erfolgt sondern auch über .h Indernet (.) äh=m tun wir in entsprechender An: wendung der Strafprozessordnung; dafür bin ich nicht zuschdändig sondern die Juschtizminister, (.)

Wie schon zuvor, erheben Baum und Schäuble in Wort *und* Bild, sowohl verbal als auch mimisch, immer wieder Einspruch gegen Künasts Ausführungen (verbal vor allem in E 01 und 02, gestisch insbesondere in E 04 und 05). Auch die Moderatorin versucht zwischendurch zu Wort zu kommen (E 05 und 06), aber schließlich kann Schäuble sich durchsetzen – und dies nicht allein verbal (E 06 und v.a. 07). Denn von einer Halbtotalen (E 07), die alle fünf am Gespräch beteiligten Personen erfasst, bewegt sich die Kamera in diesem Augenblick per Fahrt, Schwenk und Zoom mit einer vergleichsweise großen, ausholenden Geste frontal auf den Minister Schäuble zu und *unterstreicht* gleichsam das Wort, das er sich zunächst durch einfaches Weiterreden erstritten hat. Mit dieser auffälligen Geste der Kamera wird er

bildlich aus allen herausgehoben als einer, der die Dinge zurecht rücken kann; sie verleiht ihm eine Autorität, die Künstler in der gesamten Sequenz davor gerade durch die Bildregie nicht gewährt wurde.⁷

In dieser Szene wird deutlich, dass die Bildregie hier weitgehend unabhängig von der Kraft und dem Gewicht der jeweils verbal geäußerten Standpunkte einen eigenen rhetorischen Standpunkt zu formulieren vermag. Sie hebt die Position des Innenministers Schäuble dramaturgisch so heraus, dass seine Position rein choreografisch als die Überlegene erscheint. Die Bildsprache der Sendung *gewichtet* das in ihr verbal und nonverbal Gesagte und legt damit etwas nahe, das vor allem durch die *Form* der bildlichen Präsentation durchaus nachdrücklich artikuliert wird. Dies ist keineswegs ein Einzelfall. Denn generell ist von Talkshows zu sagen, dass sie alles andere als ein gefilmtes Radio darstellen, bei dem man sich das von der Bildregie Dargebotene auch sparen könnte. Denn der allein sichtbare, nicht aber hörbare Dialog der Mienen und Gesten, der Körperhaltung, je nach Sendung auch der Bewegung der Moderatoren, mit einem Wort: der bildlich bewegte und dadurch auf besondere Weise bewegende Schauplatz des verbalen Austauschs ist wesentlich für das, was sich in diesen Sendungen abspielt. Überspitzt gesagt: Das Bild macht die Musik, der Ton bewegt die Bilder.

Dies aber bedeutet: Nicht allein in dem Filmdialog des vorigen Beispiels, auch in dieser Talkshow-Szene ergibt sich der volle kommunikative Gehalt der präsentierten Gesprächssituation allein aus der Einheit von bildlicher und akustischer Dimension. Trotzdem operiert die Kamera in beiden Fällen auf höchst unterschiedliche Weise. Während die wechselnden Kamera-Einstellungen in dem Dialog aus *GoodFellas* untrennbar mit der Situation verbunden bleiben, die sie zusammen mit der Tonspur erzeugen, spielt die Kamera in dem ostentativen Schwenk und Zoom auf Schäuble gleichsam ein eigenes Spiel. Sie erzeugt Verbindungen innerhalb eines fortlaufenden Live-Gesprächs, das den Fernsehzuschauern ständig – und erkennbar – aus den Perspektiven *verschiedener* Kameras dargeboten wird, die das Geschehen aus jeweils *ihrer* Perspektive darbieten. In der gezeigten Einstellung kann die Kamera deshalb (innerhalb der von der Regie arrangierten Montage) als ein eigener Akteur auftreten, der zu dem übertragenen Gesprächsgeschehen eine auffällige Stellung bezieht. Die Kameraführung schafft hier nicht – wie in *GoodFellas* – eine eigene abgegrenzte Szene innerhalb einer bereits etablierten größeren Szene, sondern interveniert gleichsam in das Gespräch, indem sie mit ihrer Sicht *auf* es zugleich eine Ansicht *über* es formuliert.

Mithilfe meines vierten – und letzten – Beispiels möchte ich einen weiteren Vergleich vornehmen. Ich beziehe mich auf eine Passage aus einem Korrespondentenfilm, der in der Sendung *Weltspiegel* wiederum in der ARD am 7.9.2003 ausgestrahlt wurde.⁸ In seinem ersten Teil berichtet er von

⁷ Für eine ausführlichere Interpretation des Kontexts dieses Ausschnitts vgl. Keppler (2009).

⁸ Für eine Analyse des gesamten Magazinfilms vgl. Keppler (2006, S. 230–237).

einer Fahrt von Jordanien nach Bagdad. Ein Flüchtlingslager, der Grenzposten, ein Café, eine Raststätte sind die ersten Stationen. Es wird viel aus dem fahrenden Auto gefilmt; dazwischen sind Statements von Einheimischen geschnitten; dies verleiht der Eröffnung einen leicht epischen Rhythmus. Dann geht es weiter in Richtung Falludscha, der, wie es in dem Bericht schon vorher geheißen hatte, „Hochburg des Widerstands der Saddam-Anhänger gegen die Amerikaner und die Hochburg der Wegelagerer entlang dieser Autobahn“. Auf diesem Weg wird das Fernsehteam Zeuge einer Schießerei zwischen der Polizei und einer Bande von Autodieben.

Weltspiegel (ARD), 7.9.2003

Thema: Mit US-Soldaten in Bagdad auf Patrouille

Nr. Zeit	Bild	Ton
01 '7	HN: Auto, S ^{li, o} , dahinter drei Männer, lachend zu Mann mit Maschinengewehr (MG), im HG Gebäude	G: ((HG: Stimmen)) R ^m : auch der neuen Polizei wird nicht viel zugetraut (-) dabei haben sie vielleicht den gefährlichsten Job im
02 '3	T, B ^{hk} (S ^{li} , S ^{re}): Mann rennt von Straße in Steppe, duckt sich, schaut nach hinten, Kamera folgt ihm	G: ((HG: Rauschen, Schüsse)) R ^m : ganzen Irak wie wir kurze Zeit später heute nachmittag miterleben können
03 '5	A, B ^{hk} : Mann vor geöffnetem Kofferraum eines Autos, S ^{li} , S ^{re, o} , F ^v in Kniehöhe, mehrere Männer laufend	G: ((HG: Rauschen, Schüsse)) R ^m : bei Falludscha geraten wir in eine Schießerei zwischen Gangstern und einer Polizeitruppe
04 '3	HT, Z ^{h, l} : Mann mit MG rennt auf die Kamera zu, schießt	G: ((HG: Schüsse)) R ^m : Autodiebstahl.
05 '3	W, B ^{hk} : Straße, Leitplanke, dahinter mehrere Personen in Wüste, zwei Autos, Luft flimmert	G: ((HG: Fahrtwind)) R ^m : die drei Gangster werfen sogar mit Granaten auf die schwach
06 '7	T, B ^{hk} : am BR ^{li} Polizist mit MG, Auto, zwei Männer stehen vor einem liegenden Mann, winken, Auto fährt schnell durchs Bild	G: ((HG: Schreie)) R ^m : bewaffneten Polizisten (-) die Verbrecher entkommen mit ihren schnellen Autos (.) ein Polizist liegt schwer verletzt am Boden
07 '7	HN, F ^{v, hk} : fahrender Pickup, auf Ladefläche der Verwundete, Polizisten beugen sich über ihn, viel Blut um sie herum	G: ((Fahrgeräusche '1sec laut, dann HG)) R ^m : verzweifelt versuchen seine Kollegen ihn zu (.) reanimieren= vergeblich (.) diese Polizeiautos sind viel
08 '3	T, F ^{v, li, hk} : Aufnahme aus dem Auto, im Außenspiegel Blick auf Hütten	G: ((HG: Fahrgeräusche)) R ^m : zu langsam haben kein Blaulicht und keine Sirene

Nr. Zeit	Bild	Ton
09 '6	HT, S ^{re, hk} , F ^{v, hk} , S ^{li, hk} zwei Polizisten tragen Verletzten vom Krankenwagen ins Krankenhaus	G: ((HG: Schritte, Schreie)) R ^m : der Polizist (.) fünfunddreißig Jahre alt, Vater von fünf Kindern kommt nicht mehr zu Bewusstsein
10 '7	HT: Raum mit vielen Männern, Verletzter auf Liege, F ^{v, li} , Z ^v Arzt bei Herz-Massage, Arzt hebt den Arm des Toten und lässt ihn fallen	G: ((Stimmen '1sec laut dann HG)) R ^m : die Ärzte im Krankenhaus können nur noch seinen Tod feststellen (.) das war heute Nachmittag um zwei Uhr

Liest man in diesem Transkript allein den von dem – in der Anmoderation der Sendung namentlich genannten – Korrespondenten Jörg Armbruster gesprochenen Text, so könnte man denken, dieser enthalte bereits alle relevanten Informationen über den berichteten Vorgang. So sehr dieser Kommentar aber auch unabhängig von den ihm unterlegten Bildern als Nachricht verstanden werden kann, der Filmbericht geht in seinen verbalen Mitteilungen keineswegs auf. Zunächst ist festzuhalten, dass der gesprochene Kommentar zu den im Bild sichtbaren Ereignissen hier eine deutlich andere Rolle spielt als derjenige am Beginn der besprochenen Sequenz aus *GoodFellas*. Dort wird – wie im Verlauf des Films immer wieder – die Stimme aus dem Off in das Geschehen der Spielhandlung eingeflochten, um ihr eine zusätzliche zeitliche Dimension zu geben und ihr vom Ende der erzählten Geschichte her eine vorgreifend melancholische Perspektive zu verleihen. Die kommentierende Rede fügt sich auf diese Weise in den Großrhythmus der fiktionalen (wenn auch auf *wahren Begebenheiten* basierenden) Erzählung ein (wie es mit anderen Mitteln durch die Klangkollage während der Fahrt durch Los Angeles in *Zabriskie Point* geschieht). In dem Magazin-film dagegen wird nichts kunstvoll eingeflochten; der Bericht des Korrespondenten wird über das zusammengeschnittene Bildmaterial darüber gesprochen. Anders als in *GoodFellas* wird genau das kommentiert, was in den Bildverläufen zu sehen ist. Jedoch bedeutet dies kein Weniger an Interaktion zwischen Bild und Ton. Wir haben es hier lediglich mit *anderen* Korrespondenzen zwischen Bild und gesprochenem Text zu tun.

Für den Duktus dieser Kommentierung ist dabei zunächst die genaue Datierung äußerst wichtig. Der Film handelt von der gefährlichen Arbeit der Polizisten im Irak, „wie wir“ – gemeint ist das Team des Korrespondenten, angesprochen sind aber auch die Zuschauer, die hiervon noch am selben Tag im Fernsehen erfahren – „heute Nachmittag miterleben können“ (E 02). Am Ende dieser Passage wird diese Datierung noch einmal ausdrücklich wiederholt; die Zuschauer sehen eine Szene, so der Reporter, die sich „heute Nachmittag um zwei Uhr“ ereignet hat (E 10). Durch die genauen zeitlichen Angaben, das Präsens, mit dem der Bericht eingeleitet wird und nicht zuletzt durch die Kameraführung – u.a. den sichtbaren Ein-

satz einer Handkamera (E 03, 05, 07, 09) – wird ein hoher Anspruch auf Authentizität erhoben: Der Journalist ist dabei gewesen, als es auf dem Weg nach Bagdad zu tödlicher Gewalt gekommen ist. Der Umstand, dass der genaue Verlauf der Gewalthandlung in den Bildern nicht zu erkennen ist, widerspricht diesem Anspruch nicht, sondern bestätigt ihn vielmehr in ausgezeichneter Weise. Denn die plötzlich eingetretene Situation macht eine geordnete Darstellung ihres Verlaufs durch das von den Ereignissen überraschte und seinerseits bedrohte Team unmöglich. Man sieht einen Polizisten eine Straße entlang rennen (E 03), dann entnehmen weitere Polizisten Waffen aus ihrem Fahrzeug, einer von ihnen schießt eher planlos ins Weite (E 04), abseits der Straße sieht man einige Autos und Personen stehen (E 05), dann rückt der angeschossene Polizist ins Bild (E 06), über dessen blutigen Abtransport (E 07–09) in ein Krankenhaus und die dortige Feststellung seines Todes (E 10) abschließend berichtet wird.

Der entscheidende Punkt aber ist: Es *müssen* nach Kinostandards unbeholfene Bilder sein, die glaubhaft von einem Geschehen in der unmittelbaren Vergangenheit berichten. Denn die plötzlich eingetretene Situation, so kann jeder am Bildschirm es mitverfolgen, macht eine dramaturgisch wohlkalkulierte, aus dem Zentrum des Geschehens operierende Darstellung ihres Verlaufs unmöglich. Dass dies wir als Zuschauer dem Bildverlauf *ansehen* können, ist hier ein wesentlicher Teil der Botschaft, der das verbal Gesagte nicht allein illustriert, sondern mit großem Nachdruck visuell *bezeugt*. Auch hier, so meine ich daher, liegt ein nichthierarchisches Verhältnis von Bild und Ton vor. Der Korrespondent muss sich nicht auf das Hörensagen ungenannter Zeugen verlassen, sondern er hat mit eigenen Augen *gesehen*, was dort vorging, und er lässt es uns, die Zuschauerinnen und Zuschauer am Bildschirm, kraft der eingefangenen Bilder mit *unseren* Augen sehen. Diese *Übertragung* der Zeugenschaft leistet hier der bildliche *Verlauf*.

III.

Ich breche meine kleine Beispielreihe an dieser Stelle ab und komme zu meinen grundsätzlichen Überlegungen zurück. Ich habe zu Beginn die These vertreten, dass Bild und Ton in filmischen Erzeugnissen jeglicher Art *gleichursprünglich* an der Konstitution der in ihnen präsentierten Gehalte beteiligt sind. Dies, so hatte ich gesagt, ist gerade auch dort der Fall, wo eines dieser Elemente den filmischen Verlauf zu dominieren scheint. Diesem Schein war insbesondere die Interpretation des letzten Beispiels gewidmet, aber auch die Analyse der Dialogszene in Scorseses Film. Dabei hat sich gezeigt, dass das, was dominant zu sein *scheint*, in diesen beiden Fällen das gesprochene Wort, gar nicht allein dominant *ist*, ja mehr noch: es gar nicht allein dominieren *kann*. Freilich: Wenn *dominieren* in Szenen wie diesen bedeutet, als erstes aufzufallen, allein bereits wesentliche Informationen zu

enthalten und in diesem Sinn die Führung der Wahrnehmungssteuerung zu übernehmen, dann kann man sagen, dass in dem Korrespondentenfilm und vielleicht auch in der kleinen Dialogszene aus *GoodFellas* das gesprochene Wort *dominiert*. Aber eben nur dann. Denn sobald man unter dem *Dominieren* des filmischen Tons gegenüber dem filmischen Bild – oder umgekehrt des Bildgeschehens gegenüber dem Klanggeschehen, wie es die Szene aus *Zabriskie Point* nahe legen könnte – eine Relation der einseitigen Abhängigkeit der einen gegenüber der anderen Dimension verstünde, läge man zuverlässig falsch. Erst recht läge falsch, wer ein solches Dominieren, wo es denn überhaupt gegeben ist, als ein Verhältnis der Determination begreifen würde. Denn so sehr Bild und Ton je für sich genommen *wesentliche Informationen* über das jeweils Dargebotene enthalten können, *die* Botschaft, Erzählung, Atmosphäre und Geste eines filmischen Produkts formulieren sie immer *zusammen*. Das ist es, was ich meine, wenn ich von einem nicht-hierarchischen Verhältnis von Bild und Ton im Kino und Fernsehen gleichermaßen spreche. Immer geschieht Modifikation des Bildes durch Text und Ton und umgekehrt; sobald wir von filmischen Prozessen sprechen, kann von einseitigen Abhängigkeiten keine Rede sein.

Doch gerade wenn dies zutreffend sein sollte, stellt sich sogleich eine andere Frage – diejenige nach dem Unterschied zwischen der Ästhetik des Kinofilms und derjenigen des Fernsehens. Denn die Standardauffassung lautet, dass die beiden Medien sich insbesondere an der Art der Dominanz von Bild und Ton unterscheiden. In einer Erörterung dieses Verhältnisses in seinem Buch über *Visible Fictions* beispielsweise geht John Ellis von einer „unterschiedlichen Ton/Bild-Balance“ in Kino und Fernsehen aus. „Im Fernsehen verankert gewöhnlich der Ton die Bedeutung, im Film hingegen das Bild“ (Ellis 2001, S. 60).⁹ Dies klingt genau wie die Fehldeutung, die ich gerade kritisiert habe. Interessanterweise jedoch fährt Ellis fort: „In beiden Fällen handelt es sich um eine Frage der Gewichtung und nicht um einfache Abhängigkeiten des einen vom anderen. In beiden Medien existieren Bild und Ton in einem Wechselverhältnis, sie treten nicht als getrennte Einheiten auf“ (ebd.). Diese für sich genommen einleuchtende Aussage ist allerdings mit dem zuvor geäußerten Gedanken einer konträren *Verankerung* der Bedeutung einerseits im Bild und andererseits im Ton nur schwerlich vereinbar.

Worin die strukturelle Differenz von Kino und Fernsehen stattdessen besteht, möchte ich abschließend kurz andeuten. Hier ist erstens der von Raymond Williams (1974) wirkungsmächtig analysierte *Flow*-Charakter der Fernseherfahrung zu nennen. Dieser ist dadurch gekennzeichnet, dass das Fernsehen seinen Zuschauern die seit längerem zunehmend genutzte Möglichkeit bietet, nicht so sehr einzelne Sendungen (einschließlich der

⁹ Vgl. ebd. S. 66: „Die unterschiedliche Gewichtung zwischen Sehen und Hören erzeugt eine qualitativ andere Beziehung zur Fernsehsendung. Nicht dass die Erfahrung weniger intensiv wäre als im Kino; sie hat vielmehr ihre eigene Form.“

gezeigten Spielfilme), sondern *Segmente* diverser Sendungen in einem beliebigen Wechsel zu verfolgen. Hier ist zweitens daran zu erinnern, dass das Fernsehen seinen Zuschauern stets – simultan wie sukzessiv – ein ganzes Kaleidoskop von Gattungen präsentiert, die in ausdrücklicher oder stillschweigender Interferenz auch dort miteinander verbunden bleiben, wo die Betrachter von Anfang bis Ende bei *einer* Sendung verweilen (vgl. Keppler 2006, S. 82 ff.). Drittens spielt – durchaus! – der Ton eine besondere Rolle in der Fernsehkommunikation. In einer subtilen Untersuchung über den *Television Sound* hat Rick Altman diesem die Funktion zugeschrieben, die oft durch vielerlei häusliche Aktivitäten abgelenkten Zuschauer „zum Bild zurückzurufen“ (Altman 2001, S. 400). Das Fernsehen, sagt Altman, gibt den Zuschauern, auch wenn sie in der Küche oder anderswo im Haushalt zugange sind, „das Gefühl (...), alles wirklich Wichtige“ werde im Fernsehen durch den Ton „angekündigt“ (Altman 2001, S. 394). *Ankündigung* aber meint hier keineswegs, dass im Fernsehton tatsächlich alles „wirklich Wichtige“ *zum Ausdruck* käme, dass also der Gehalt von Sendungen, die sich so bemerkbar machen, vorwiegend in ihrer akustischen Dimension *verankert* wäre. Vielmehr gilt der von Altman analysierte Lockruf des Fernsehtons gerade der *Einheit* von Bild und Ton, deren Attraktionen auch die vom Schauen abgelenkten Zuschauer erreichen sollen. In diesem Sinn – aber auch *nur* in diesem – macht der Ton weniger *im* als vielmehr *beim* Fernsehen durchaus eine besondere Musik.

Anhang: Transkriptionssystem

Beschreibungsinventar für das tabellarische Filmprotokoll und Erläuterung der Abkürzungen für visuelle und auditive Elemente

1. Visuelle Dimension

1.1 Kameraoperationen

1.1.1 Einstellungsgrößen

D Detailaufnahme: eng begrenzter Bildausschnitt, Großaufnahme von Gegenständen.

G Großaufnahme: Konzentration auf den Kopf / das Gesicht bis zum Hals.

N Nahaufnahme: Brustbild; Darstellung von Personen vom Kopf bis Mitte des Oberkörpers; neben den mimischen werden auch gestische Elemente sichtbar. Oft für die Darstellung von Diskussionen und Gesprächen verwendet.

HN Halbnahe: Darstellung Kopf bis zur Taille; Aussagen über die unmittelbare Umgebung der abgebildeten Personen werden möglich. Oft zur Darstellung von Personen im Dialog.

- HT Halbtotale Einstellung: Menschen von Kopf bis Fuß, oft zur Darstellung von Personengruppen verwendet.
- T Totale: ganze Person mit Umgebung; gibt einen Überblick über den Handlungsraum.

1.1.2 Kamerabewegungen

- Z Zoom
- F Fahrt
- S Schwenk
- TS Tiefenschärfe
- hk Handkamera
- B Bewegte Kamera

Richtung der Kamerabewegung (jeweils hochgestellt hinter Kamerabewegung)

- | | | | |
|----|-------------|----|-------------|
| v | nach vorn | o | nach oben |
| h | nach hinten | u | nach unten |
| li | nach links | re | nach rechts |

1.1.3 Kameraperspektive (in Klammer hinter Einstellungsgröße)

- AS Aufsicht / Vogelperspektive
- US Untersicht / Froschperspektive
- l leicht
- s stark

1.2 Schnitt

- Ü Überblende: Einzelbilder von Einstellung A überlappen mit Einzelbildern von Einstellung B
- Schnitt: auf letztes Einzelbild von Einstellung A folgt unmittelbar erstes Einzelbild von Einstellung B

1.3 Elemente der Bildkomposition

1.3.1 Insert

Inhalt des Inserts kursiv; Besonderheiten (Groß- / Fettschrift etc.) werden übernommen

1.3.2 Lokalisierung von Personen oder Gegenständen im Raum

VG	Vordergrund	BR	Bildrand
HG	Hintergrund	BH	Bildhälfte
BM	Bildmitte		

2. Akustische Dimensionen

2.1 SprecherInnen

NN bei bekannten SprecherInnen Abkürzung des Vornamens und des Nachnamens

(A) SprecherIn ist vermutlich SprecherIn A

() SprecherIn nicht identifizierbar

2.2 Gesprochene Sprache (wird nach Bedarf erweitert nach den Regeln des GAT-Transkriptionssystems für gesprochene Sprache (vgl. Selting et al. (1998))

(.)	Mikropause unter 0.25 Sek. Dauer
(-), (--), (---)	geschätzte Pausen von ca. 0.25 - 0.75 Sek. Dauer
(2)	geschätzte Pause in Sekunden ab Pausendauer von ca. 1 Sek.
[Beginn einer Überlappung bzw. gleichzeitiges Sprechen
ja:::	Dehnung; Anzahl der Doppelpunkte entspricht Länge der Dehnung
?	stark steigende Intonation
,	schwach steigende Intonation
;	schwach fallende Intonation
.	stark fallende Intonation
=	Verschleifung innerhalb von Einheiten, z.B. und=äh
hm, ja, nee	einsilbige Rezeptionssignale
hm=hm, ja=a, nei=ein	zweisilbige Signale
,hm'hm	mit Glotalverschlüssen, meistens verneinend
waru'	Abbruch eines Wortes oder einer Äußerung
.h, .hh, .hhh	hörbares Einatmen, je nach Dauer
h, hh, hhh	hörbares Ausatmen, je nach Dauer

a(h)ber	Aspirationslaut oder Lachpartikel beim Sprechen
((lacht))	Umschreibung von para-linguistischen Information (Lachen, Husten, Räuspern etc.), äußere Klammer kennzeichnet Anfang und Ende
()	unverständliche Textpassage
(und)	vermuteter Wortlaut, nicht sicher rekonstruierbar
<<f >>	forte, laut, (z.B. <<f> Hilfe:>)
<<ff >>	fortissimo, sehr laut
<<p >>	piano, leise
<<pp >>	pianissimo, sehr leise

Es ist wichtig, hier zu beachten, dass Satzzeichen in den Filmprotokollen nicht im üblichen Sinn zur Interpunktion verwendet werden, sondern dazu, steigende bzw. fallende Intonation am Satzende zu markieren! Um die Lesbarkeit der Filmprotokolle zu verbessern, haben wir in Abweichung von den GAT-Konventionen im gesprochenen Text die Groß- und Kleinschreibung beibehalten; dies betrifft jedoch nicht die Markierung von Satzenden bzw. -anfängen.

Literatur

- Adelmann, Ralf et al. (Hg.) (2009): Grundlagen zur Fernsehwissenschaft: Theorie, Geschichte, Analyse. Konstanz.
- Albrecht, Gerd (1964): Die Filmanalyse – Ziele und Methoden. In: Everschor, Franz (Hg.): Filmanalysen 2. Düsseldorf, S. 233–270.
- Altman, Rick (2001): Fernsehton. In: Adelmann (Hg.), S. 388–412.
- Bonfadelli, Heinz (2002): Medieninhaltsforschung: Grundlagen, Methoden, Anwendungen. Konstanz.
- Deleuze, Gilles (1991): Das Zeit-Bild. (= Kino 2). Frankfurt/M.
- Ellis, John (2001): Fernsehen als kulturelle Form. In: Adelmann (Hg.), S. 44–73.
- Goodman, Nelson (1995): Sprachen der Kunst. Frankfurt a.M.
- Grimm, Jürgen (2006): Super Nanny. Ein TV-Format und sein Publikum. Konstanz.
- Heritage, John/Clayman, Steven E./Zimmerman, Don H. (1988): Discourse and message analysis. The micro-structure of mass media messages. In: Hawkins, Robert P./Wiemann, John M./Pingree, Suzanne (Hg.): Advancing communication science: Merging mass and interpersonal processes. London, S. 77–109.
- Keppler, Angela (2005): Die Einheit von Bild und Ton. Zu einigen Grundlagen der Filmanalyse. In: Mai, Manfred/Winter, Rainer (Hg.): Das Kino der Gesellschaft – die Gesellschaft des Kinos. Interdisziplinäre Positionen, Analysen und Zugänge. Köln. S. 60–78.

- Kepler, Angela (2006): *Mediale Gegenwart. Eine Theorie des Fernsehens am Beispiel der Darstellung von Gewalt*. Frankfurt a.M.
- Kepler, Angela (2009): *Das Gesagte und das Nichtgesagte. Was die Dramaturgie politischer Talkshows zeigt*. In: Girnth, Heiko/Michel, Sascha (Hg.): *Multimodale Kommunikation in Polit-Talkshows*. [Ersch. demn.]
- Merten, Klaus (1995): *Inhaltsanalyse: Einführung in Theorie, Methode und Praxis*. 2., verb. Aufl. Opladen.
- Reichertz, Jo (2000): *Die frohe Botschaft des Fernsehens. Kulturwissenschaftliche Untersuchung medialer Diesseitsreligion*. Konstanz.
- Selting, Margret et. al. (1998): *Gesprächsanalytisches Transkriptionssystem*. In: *Linguistische Berichte* 173, S. 91–122.
- Williams, Raymond (1974): *Television, technology and cultural form*. London.