

Michael Niehaus

Voice over. Eine filmnarratologische Bestandsaufnahme

Der Beitrag beschäftigt sich mit dem Phänomen des voice over im Hinblick auf seine Stellung im Narrationsprozess im Film. Anhand einer kritischen Lektüre des neuesten filmnarratologischen Systematisierungsversuches von Markus Kuhn wird gezeigt, dass sich der Ort der voice-over-Stimme im Film weder auf seine narrative Funktion reduzieren noch im Rahmen der binären Oppositionen erzählen/zeigen, sprachlich/visuell und intradiegetisch/extradiegetisch zureichend bestimmen lässt. Die Kategorie der (narrativen) Instanz wird, so der Befund, in der Filmnarratologie nicht zureichend bestimmt.

Die Stimme, die über den Film gelegt wird oder ihn zu steuern vorgibt, die in ihn einbricht oder sich ihm aufpropft, ist trotz ihres vielerorts schlechten Rufes nicht mundtot zu machen. Solange man der Auffassung huldigte, dass das eigentliche Reich des Films das Visuelle mit einer ihm eigentümlichen Sprache sei, lag es nahe, die Stimme, die keinem Körper zugehört, als einen Fremdkörper aufzufassen – als ein diskursives Element, das die Bilder eher zum Schweigen bringt denn anreichert.¹ Es schien, als versuche eine Erzählstimme das Medium Film an eine literarische Instanz zurückzubinden, von der es sich befreien musste. Es schien auch, als zwingte diese Instanz die optischen wie die akustischen Bilder unter eine Linearität, die ihnen nicht gemäß sei. Das *Verhältnis* zwischen der Instanz der Stimme und den gezeigten Bildern blieb dabei jedoch weitgehend ungedacht.

¹ Die tendenzielle Unterschlagung des Diskursiven beginnt freilich schon mit der Geringschätzung der Zwischentitel im Stummfilm. Jerome Philipp Schäfer bemerkt, dass „eine stiefmütterliche Behandlung von Sprache im Film ein echtes Problem darstellt“ (Schäfer 2008, S. 23). Es ist allerdings vielsagend, dass Schäfer, um im Anschluss an Lotman die Bedeutung der sprachlichen Elemente zu betonen, auf die ikonische Dimension von Textinserts verweist.

Die noch junge Filmnarratologie, die ein begriffliches Instrumentarium bereitstellt, mit dem beschrieben werden kann, *wie* Filme erzählen, eröffnet hier eine vorurteilsfreiere Betrachtungsweise. Seit 2011 ist mit dem Buch *Filmnarratologie. Ein erzähltheoretisches Analysemodell* von Markus Kuhn eine erste systematische Untersuchung im deutschsprachigen Raum verfügbar, die den Anspruch erhebt, das gesamte Feld der Filmnarratologie zu entfalten sowie die bisherigen Konzepte zusammenzufassen und einer kritischen Würdigung zu unterziehen.² Die folgenden Überlegungen sind schon deshalb gehalten, sich an dem mustergültigen Versuch von Markus Kuhn zu orientieren, weil in ihm, anders als in den bisherigen Ausführungen zum *voice over*, die Vielgestaltigkeit des Phänomens – wenn man so sagen darf – vor Augen gestellt wird. Im Durchgang durch diese Vielgestaltigkeit wird sich dann zeigen, um welchen Preis filmnarratologische Beschreibungen des Phänomens Stimme Herr werden können. Der Durchgang dient also zugleich dazu, nach den Grundlagen der Filmnarratologie zu fragen. Dabei geht es vor allem um das Verhältnis von Medium und Instanz.

I. Narratologische Voraussetzungen

Überträgt man die klassische Frage von Gérard Genette „Wer spricht?“, die bei ihm unter der Kategorie *Stimme* verhandelt wird, auf den Film, so scheint sie zunächst lediglich auf „fakultativ vorhandene sprachliche Erzählinstanzen“ anwendbar zu sein – solche eben, die man tatsächlich sprechen hört. Die Frage muss Kuhn zufolge daher erweitert werden zu der „medienneutraleren“ Frage „Wer erzählt?“ (Kuhn 2011, S. 73). Kuhns Antwort, dass Genettes Kategorie der *Stimme*, um auch „audiovisuelle Merkmale“ umfassen zu können „von der metaphorischen Implikation befreit gelöst und zum Begriff einer *narrativen Instanz* oder *Erzählinstanz*“ erweitert werden muss, scheint freilich ans Tautologische zu grenzen. Im Übrigen wäre das ein Schritt, den auch die Erzähltheorie der Literatur

² Den Stellenwert des Buches von Kuhn bestätigen – trotz kritischer Einwände – auch die beiden umfangreichen Rezensionen in IASL; vgl. Brüssel 2011 und Palmier 2011.

zu betreffen hätte, wo der Begriff der Stimme nicht weniger metaphorisch ist.

In der Regel operieren die Narratologen mit der (von der platonischen Unterscheidung zwischen *diegesis* und *mimesis* hergeleiteten) Opposition *eigentliche Erzählung* vs. *szenische Erzählung*, *berichtende Erzählung* vs. *szenischer Darstellung* oder *telling* vs. *showing* usw. (vgl. etwa Martinez/Scheffel 2005, S. 48). Man nimmt dann allerdings, was das zweite Glied angeht, zwei einander eigentlich ausschließende Metaphoriken in Anspruch: So, wie im schriftlichen Diskurs nur im übertragenen Sinne eine Stimme zu hören ist, so wird auch nur im übertragenen Sinne eine Szene gezeigt. Denn genau das, was am *showing*, an der *szenischen Darstellung* nicht wörtlich genommen werden darf, ist es, was der *Stimme* im übertragenen Sinne zugeordnet werden kann. Das Szenische, das an seinem angestammten Ort im *Drama* kein ‚vermittelndes Kommunikationssystem‘ benötigt,³ ist in der *szenischen Erzählung* ebenfalls durch eine ‚Stimme‘ vermittelt bzw. gerahmt. Aber je ausschließlicher sich die Erzählung am Szenischen orientiert (die Kurzgeschichten von Hemingway sind hier das klassische Beispiel), desto mehr dissimuliert sich die metaphorische Stimme. Die filmnarratologische Entscheidung, sich für die Beschreibung der audiovisuellen Merkmale von der Kategorie Stimme zu verabschieden, ist bereits in der literarischen Erzähltheorie (und in den literarischen Erzähltechniken) angelegt. Im Modell von Markus Kuhn führt dies zu der strategischen Entscheidung, eine „analytische Trennung von *sprachlichem* und *kinematographischem* (bzw. *audiovisuellem*) Erzählen im Film“ (Kuhn 2011, S. 75) vorzunehmen. Das ist Kuhns Antwort auf die Frage: „Gibt es einen Film-erzähler?“ (Ebd.) Im Falle einer bejahenden Antwort auf diese Frage gibt es natürlich eine ganze Reihe von Benennungsvorschlägen. Kuhn zählt sie auf: „grand imagier“ (Laffay), „Enunziator“ (Metz), „intrinsic narrator“ (Black), „film narrator“ (Lothe), „filmic composition device“ (Jahn) „fingiertes Aussagesubjekt“ (Lohmeier) usw. (vgl. Kuhn 2011, S. 76).⁴

³ Vgl. als *locus classicus* hierfür Pfister 1982, S. 47f.

⁴ Vgl. für einen neueren Überblick über die Diskussionen dieser Frage auch Schäfer 2008. Insgesamt ist festzustellen, dass bei den meisten

Näher setzt sich Kuhn mit dem Begriff des „cinematic narrator“ von Seymour Chatman auseinander. Denn in dessen Konzeption findet er den eigenen Vorschlag einer analytischen Trennung von sprachlicher und visueller Erzählinstanz gleichsam vorbereitet. Chatman definiert den *narrator* „in a broad sense“ als einen „Presenter“ und unterteilt ihn in den „Tell-er“ und den „Show-er“ als die beiden grundlegenden Möglichkeiten, eine Geschichte zu präsentieren: „The difference between telling and showing then comes down simply to the implied author’s choice of signs – analogous or motivated for mimetic narratives, arbitrary ‚symbolic‘ for diegetic narratives, and a mixture for mixed narratives.“ (Chatman 1990, S. 113f.) Diese Unterscheidung ist zunächst einmal nicht nur auf den Film gemünzt. Zum mimetisch Narrativen (das dem „Show-er“ zugeordnet ist) zählt Chatman zwar auch den Film, vor allem aber das Drama. Denn das Drama muss sich aus dieser Perspektive (wie bei Platon) als der reine Fall des „showing“ darstellen, während das Besondere, dass das filmische Zeigen sich über eine vermittelnde Kamera vollzieht, für diese Unterscheidung zunächst keine Rolle spielt.

Was den „cinematic narrator“ angeht, so unterscheidet Chatman zunächst zwischen dem „auditory channel“ und dem „visual channel“, um dann aufzuzählen, dass ersterer sich in seiner Art unterscheiden könne (*noise, voice, music*) und in seinem Ursprung (*on-screen, off-screen*); beim komplexeren *visual channel* wird unter anderem das kinematographische Bild unterschieden, das in die Parameter *lighting, color, camera* und *mise-en-scène* zerfällt usw. Für Chatman ist der *cinematic narrator* „the composite of all these plus other variables“ (Chatman 1990, S. 135) – also das Konglomerat aller filmischen Mittel, denn alle filmischen Mittel haben Teil am spezifisch filmischen Erzählen. Wie kann der *cinematic narrator* dann aber eine Instanz, bzw. – nach Chatman – ein „agent“ (vgl. Chatman 1990, S. 119) sein? Antwort: der „cinematic narrator“ ist nichts als das formale Nadelöhr, durch welches die „guiding intelligence“ des „implied author“ (ebd.) das Erzählte ins Werk setzt. Das von der Narratologie immer wieder aus verschiedenen Gründen problematisierte

Überlegungen auf diesem Gebiet die Frage nach dem *voice-over-narrator* nur eine untergeordnete Rolle spielt.

Konzept des „implied author“ ist für diese Theorie unverzichtbar. Entsprechend gilt, dass „the synthesis as the narrator, of course, is achieved by the semiotic processes performed by the viewer“ (Chatman 1990, S. 135).⁵

Auf diese Weise ist freilich die Entpersonalisierung des „cinematic narrator“ bis zu einem Endpunkt getrieben. Denn einen derartigen „cinematic narrator“ wird sich der „viewer“ im Rezeptionsvorgang sicherlich *nicht* vorstellen. Was ihn möglicherweise umtreibt, ist der „implied author“, insofern sich der Rezipient nämlich darüber Gedanken machen kann, ‚was das Ganze soll‘. Die mit dem Etikett des „cinematic narrator“ versehene Instanz ist daher im Grunde nichts anderes als die *Entscheidung* eines „implied author“, ein *narratives* (und nicht zum Beispiel ein argumentatives oder ein deskriptives) Gebilde zu erschaffen.⁶ Dass alle filmischen Mittel, die in einem Film zum Einsatz kommen, irgendwie Einfluss auf die Art und Weise haben, wie der Film erzählt, dürfte unstrittig sein. In seinem früheren Buch *Story and Discourse* (1978) hatte Chatman die Notwendigkeit, in allen narrativen Gebilden einen *narrator* anzunehmen, anzunehmen, noch zurückgewiesen,⁷ jetzt glaubt er aus systematischen Gründen nicht auf ihn verzichten zu können; freilich fügt er hinzu: „But the narrator need not to be a ‚someone‘. Every narrative statement is presented by a narrator, and the narrator may be not a someone but a something. The agent of presentation need not to be human to merit the name ‚narrator‘.“ (Chatman 1990, S. 116) Das ist nur scheinbar unmissverständlich. Denn in diesem Konzept *muss* der *narrator* nicht nur kein „someone“ sein, vielmehr *kann* er – im

⁵ Für das Verhältnis von „implied author“ und „cinematic narrator“ gilt nach Chatman: „The cinematic narrator presents what the cinematic implied author requires.“ (Chatman 1990, S. 130) „In short, in cinema as in literature, the implied author is the agent intrinsic to the story whose responsibility is the overall design – including the decision to communicate it through one or more narrators“ (Chatman 1990, S. 132).

⁶ Chatman bezieht sich auf diese Modi im Rahmen seiner Unterscheidung verschiedener Weisen der Texterzeugung (vgl. Chatman 1990, S. 111ff.).

⁷ Dort heißt es etwa sehr plausibel: „The narrators presence derives from the audience’s sense of some demonstrable communication. If it feels being told something, it presumes a teller“ (Chatman 1978, S. 147).

Film, nicht in der Literatur – überhaupt kein „someone“ sein. Der *narrator* ist tatsächlich ein Deckname für ein *Etwas* – ein Etwas, das aus Komponenten besteht.

Dies wird an der Überlegung deutlich, dass die „different components of the cinematic narrator“ normalerweise „in consort“ zusammenwirken, es aber nicht müssen (Chatman 1990, S. 135). Es gebe nämlich bisweilen eine „ironic tension between two of them“ (ebd.). Man fragt sich, wie es möglich ist, dass zwischen verschiedenen Komponenten eine ironische Spannung entsteht. Gemeint ist damit nicht, dass etwa der Schnitt eines Films in einem Spannungsverhältnis zur Beleuchtung stehen kann, oder die Kameraführung zum Aussehen der Akteure usw. Vielmehr komme es sogar im Hollywoodfilm vor, dass „the visual track“ die Geschichte untergrabe, die von einem „character narrator’s voice over“ erzählt werde (Chatman 1990, S. 136).⁸ Hier befinden wir uns also an dem Punkt, an dem sich der Status des *voice over* aus filmnarratologischer Perspektive als strukturell bedeutsam erweist. Zuvor hat Chatman betont, dass der *narrator* eines Films keineswegs mit dem *voice-over-narrator* verwechselt werden dürfe, da dieser eben nur „one component of the total showing“ (Chatman 1990, S. 134) sei. Aus systematischen Gründen kann die hörbare Erzählerstimme eines Films für Chatman niemals mit dem *cinematic narrator* zusammenfallen, allenfalls kann sie eine Komponente darstellen, die mit den übrigen Komponenten zusammenstimmt.

Theoretisch lässt sich dann aber nicht rechtfertigen, warum sie die *einzig*e Komponente darstellt, die sich den übrigen Komponenten entgegensetzen kann. In der Regel sei es so, dass der *voice-over*-Erzähler durch das, was im „visual track“ gezeigt werde, Lügen gestraft werde, nicht aber umgekehrt. Dass es überhaupt eine solche „partial unreliability“ geben könne, sei der Besonderheit von „two-track media such as the cinema“ (Chatman 1990, S. 136) geschuldet. Daraus müsste dann aber gefolgert werden, dass der *cinematic narra-*

⁸ In dem kurz zuvor erschienenen und von Chatman bereits rezipierten Buch *Invisible Storytellers* von Sarah Kozloff (dem einschlägigen Buch zum Phänomen *voice over*) gibt es ein ganzes Kapitel über „Irony in Voice-Over Films“ (vgl. Kozloff 1988, S. 102-127).

tor keineswegs einfach aus verschiedenen Komponenten besteht, sondern dass er eine *interne Struktur* aufweist. Es gäbe nämlich die Komponenten des *visual track* bzw. *channel* und die des *auditory track* bzw. *channel*, die einander irgendwie widersprechen können – wobei Chatman im Prinzip eine hierarchische Beziehung zwischen den beiden unterstellt: Der *auditory channel* ist dem *visual channel* untergeordnet (was aber eine empirische Feststellung zu sein scheint⁹).

Die systematische (logische oder medientechnische) Unterscheidung von *auditory channel* und *visual channel*, die hier durchscheint, wird von Chatman nicht theoretisch reflektiert. Markus Kuhn setzt an diesem Punkt an, ohne allerdings zu sehen, welche theoretischen Probleme er sich damit einhandelt. Ihm zufolge müsste Chatman lediglich die Unterscheidung zwischen „telling“ und „showing“ für die interne Strukturierung des *cinematic narrator* fruchtbar machen. Kuhn stellt fest, dass es bei Chatman, was die Stellung des *voice-over*-Erzählers angeht, eine systematische Unklarheit gibt. Sie komme zum Ausdruck wenn Chatman einräumt, dass der *voice-over*-Erzähler bisweilen den *visual track* zu dominieren scheint (Chatman nennt *Le Journal d'un Curé de Campagne* (1964) von Robert Bresson). Diese Dominanz drückt aber keine Hierarchie aus, wie Kuhn meint, sondern bringt nur eine bestimmte Art des Zusammenspiels der Komponenten zum Ausdruck. Kuhn möchte die „auf den Kopf gestellte“ Hierarchie“ als „Argument“ für die von ihm vorgeschlagene „analytische Trennung zweier (oder mehrerer) nicht zwangsläufig hierarchisch anzunehmender visueller oder sprachlicher narrativer Instanzen“ (Kuhn 2011, S. 80) nehmen. Voraussetzung für den gesamten Ansatz von Kuhn ist eine unabhängige Betrachtung zweier narrativer Instanzen im Film. Die spezifische Form der *Asymmetrie* – und der Verzahnung – dieser „Instanzen“ muss dabei jedoch aus dem Blick geraten. Darüber hinaus stellt sich die Frage, wie diese beiden Instanzen überhaupt definiert sind. Denn mit der Unter-

⁹ Chatman erwähnt, dass Sarah Kozloff *einen* Film gefunden habe, in dem das *voice over* den *visual track* gleichsam korrigiert: die Eröffnungssequenzen in *Ein Amerikaner in Paris* (1951) von Vicente Minelli; hier ist die Korrektur aber Teil eines metaleptischen Spiels.

scheidung zwischen *visual* und *auditory track* deckt sich die Unterscheidung von *telling* und *showing* natürlich nicht (Chatman verbucht den *voice-over-narrator* ja sogar explizit als „one component of the total *showing*“).

II. Die sprachliche und die visuelle Erzählinstanz

Kuhn modelliert die narrativen Instanzen im Film folgendermaßen: Der „*implizite Autor*, der selbst über keine semiotischen Zeichensysteme ‚verfügt‘“, setzt „eine *visuelle Erzählinstanz* und eine oder mehrere (oder auch keine) *sprachliche Erzählinstanz(en)* auf extradiegetischer Ebene ein, um ‚filmisch erzählen zu lassen“ (Kuhn 2011, S. 85) Während von dieser Seite aus betrachtet der implizite Autor als derjenige *konzipiert* ist, der andere Instanzen ‚einsetzt‘, kann man ohne diese problematische Kategorie auskommen, wenn man sagt, dass die sprachliche und die visuelle Erzählinstanz gleichsam automatisch entstehen und zusammenwirken. Daher die alternative Formulierung:

Der Prozess des filmischen Erzählens entsteht im Zusammenspiel einer visuellen Erzählinstanz, die durch audiovisuelles Zeigen bzw. Vorführen von Szenen erzählt, mit einer oder mehreren (oder auch keiner) sprachlichen Erzählinstanz(en), die wortsprachlich erzählen und der visuellen Erzählinstanz untergeordnet sein können, aber nicht müssen. (Ebd.)

Von der visuellen Erzählinstanz (VEI) heißt es, sie vermittele durch Vorzeigen „von aneinandergereihten Einstellungen und ihren Einstellungen und ihren Relationen zueinander [...] die Ereignisse und Handlungen der diegetischen Welt“ (Kuhn 2011, S. 92). Man müsste also sagen, dass die VEI im Grunde mit der Diegese zusammenfällt (wobei natürlich innerhalb der Diegese sprachliche Erzählinstanzen zweiter Ebene vorkommen können). Bezeichnenderweise spricht Kuhn das aber nicht wirklich aus. Das liegt unter anderem daran, dass dann die sprachliche Erzählinstanz als extradiegetisch klassifiziert wäre. Und das ist problematisch.

Die VEI ist Kuhn zufolge im Regelfall die „logisch privilegierte Vermittlungsinstanz“ (Kuhn 2011, S. 92). Das liegt zum einen daran, dass ohne sie kein Film gezeigt würde. Zum anderen entspricht die

VEI dem, was in der auf Literatur bezogenen Erzähltheorie die extra-heterodiegetische Erzählinstanz heißt, deren Behauptungen „im Rahmen der erzählten Welt, nicht nur wahr, sondern notwendig wahr“ (Martinez/Scheffel 2005, S. 96) sind. Das heißt, sie kann im Regelfall nicht „*mimetisch teilweise unzuverlässiges Erzählen*“ (ebd., S. 102) sein. Kuhn fasst zusammen: „Im Film ist das logische Prinzip im Normalfall für die visuelle Erzählinstanz anzunehmen, weil diese in den meisten Fällen *extra-* und *heterodiegetisch*, also nicht an die diegetische Welt gebunden ist.“ (Kuhn 2011, S. 94) Was besagt aber diese Begründung? Warum soll das, was nicht an die ‚diegetische Welt‘ gebunden ist, logisch privilegiert sein? Der Grund kann nur darin liegen, dass die VEI nicht als Person, sondern eben nur als *Instanz* aufzufassen ist. Und eine *Instanz* kann keinen *Grund* haben, mimetisch unzuverlässig zu erzählen. Dort aber, wo es keinen Grund gibt, kann der Rezipient auch keinen Grund finden außer dem, mimetisch unzuverlässig erzählen zu *wollen*.¹⁰ Anders verhält es sich nur, wenn die Unzuverlässigkeit jemandem zugerechnet werden kann (wenn sie interpretiert werden kann).

Kuhn macht seine Bemerkungen zum logisch privilegierten Status der VEI im Hinblick auf „einige Formen des Zusammenspiels von lügenden sprachlichen Erzählinstanzen mit sich verschieden dazu verhaltenden visuellen Erzählinstanzen“ (Kuhn 2011, S. 94). Hieraus kann man grundlegende Schlüsse für das Verhältnis von VEI und SEI ziehen (was Kuhn aber nicht tut): Die sprachliche Erzählinstanz kann *lügen*, die audiovisuelle Erzählinstanz kann es nicht. Man kann nur mit Worten lügen. Man kann nicht lügen, indem man *zeigt*. Das Gezeigte kann nicht Lüge, es kann nur Trug sein. Wenn Bilder lügen können sollen, muss ihnen eine SEI dabei helfen.

Von einer sprachlichen Erzählinstanz kann man laut Kuhn immer dann sprechen, wenn auf irgendeine Weise sprachlich mindestens eine Minimalgeschichte erzählt wird“ (Kuhn 2011, S. 96). Da das nicht der Fall sein muss, nennt Kuhn diese SEI fakultativ. Aber ab

¹⁰ Auf die weiteren Probleme, die hier involviert sind, kann an dieser Stelle nicht eingegangen werden. Die neuere und neueste Literatur zum Thema ‚unzuverlässiges Erzählen‘ ist reichhaltig; vgl. etwa das Themenheft 1/2011 des *Journal of Literary Theory*.

wann ist es der Fall? Was heißt hier ‚Minimalgeschichte‘? Genügt es schon, wenn am Anfang ein einziges Mal ein *voice over* zu hören ist (wie etwa in *No Country for Old Men* (2007) oder *True Grit* (2011) von den Coen-Brüdern)? Genügt es, wenn diese Stimme nur *ankündigt*, eine Geschichte erzählen zu wollen? Die SEI muss überdies nicht „phonetisch/auditiv“ sein. Auch Zwischentitel und Schrifttafeln können eine SEI bilden, die dann „graphemisch/visuell“ ist (ebd.). In dieser Hinsicht ist noch schwieriger, zu entscheiden, ab wann man von einer sprachlichen Erzählinstanz sprechen soll, denn schließlich erfüllen Zwischentitel zunächst einmal eine paratextuelle Funktion. Soll man von einer SEI sprechen, wenn mittels Zwischentiteln die Akte durchgezählt, in Kapitel eingeteilt, klärende Orts- und Zeitangaben gemacht werden?¹¹

In Filmen kann es eine ganze Reihe von Elementen geben, die zweifellos nicht zur Diegese gehören und gleichwohl auch nicht unter die Kategorie einer sprachlichen Erzählinstanz gebracht werden können. Kuhn selbst räumt in einer Fußnote ein, dass von dem, was der „cinematic narrator“ im Sinne Chatmans auf dem auditiven Ka-

¹¹ In anderem Zusammenhang erklärt Kuhn, dass Zeitangaben wie „Rom im Jahre 50 nach Christus“ oder „New York, Oktober 2166“ in Filmen nicht als eigenständige SEI analysiert werden müssten, „weil sie keine Minimalgeschichte repräsentieren und eindeutig der VEI untergeordnet sind“. Darüber hinaus betreffen solche Markierungen nur die *histoire* und hätten „keinen Aussagewert bezüglich des Verhältnisses der *narration* zur *histoire*“ (Kuhn 2011, S. 244). Das zweite Argument ist interessant, insofern es die Frage aufwirft, ob es sich mit Angaben wie „Drei Wochen später“ anders verhält. Das erste Argument legt die Problematik der Kategorie ‚Unterordnung‘ noch einmal frei. Inwiefern kann man sagen, dass solche Zeitangaben der VEI untergeordnet sind, da sie doch offensichtlich von einer logisch privilegierten Instanz aus erfolgen, deren Angaben nicht sinnvoll bezweifelt werden können? Die Begründung, die Angabe sage doch nur, was man sieht, ist hier unzureichend, denn offensichtlich wird es für wichtig befunden, noch einmal durch eine *Instanz* zu *bestätigen*, was man *sieht*. Gerade in der Verdopplung tritt der Status der Instanz zutage. Eine VEI, die *nicht* gerahmt ist, lässt sich im übrigen überhaupt nur mittels Markierung durch eine sprachliche Instanz als ‚spätere Narration‘ (im Sinne Genettes) ausweisen. Eine VEI kann sich, mit anderen Worten, nicht selbst *verorten*.

nal alles zu bieten hat, „hier nur der Aspekt der Stimme“ untersucht werde. Wenn der „Analyseschwerpunkt auf der narrativen Funktion der außersprachlichen Tonebene und dem Verhältnis derselben zur visuellen Ebene“ liegen solle, so sei es möglicherweise sinnvoll, „neben einer *visuellen* und einer fakultativ *sprachlichen* Erzählinstanz eine *auditive Erzählinstanz (AEI)* anzunehmen, der alle außersprachlichen Tonelemente zugeschrieben werden“ (Kuhn 2011, S. 94f.). Das klingt nun so, als könne man die Instanzen so vermehren, wie man es gerade für die Analyse braucht. Es ist ja in diesem Zusatz nicht einmal klar, ob diese „AEI“ nur die außersprachlichen extradiegetischen Tonelemente umfassen würde (wie vor allem die Filmmusik) oder die Tonspur im Allgemeinen (womit es dann eine große Schnittmenge mit der VEI gäbe).

Es ist im Gegenteil gerade der entscheidende Punkt des Modells von Kuhn, dass sich die Instanzen nicht vermehren dürfen. Zwar kann sich die SEI selbst verdoppeln (weil in einem Film mehr als ein *voice-over*-Erzähler vorkommen kann wie in *All About Eve* (1950) von Joseph Mankiewicz), und auch die VEI kann sich aufspalten (durch einen *split screen* wie etwa in *Timecode* von Mike Figgis). Aber die Unterscheidung selbst beruft sich ja auf die Opposition von *showing* und *telling*, die vorgibt, kein Drittes zuzulassen. Und sie bezieht sich auf die Opposition *intradiegetisch* vs. *extradiegetisch*, die ebenfalls kein Drittes zuzulassen scheint. Aber diese beiden Oppositionen sind weder kongruent noch lassen sie sich mit der Unterscheidung zwischen SEI und VEI zur Deckung bringen. Insofern muss jede Definition der VEI und der SEI prekär bleiben. Dessen eingedenk, kann man gleichwohl sehr gut mit dieser Unterscheidung arbeiten.

III. „Zusammenspiel“

Wenn man analytisch zwischen audiovisueller Erzählinstanz (VEI) und sprachlicher Erzählinstanz (SEI) unterscheidet, ohne auf den unterschiedlichen Status dieser Instanzen einzugehen, stellt sich deren Verhältnis vorab als formal symmetrisch dar, als ein „Zusammenspiel“ (Kuhn 2011, S. 97). Kuhn setzt sich zunächst mit Sarah Kozloff auseinander, die drei grundsätzliche Fälle unterscheidet: ein *Überlappen* von „narration“ und „images“, ein *komplementäres Ver-*

hältnis zwischen ihnen, und ein disparates Verhältnis, wobei es keine klaren Grenzziehungen gebe und das Verhältnis auch keineswegs statisch sein müsse (Kozloff 1988, S. 103). Kuhn differenziert weiter mit Hilfe der schematischen Abstufung *widersprüchlich – verschieden – sich ergänzend – verzahnt – polarisierend – illustrierend/umschreibend – paraphrasierend* (Kuhn 2011, S. 99). In jedem Fall könnten sowohl die SEI wie auch die VEI dominant sein oder es könne Gleichberechtigung herrschen (vgl. Kuhn 2011, S. 100). So könne beim überlappenden Zusammenspiel die VEI illustrieren, was die dominierende SEI erzählt oder die SEI umschreiben, was die dominierende VEI zeigt (ebd.).

Ob man eine solche formale Symmetrie postulieren kann, müsste sich natürlich beim ‚widersprüchlichen‘ Verhältnis zwischen SEI und VEI zeigen, womit man wieder bei der Frage nach der Unzuverlässigkeit angelangt wäre. Als Beispiel für die Dominanz der VEI nennt Kuhn den Film *All about Eve* von Joseph Mankiewicz, wo eine „zuverlässige extradiegetische VEI [...] die unzuverlässige extrahodiegetische SEI“ entlarve (Kuhn 2011, S. 97). Diese Entlarvung (wenn man sie so nennen will) findet in diesem Film vor allem deshalb statt, weil es *mehr als eine* SEI gibt und weil die SEI keine unparteiischen Beteiligten des Geschehens sind. Das, was die SEI sagt, steht dabei nicht in einem tatsächlichen Widerspruch zum Gezeigten. Es wird nur lesbar als eine tendenziöse Interpretation. Mimetische Unzuverlässigkeit kann man den *voice-over*-Erzählern Addison DeWitt, Karen und Margo in diesem Film nicht nachsagen (vgl. Kozloff 1988, S. 62ff.).

Mimetische Unzuverlässigkeit einer SEI würde voraussetzen, dass die SEI entweder nicht weiß, was die VEI zeigt oder dass sie sich wissentlich in Widerspruch zur VEI setzt (ohne aber Einspruch gegen die VEI zu erheben). Im Fall von *All about Eve* wird zu Anfang die Herrschaft der *voice-over*-Stimme von Addison DeWitt über das Gezeigte geradezu zelebriert: Er ist es, der uns erklärt, was wir sehen, und der sich uns zugleich als ‚der Zeigende‘ vorstellt. Zu den Asymmetrien zwischen der SEI und der VEI gehört es, dass das homodiegetische *voice over* sich in einer Position befinden kann, in der es auf die VEI zeigt (also auf das Gezeigte zeigt), während die VEI die SEI allenfalls Lügen strafen kann, *ohne* auf sie zu zeigen. Wenn

uns etwas gezeigt wird und wir dabei eine extradiegetische Stimme hören, so gehen wir vielleicht nicht unbedingt davon aus, dass es die Stimme ist, die uns zeigt (dann wäre die VEI keine unabhängige Instanz mehr), wohl aber davon, dass die Stimme darüber Bescheid weiß, was uns gezeigt wird.¹² Dies ist sozusagen der unmarkierte Fall: Wenn es anders sein soll (wenn der Film unter einer anderen Voraussetzung gesehen werden soll), so muss uns das signalisiert werden. Es macht also, solange es einen *Bezug* der SEI auf die VEI gibt, keinen Sinn, wenn man ein Erzählen der SEI annimmt, das von der VEI als mimetisch unzuverlässig enthüllt wird. Unsere Unterstellung, dass die SEI darüber Bescheid weiß, was uns die VEI zeigt, impliziert zugleich die Vorstellung, dass es die SEI ist, die ein sich ergänzendes, verzahnt, polarisierendes, illustrierendes, umschreibendes oder paraphrasierendes Verhältnis zwischen ihr und der VEI *zulässt*. Im gewöhnlichen Sprechen über Filme drückt sich das zum Beispiel in der Redewendung aus, dass sich der *voice-over*-Erzähler ‚zurückzieht‘. Das heißt ja nichts anderes, als dass er aus freien Stücken schweigt – und anders können wir uns sprachliche Handlungen nicht denken. Brian Henderson hat umgekehrt behauptet, die Erzählstimme im Film sei eine „puppet of the narration“: „The puppet narrators of cinema are jerked on and off stage in a manner that is quite undignified“. (Henderson 1983, S. 16) Das mag eine richtige Beobachtung für die *Funktion* der Erzählstimme im konventionellen Erzählkino sein, aber im Hinblick auf die Frage nach der *Instanz* im Film verhält es sich genau umgekehrt: Es ist zunächst einmal die SEI, die zurücktritt und der Diegese den Vortritt *lässt*.¹³ Empirisch gesehen *muss* sich die SEI freilich an bestimmten

¹² Wenn man versucht, das Verhältnis zwischen *voice-over*-Erzähler und dem audiovisuell Gezeigten auf dieser Ebene zu bestimmen, drängt sich die Feststellung, dass die Stimme die Bilder nicht produziert, förmlich auf: *voice-over*-narrators „are invariably swallowed up in the overall narrational process of the film, which they do *not* produce“ (Bordwell 1985, S. 61). Sie ist aber genau genommen sinnlos, denn wie sollte eine Stimme tatsächlich etwas ‚produzieren‘ können?

¹³ Auf eine analoge Depotenzierung läuft es hinaus, wenn man aus der zweifellos richtigen Beobachtung, dass die *voice-over*-Stimme ‚weggelassen‘ werden kann, die falschen Schlüsse zieht – wenn man etwa

Punkten (etwa am Ende oder an bestimmten Nahtstellen) mit diesem Anspruch in Erinnerung rufen.

Dass eine extra-*homodiegetische* SEI durch die VEI ‚entlarvt‘ werden kann, steht außer Frage, weil die SEI, die einer Figur zugeordnet werden kann, stets eine *Deutung* der Diegese vorträgt, die sich tendenziell als tendenziös oder verzerrend auffassen lässt. Eine mimetische Unzuverlässigkeit der SEI, also einen expliziten Widerspruch zwischen dem, was gezeigt wird, und dem, was von der SEI gesagt wird, scheint es in der Logik des Erzählkinos jedoch nur unter bestimmten Umständen geben zu können. Dass uns ein extra-*homodiegetischer voice-over*-Erzähler mitteilt, an jenem Abend habe er eine Krawatte getragen, während wir zugleich gezeigt bekommen, dass die betreffende Figur an diesem Abend eine Fliege getragen hat, ist durchaus möglich. Durch den expliziten Widerspruch bekommen wir signalisiert, dass sich der Erzähler falsch erinnert. Und zugleich ist damit ein bestimmtes Verhältnis zwischen SEI und VEI instituiert: Das was die VEI zeigt, sind dann *nicht* die Erinnerungen der SEI, wie es in den konventionellen Rückblenden der Fall ist. Der Zuschauer kann in diesem Fall einen Grund dafür benennen, wie es zu diesem Widerspruch kommt. Bei einem extra-*beterodiegetischen Voice over*-Erzähler können wir einen solchen Grund prinzipiell *nicht* namhaft machen, weil dieser nur als Instanz und nicht als Subjekt spricht.

Kann auch umgekehrt die VEI von einer SEI als falsch entlarvt werden? Kuhn weicht dieser Frage in seinen Überlegungen zum „Zusammenspiel“ der VEI und der SEI aus, obwohl die Antwort aus der Statuierung des ‚logisch privilegierten Status‘ der extradiegetischen VEI folgt. Er erklärt, dass „auch die extradiegetische VEI [...] unzu-

sagt, dass der *voice-over*-Erzähler „im Film nie als unhintergehbare Vermittlungsinstanz fungiert, sondern lediglich als eine – kommunikationstheoretisch weglassbare – Stimme neben den Figurenstimmen einsetzbar ist“ (Mahler 2001, S. 264f.). Der Umstand, dass die Erzählerstimme in ihrem Auftauchen einen bestimmten Status implizieren muss, wird hier außer Acht gelassen. Das *voice over* „als sprachliches Phänomen“ ist daher, insofern sie *erzählt*, keineswegs „einfach eine unter vielen Möglichkeiten der filmischen Kommunikation“ (Schäfer 2008, S. 28).

verlässig sein“ (Kuhn 2011, S. 97) sein könne und verweist auf das Beispiel *Fight Club* (1999) von David Fincher und *Stage Fright* (1950) von Alfred Hitchcock. In *Fight Club* erweist sich die VEI zwar als unzuverlässig (man sieht Tyler Durden, obwohl er nur in der Vorstellung des Protagonisten existiert), aber dies wird nicht durch die SEI entlarvt, sondern zeigt sich innerhalb der Diegese.¹⁴ *Stage Fright* wird enthält eine berühmte (von Hitchcock später bedauerte) ‚lügende Rückblende‘, die ebensowenig durch eine SEI entlarvt wird. Als Beleg für die Mitwirkung sprachlicher Erzählinstanzen an der Überführung der VEI liefert Kuhn ersatzweise den kanonischen Film *Rashomon* (1950) von Akira Kurosawa – ein in mehrfacher Hinsicht ungeeignetes Beispiel. Denn dieser Film lässt Aussagen von Figuren sehen, die sich widersprechen (Niehaus 2006). Weil die Rückblenden in *Rashomon* als visualisierte Aussagen mehrerer intradiegetischer Erzähler zu betrachten sind, findet deren Entwertung gerade nicht durch die sprachlichen Erzählinstanzen statt, sondern dadurch, dass in den metadiegetisch visualisierten Erzählungen der Beteiligten nur der Ausgang der Geschichte übereinstimmt, sie ansonsten aber nicht miteinander vereinbar sind.

Eine visualisierte Erzählung von einer Figur innerhalb der VEI (also einer intradiegetischen SEI) klassifiziert Kuhn zu Recht als extradiegetisch, weil die Visualisierung *nicht* Teil der erzählten Welt ist; Kuhn prägt in diesem Zusammenhang die schöne Wendung vom „latenten visuellen Ebenenkurzschluss“ (Kuhn 2011, S. 281 u. 310ff.). Wenn sie aber extradiegetisch ist, hat sie aber auch einen logisch privilegierten Status. Deswegen ist nicht die extradiegetische VEI unzuverlässig, sondern die intradiegetische SEI. Die extradiegetische VEI darf mit der SEI aber nur ‚gemeinsame Sache‘ machen, weil mehrere Versionen desselben Geschehens gezeigt werden; wird nur eine einzige gezeigt (wie in *Stage Fright*), wird sie zur lügenden Rückblende.

Kuhn liefert für diesen Fall also nur ungenügende Beispiele. Und das liegt schlicht und ergreifend daran, dass die VEI im Film nicht durch eine SEI als unzuverlässig vorgeführt werden kann. Ein derartiges Zusammenspiel von SEI und VEI ist ausgeschlossen. Warum das so

¹⁴ Kuhn spricht hier von einer „mentale[n] Metalepse“ (Kuhn 2011, S. 273).

ist, ließe sich nur innerhalb einer Theorie der Instanzen erklären. Eine VEI kann eine SEI *entkräften*, eine SEI kann einer VEI hingegen lediglich *widersprechen*. Aber nur dann, wenn die SEI wirklich *erzählt*.

IV. *Voice over* – Erzählen, Zeigen?

Zur Erinnerung: Die SEI unterscheidet sich von der VEI nicht dadurch, dass man sie hört, während man die VEI sieht. Die VEI ist – im Tonfilm – audiovisuell, man kann sie auch hören; und die SEI braucht man nicht zu hören, da sie auch durch Zwischentitel und Schrifttafeln operieren kann. Der Unterschied zwischen SEI und VEI besteht auch nicht einfach darin, dass die VEI die Diegese definiert, während die SEI extradiegetisch ist. Nach dem ‚visuellen Ebenenkurzschluss‘ sind alle Rückblenden extradiegetische VEI; und die SEI realisiert sich zwar häufig über ein *voice over*, das definitionsgemäß nicht zur Diegese gehört, aber daraus folgt nicht, dass alles, was nicht zur Diegese gehört, der SEI zuzuschlagen wäre, wie etwa extradiegetische Filmmusik deutlich macht. Die SEI soll „vergleichbar mit Genettes Kategorie der Stimme“, die – wenn „phonetisch/auditiv“ – technisch realisiert wird als *voice over*, oder – bei Erzählen auf zweiter Stufe als *voice off* oder durch Erzählen im Figuraldialog (Kuhn 2011, S. 95). Das strukturelle Problem dieser Kategorien resultiert daraus, dass es nur durch die Basisunterscheidung zwischen Erzählen (*telling*) und Zeigen (*showing*) gerechtfertigt ist: Die SEI erzählt, während die VEI zeigt (Kuhn 2011, S. 86). Verschiedene filmische Phänomene, die Kuhn keineswegs verschweigt, bringen diese Zuordnung durcheinander. Unter anderem trifft es nicht zu, dass jedes *voice over* erzählt. Aber was tut es dann?

Erzählen ist zunächst einmal ein *sprachliches* Verhalten, das – anders als filmisches Erzählen – in der Alltagswelt vorkommt und jedem jederzeit zur Verfügung steht.¹⁵ Weil das so ist, können zum Beispiel

¹⁵ Kuhn wendet sich gegen die Narrativitätskonzeption etwa von Werner Wolf, nach dem bestimmte Signale das Narrative als kognitives Schema aktivieren. Gegen dieser seiner Vorannahme entgegen Wolf im Ergebnis doch dem „dominierenden Sprachbezug der Narratologie“ nicht und

in Filmen Figuren vorkommen, die erzählen (was immer man auch als Kriterium für eine Minimalerzählung ansetzt). Gerade der jederzeit mögliche fließende Übergang zum Erzählen ist das entscheidende Argument dafür, dass Erzählen zunächst einmal als ein sprachliches Verhalten unter anderen betrachtet werden muss. Wenn eine Figur gezeigt wird, die spricht, ist sie ein Teil der VEI. Wenn sie zu erzählen beginnt, wird sie der Theorie nach zu einer intradiegetischen SEI. Aber sie wird gleichzeitig auch gezeigt, bleibt also weiterhin Bestandteil der VEI. Das ist trivial, macht aber deutlich, dass es kein *telling* ohne *showing* gibt. In der Literaturwissenschaft haben Narratologen, die dem Erzähler einen großen Stellenwert in ihrer Theorie einräumen und für jede Erzählung einen Erzähler postulieren, darauf beharrt, dass aus jedem Erzählen Rückschlüsse auf den Erzähler gezogen werden können.¹⁶ Auf die Disjunktion von Erzählen und Zeigen übertragen hieße das, dass jede Erzählung bis zu einem bestimmten Grad auch ihren Erzähler zeigt. Bei der literarischen Erzählung kann man sich auf den Standpunkt stellen, dass eine solche Auffassung kontraintuitiv und nutzlos ist, da es im Standardfall nicht der Rezeptionsweise des Lesers entspricht, seine Aufmerksamkeit auf einen nahezu vollständig abwesenden und nicht wahrnehmbaren Erzähler zu richten¹⁷, im Film verhält sich die Sache aber anders, denn hier ist die SEI, sobald sie nicht auf Schrifttafeln beschränkt ist, tatsächlich wahrnehmbar. Es wird uns in einem übertragenen Sinne das akustische Bild einer ganz unmetaphorischen Stimme ‚gezeigt‘, aus dem wir das Geschlecht, das Alter, die

postuliere einen „verbalen Prototyp[]“ des Erzählens (Kuhn 2011, S. 54; vgl. Wolf 2002). Wolfs Position wäre entgegenzuhalten, dass sich die kognitiven Schemata nur ausbilden können, weil Erzählen ein grundlegendes sprachliches Verhalten ist, während Kuhn den unnötigen Schluss zieht, der dominierende Sprachbezug der Narratologie hänge damit zusammen, dass Erzählen zunächst einmal ein grundlegendes sprachliches Verhalten sei. Erzählen durch Zeigen ist ein fundamental *artifizielles* Erzählen.

¹⁶ So etwa Wolf Schmid in den *Elementen der Narratologie* (Schmid 2005, S. 72ff.).

¹⁷ Dietrich Weber spricht davon, dass der Standardtyp des Erzähltextes autor- und erzählerverleugnend sei (vgl. Weber 1998, S. 90ff.).

Gefühlsverfassung und anderes mehr entnehmen können. Man kann argumentieren, dass diese Stimme, weil sie nicht zur Diegese gehört, narratologisch nicht relevant ist; wenn man aber – wie Kuhn – davon ausgeht, dass alle filmischen Mittel zur spezifischen Erzählweise eines Films beitragen, so liegt bereits hier ein blinder Fleck der narratologischen Beschreibung.

Dies ist aber nur ein Teil der Probleme, die aus der Identifizierung der extradiegetischen SEI mit dem Erzählen erwachsen. Gerade das Phänomen der Stimme zeigt, dass der Begriff der Diegese selbst nicht so trennscharf ist, wie er zu sein scheint.¹⁸ Wenn eine Figur mit sich allein ist und der Zuschauer Zeuge eines Selbstgesprächs wird, das sie führt, dann ist dieser Sprechakt gewiss Teil der Diegese. Wenn wir die Stimme der Figur hören, ihre Lippen sich aber sichtbar nicht bewegen, so handelt es sich hingegen nach den eingeführten Kategorien um ein extradiegetisches *voice over* (auch wenn die Stimme dasselbe spricht). Wenn schließlich die Kamera einer Figur über die Schulter sieht, während sie schreibt, und ihre Stimme das Schreiben begleitet, so wissen wir nicht, ob die Stimme Teil der Diegese ist oder nicht; für diesen Fall reserviert Kuhn den Terminus „Pseudo-voice-over“ (Kuhn 2011, S. 191). In Eric Rohmers *La Marquise d'O* (1976) kommen alle drei Formen hintereinander vor (vgl. Kuhn 2011, S. 274).

Das legt zum einen nahe, dass die Rezeption eines Films keine deutliche Unterscheidung zwischen intradiegetisch und extradiegetisch erfordert. Zum anderen bleibt unklar, welchen Status der *voice-over*-Monolog dann eigentlich hat. In einem literarischen Text würde es einem Erzähltheoretiker nicht einfallen, Gedanken einer Figur als extradiegetisch zu bezeichnen, weil sie in der erzählten Welt nicht wahrnehmbar sind. Wenn man sich dieses Kriterium zu eigen machte, müsste man auch die vernebelte subjektive Einstellung eines Be-

¹⁸ Natürlich ist der – von Kuhn nicht weiter erörterte – Begriff der Diegese umstritten und kann verschieden gefasst werden (vgl. etwa Hartmann 2007). Nicht darauf kommt es hier an, sondern auf die Konsequenzen des Versuchs, überhaupt mit festen Grenzen zwischen intra- und extradiegetisch zu operieren, für die Frage nach der Lokalisierung der Stimme.

trunkenen oder das auf der Tonspur hörbare Geräusch eines aufgeregt schlagenden Herzens als extradiegetisch bezeichnen. Vor allem jedoch ist der Monolog, wenn er lippensynchron ist, eine sprachliche Äußerung innerhalb der VEI, wird aber, wenn er nicht lippensynchron ist, aufgrund der kategorialen Mechanik zu einem Element der SEI. Das gilt für alle Formen von inneren Stimmen: „Bei der temporären *inneren Stimme* [...] handelt es sich um eine extrahomodiegetische SEI, die intern auf das *erzählte Ich* fokalisiert, das dem zeitgleich von der VEI *gezeigten Ich* (der szenischen Figur) entspricht.“ (Kuhn 2011, S. 271f.) Diese Kategorisierung widerspricht der literaturwissenschaftlichen Einordnung, die hier ja gerade nicht von Erzähler-, sondern von Figurenrede sprechen würde – die Unterscheidung von *telling* und *showing* wurde ja gerade aus dieser platonischen Differenzierung abgeleitet. Kuhn hingegen muss daraus, dass er die innere Stimme als extradiegetisch klassifiziert, schließen, dass sie auch *erzählen* muss. Tatsächlich findet dort, wo im Film eine innere Stimme spricht, weder Erzählen statt, noch spricht eine *Instanz*.

Bei dem, was Kuhn als SEI definiert, wird nicht nur erzählt, sondern auch gezeigt. Wenn wir die innere Stimme einer Figur hören, wird uns nicht erzählt, was sie denkt, sondern es wird uns gezeigt – entsprechend könnte man durchaus dafür argumentieren, dass der Zuschauer beim „Diegetisieren“ (Hartmann 2007, S. 56f.) diese innere Stimme in die dargestellte Welt integriert. Ein anderes Beispiel. Kuhn kommt selbst in einem kurzen Abschnitt auf sogenannte „Voice-over-Dialoge und Voice-over-Szenen“ zu sprechen (Kuhn 2011, S. 164). In *Pierrot le Fou* (1965) von Jean-Luc Godard etwa gibt es einen nicht ans Bild gekoppelten Dialog zweier *voice-over*-Stimmen. Im Prinzip kann der Film eine asynchrone auditive Ebene einziehen, „vergleichbar mit akustischen Ereignissen und Räumen des Hörspiels“ (ebd.). Eine solche Nutzung der Tonspur kann offenbar weder der VEI noch der SEI zugeschlagen werden, worüber Kuhn jedoch kein Wort verliert. In dem von Kuhn angeführten als Beispiel Film *Kleinruppin forever* (2004) wird zu Beginn des Films ein Säuglingspaar von den Eltern versehentlich auf einer Autobahnraststätte stehen gelassen. Während das Bild bei den schlafenden Babys bleibt, folgt die Tonspur dem sich streitenden Ehepaar, das im

Auto schließlich einen tödlichen Unfall erleidet. Ähnlich wie bei einem *split screen* zeigen Tonspur und Bild hier zwei von einander abgekoppelte *Szenen* (wobei die akustische Szene das Ereignis ‚zeigt‘, die visuelle Szene kein Ereignis, sondern diejenigen, für die das Ereignis von Bedeutung sein wird).¹⁹ Niemand erzählt etwas – es gibt keine SEI. Der berühmte letzte Film *Blue* (1993) von Derek Jarman zeigt von Anfang bis Ende einen unverändert blauen Schirm. Alles, was geschieht, geschieht auf der Tonspur – ein Geflecht aus Szenen, Klängen, Reflexionen, Musik. In einem solchen Film hätte die Unterscheidung zwischen Erzählen und Zeigen (die man wohl anwenden kann, auch wenn sie unscharf ist) mit derjenigen zwischen SEI und VEI nichts mehr zu tun.

Man könnte einwenden, dass es sich um Ausnahmen handelt – um einen originellen filmischen Einfall für den Beginn eines Films im Falle von *Kleinruppin forever*, um ein existenzielles Experiment im Fall von *Blue*. Aber darum geht es nicht. Die Ausnahmen machen nur manifest, dass die Filmnarratologie aufgrund ihres Erzählbegriffs blind bleiben muss für alle jene Manifestationen der Stimme, die weder Erzählen sind noch Zeigen, und weder Erzähltes noch Gezeigtes. Und – damit zusammenhängend – bleibt auch der Begriff der Instanz im Dunkeln. Es handelt sich hierbei nicht um ein einfach zu behebendes Problem des erzähltheoretischen Analysemodells, sondern um ein systematisches Defizit, das in den narratologischen Kategorien begründet ist.

¹⁹ Es gäbe in diesem Fall also zwei simultane VEI, von denen eine ein akustisches Bild, die andere hingegen ein visuelles Bild lieferte. Entsprechend können sich die VEI beim *split screen* vermehren (müssen es aber nicht), worauf Kuhn aber in seinen Ausführungen zum *split screen* nicht zu sprechen kommt (vgl. Kuhn 2011, S. 165f.). Diese Möglichkeiten gehören in den größeren, narratologisch anscheinend noch weitgehend unerschlossenen Bereich der *Simultaneität*. Während man nur eine Sache zu gleicher Zeit erzählen kann, kann man mehr als eine Sache zu gleicher Zeit zeigen. Für eine *intermediale* Erzähltheorie wären grundsätzliche Überlegungen zur Simultaneität unerlässlich (vgl. Grubačić 1981). Auch das *Voice over* partizipiert auf spezifische Weise an der Frage nach der Simultaneität (da Zeigen auch gleichzeitig mit Erzählen stattfinden kann).

Literatur

- Bordwell, David: *Narration in the Fiction Film*. Madison: University of Wisconsin Press: 1985.
- Brössel, Stephan: „Die Narratologie des Films als Beitrag zu einer transmedialen Erzähltheorie“. (Rezension über: Markus Kuhn: *Filmnarratologie. Ein erzähltheoretisches Analysemodell*. Berlin/New York: Walter de Gruyter 2011). In: *IASLonline*. URL: http://www.iaslonline.de/index.php?vorgang_id=3455 (zit. am 12.10.2011).
- Chatman, Seymour: *Story and Discourse. Narrative Structure in Fiction and Film*. Ithaca: Cornell 1978.
- Chatman, Seymour: *Coming to Terms. The Rhetoric of Narrative in Fiction and Film*. Ithaca: Cornell 1990.
- Grubačić, Slobodan: „Simultane Welt“. In: Rolf Kloepfer/Gisela Janetzke-Pillner (Hg.): *Erzählung und Erzählforschung im 20. Jahrhundert*. Stuttgart u.a.: Kohlhammer 1981, S. 477-489.
- Hartmann, Britta: „Diegetisieren, Diegese, Diskursuniversum“. In: *Montage AV* 16/2 (2007), S. 53-69.
- Henderson, Brian: „Tense, Mood and Voice in Film. Notes after Genette“. In: *Film Quarterly*. Vol 36/4 (1983), S. 4-17.
- Kozloff, Sarah: *Invisible Storytellers. Voice-Over Narration in American Fiction Film*. Berkeley u.a.: University of California Press 1988.
- Kuhn, Markus: *Filmnarratologie. Ein erzähltheoretisches Analysemodell*. Berlin – New York: Walter de Gruyter 2011.
- Mahler, Andreas: „Erzählt der Film?“ In: *Zeitschrift für französische Sprache und Literatur* 111 (2001), S. 260-269.
- Martinez, Matias/Scheffel Michael (2005): *Einführung in die Erzähltheorie*. 6. Aufl. München: C.H. Beck 1999.
- Niehaus, Michael: „Aussagen sehen. Einfältige Gedanken zur Zweideutigkeit der Rückblende“. In: *Kulturrevolution*, 1/2006: mediale sichtbarkeiten, mediale blicke (2006), S. 22-29.

- Palmier, Jean-Pierre (2011): „Von Stimmen und Bildern. Ein erzähltheoretisches Analysemodell für bestimmte Filme.“ (Rezension über: Markus Kuhn: *Filmnarratologie. Ein erzähltheoretisches Analysemodell*. Berlin/New York: Walter de Gruyter 2011.) In: *IASLonline* [20.09.2011] URL: http://www.iaslonline.de/index.php?vorgang_id=3454 (zit. am 12.10.2011).
- Schäfer, Jerome Philipp (2008): „Grand Imagier oder Kamera? Zur Erzählinstanz im filmischen Kommunikationssystem“. In: *Medienobservationen*. URL: <http://www.medienobservationen.uni-muenchen.de> (zit. 12.10.2011)
- Schmid, Wolf: *Elemente der Narratologie*. Berlin – New York: Walter de Gruyter 2005.
- Weber, Dietrich (1998): *Erzählliteratur. Schriftwerk – Kunstwerk – Erzählwerk*. Göttingen: Vandenhoeck & Ruprecht 1998.
- Wolf, Werner (2002): „Das Problem der Narrativität in Literatur, bildender Kunst und Musik. Ein Beitrag zur intermedialen Erzähltheorie“. In: *Erzähltheorie transgenerisch, intermedial, interdisziplinär*. Hg. v. Ansgar Nünning / Vera Nünning. Trier: WVT 2006, S. 23-104.