

Steffen Lepa; Christian Floto

Audio-Vision als Konstruktion – Grundzüge einer funktionalistischen Audioanalyse von Film und Multimedia

2005

<https://doi.org/10.25969/mediarep/14252>

Veröffentlichungsversion / published version
Sammelbandbeitrag / collection article

Empfohlene Zitierung / Suggested Citation:

Lepa, Steffen; Floto, Christian: Audio-Vision als Konstruktion – Grundzüge einer funktionalistischen Audioanalyse von Film und Multimedia. In: Harro Segeberg, Frank Schätzlein (Hg.): *Sound. Zur Technologie und Ästhetik des Akustischen in den Medien*. Marburg: Schüren 2005 (Schriftenreihe der Gesellschaft für Medienwissenschaft (GfM) 12), S. 347–365. DOI: <https://doi.org/10.25969/mediarep/14252>.

Nutzungsbedingungen:

Dieser Text wird unter einer Deposit-Lizenz (Keine Weiterverbreitung - keine Bearbeitung) zur Verfügung gestellt. Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Terms of use:

This document is made available under a Deposit License (No Redistribution - no modifications). We grant a non-exclusive, non-transferable, individual, and limited right for using this document. This document is solely intended for your personal, non-commercial use. All copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute, or otherwise use the document in public.

By using this particular document, you accept the conditions of use stated above.

Audio-Vision als Konstruktion

Grundzüge einer funktionalistischen Audioanalyse von Film und Multimedia

Der vorliegende Beitrag versteht sich als ein Schritt, die wissenschaftliche Beschäftigung mit Klang und Geräusch aus ihrer Nischenposition¹ herauszuholen und endlich als wichtigen, leider oft sträflich vernachlässigten Teil der Medienwissenschaften zu begreifen.

Innerhalb des Braunschweiger Forschungsansatzes CRIMP², der sich mit der Evaluation von multimedialer Lernsoftware beschäftigt, versucht man in diesem Sinne neben der empirischen Forschungsarbeit zu Multimedia einerseits festzustellen, welchen Beitrag ganz unterschiedliche ‚Fremddisziplinen‘ wie die Pädagogik, Neurokognitionsforschung, Marketingpsychologie, Design- aber auch „Audiowissenschaft“ zu medienwissenschaftlichen Modellen von Multi-Mediawirkungen liefern können. Andererseits soll aber auch herausgearbeitet werden, inwiefern die Medienwissenschaften als interdisziplinäres Feld eine alternative Sichtweise für den mediendidaktischen Kontext auf die Herausforderung ‚Multimedia‘ in Zukunft liefern können. Die Untersuchung von Audio nimmt hierbei, wie festzustellen sein wird, aufgrund der ihr eigenen typischen Fallstricke eine prominente Rolle ein.

Weiter soll deutlich werden, dass die Analyse von Audiowirkungen ein entscheidend neues Licht auf die Analyse von multimedialen Wirkungen im Ganzen werfen kann. Kronzeugen für die Verifikation dieser Behauptung sollen im Folgenden Ergebnisse der neurokognitiven Grundlagenforschung und die Ausführungen des französischen Film- und Klangtheoretikers Michel Chion in seinem vielbeachteten Werk *Audio-Vision. Sound On Screen*³ sein.

1 Wenn nicht sogar ihrer „splendid isolation“.

2 Projektgruppe VASE IV: „Criteria for Audiovisuals in Multimedia Production (CRIMP)“ im Rahmen des Forschungs- und Kompetenzverbundes „Learning Lab Lower Saxony“, nähere Informationen dazu finden sich unter: <http://www.learninglab.de/vase4/>.

3 Michel Chion: *Audio-Vision. Sound on Screen*. New York 1994.

Der „Audiovisuelle Kontrakt“: Film, Fernsehen und Multimedia als „Bauchredner-Effekt“

I use the phrase ‚audiovisual contract‘ as a reminder that the audiovisual relationship is not natural, but a kind of symbolic contract that the audio-viewer enters into, agreeing to think of sound and image as forming a single entity.⁴ (*Chion*)

Wir sind es bisher gewohnt, bei der Analyse von Film oder multimedialen Produkten die verschiedenen Modalitäten getrennt voneinander zu analysieren. So untersucht man einerseits zum Beispiel Bildästhetik, Kameraeinstellungen und visuelle Narration und wendet sich getrennt davon dann der Tonspur zu, um diese in Hinblick auf Musik-, Sprach- und Geräuscheinsatz zu untersuchen. Am Schluss steht schließlich eine Betrachtung der Dramaturgie und Gesamtwirkung der filmischen Narration. Dies geschieht, je nach Untersuchungsparadigma und Intention der Analyse, entweder nach psychoanalytischen, semiotischen, kognitiven oder pädagogischen Aspekten. Bis Anfang der neunziger Jahre schien dieses Vorgehen auch durchaus seine Berechtigung zu haben, ging man doch auch innerhalb der Kognitionsforschung lange davon aus, das die unterschiedlichen sensorischen Einspeisungen unserer Sinnesorgane im Gehirn zwar durchaus konvergieren, dies aber erst relativ spät und nur auf semantischer Ebene, im Sinne referentieller Verknüpfungen zwischen beispielsweise visuell identifizierten Objekten und gleichzeitig gehörter Sprache.⁵

Das in dieser Hinsicht bekannteste multimodale Phänomen ist der „Ventriloquismus“, also der sogenannte „Bauchrednereffekt“⁶, der bis heute die Wahrnehmung von Fernsehen beherrscht.⁷ Es wurde und wird anscheinend wie selbstverständlich davon ausgegangen, dass die mehrheitlich faktisch ‚falsche‘ räumliche Positionierung des Klanges vom Gehirn des Rezipienten auf semantischer Ebene ‚richtiggestellt‘ wird, und somit unser erlerntes kulturelles Wissen dazu führt, dass wir den Klang der Sprache des Fernsehmoderators kausal auf dessen Bild bzw. Mund beziehen oder Geräusche von z. B. Füßen auf Treppenstufen mit dessen visuellem Pendant verknüpfen. Chion spricht hier, im Gegensatz dazu, von einer mental konstruierten „räumlichen Magnetisierung“ („spatial magnetization“) des Klanges durch das Bild.⁸

4 Ebd. S. 216.

5 A. Paivio: *Mental representations: A dual coding approach*. Oxford 1986.

6 Vgl. P. Bertelson/M. Radeau: „Cross-modal bias and perceptual fusion with auditory-visual spatial discordance“. In: *Perception & Psychophysics* (1981), Nr. 29 (6), S. 578–584.

7 Für 99% der Rezipienten natürlich auch beim Surround-Klang.

8 Chion: *Audio-Vision. Sound on Screen*. S. 70.

Neuronale Mechanismen multimodaler Integration

Aktuelle Untersuchungen aus dem Bereich der Neurokognition legen ebenfalls nahe, dass die Auffassung über das Zusammenspiel der Sinne als rein „semantisch“ und „kulturell erlernt“ in ihrer Allgemeinheit etwas zu kurz gegriffen war: Multimodale Integration findet offenbar bereits auf sehr frühen Stufen der Wahrnehmung, teils noch vor der Objekterkennung statt.

In einer Pionierstudie wiesen Stein und Meredith durch Ableitung der elektrischen Potentiale einzelner Nervenzellen erstmals die Existenz von „multimodalen“ Neuronen im Superior Colliculus der Katze nach.⁹ Ihre Untersuchungen zeigten, dass sensorische Erregungen aus unterschiedlichen Sinnesorganen zu überadditiven, gleichsam „multiplikativen“ Erregungen in den von den Forschern gefundenen „intersensorischen“ bzw. heteromodalen Arealen des Gehirns führen können. Die jeweiligen rezeptiven Felder¹⁰ der hier einspeisenden modalitätsspezifischen sensorischen Hirnareale weisen dabei starke, wenn auch nicht exakte, räumliche Überlappungen auf. Ergebnis dieser Untersuchungen¹¹ war die empirische Bestätigung zweier bestimmender, bis dato nur theoretisch postulierter Prinzipien bei der multimodalen Integration auf neuronaler Ebene: Erstens die sogenannte „Raumregel“, nach der räumliche Nähe, und zweitens die „Zeitregel“, nach der gemeinsames zeitliches Auftreten von Reizen unterschiedlicher Modalitäten zur Wahrnehmung eines integrierten multimodalen Objektes führt. Wohlgermerkt: Eine stark „unscharfe“ Erfüllung beider Bedingungen ist prinzipiell hinreichend. Dies macht auch Sinn, da Schall sich durch vielfache, unterschiedliche Reflektionen nicht wie Visuelles linear durch den euklidischen Raum bewegt, sondern, bildlich formuliert, sich eher wie Gas diffus im Raum ausbreitet.

Außerdem schließen die unterschiedlich langen Signalwege der verschiedenen Sinnesempfindungen ohnehin eine echte ‚Gleichzeitigkeit‘ in der Wahrnehmung aus. Dies lässt sich auch empirisch zeigen, z. B. anhand des von Sekuler et al. gefundenen „Bouncing“-Paradigmas.¹² Bei diesem Versuch werden zwei Lichtreihen kreuzweise aufeinander zugeführt. Versuchspersonen interpretieren diese basale Form der Animation normalerweise als ein Überkreuzen. Bietet man jedoch im Moment der Überschneidung gleichzeitig einen auditiven oder

9 Barry E. Stein/M. Alex Meredith: *The Merging of the Senses*. Cambridge, Massachusetts (USA) 1993.

10 Durch die jeweiligen Neuronen sensorisch abgedeckte Bereiche der „Außenwelt“.

11 Calvert et al. (G. A. Calvert/R. Campbell/M. J. Brammer: „Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex“. In: *Current Biology* (2000), Nr. 10, S. 649–657) wiesen in Folge nach, dass analoge Integrationsprinzipien auch bei Primaten und im menschlichen Kortex stattfinden.

12 Vgl. R. Sekuler/A. B. Sekuler/R. Lau: „Sound alters visual motion perception“. In: *Nature* (1997), Nr. 385, S. 308.

taktilen Kurzreiz wie einen Piepton dar, ändert sich die Interpretation der meisten Versuchspersonen dergestalt, dass plötzlich eine Kollision wahrgenommen wird. Variiert man nun Intensität und Zeitpunkt des Pieptons oder taktilen Reizes, finden sich die Geschwindigkeitsunterschiede in der Verarbeitung der verschiedenen Modalitäten und die Unschärfe der Gleichzeitigkeitsempfindung in den psychophysischen Kurven ihrer Perzeption wieder.¹³

Zu den zwei bereits zuvor theoretisch bekannten „Raum- und Zeitregeln“ kam durch die Untersuchungen von Stein und Meredith die sogenannte „Effektivitätsregel“ hinzu, womit gemeint ist, dass gerade bei relativ schwachen, aber koinzidenten bimodalen Reizen besonders starke übersummativ Effekte auftreten. Inzwischen zeichnet sich bereits ein viertes Prinzip ab, das allerdings hier erst später erläutert werden wird. Entscheidend für die weitere Betrachtung ist, dass der Zeitpunkt der von ihnen und anderen Forschern gemessenen multimodalen Effekte teilweise so früh ist, dass man den Einfluss tieferer, semantischer Verarbeitung ausschließen kann.¹⁴ Multimodale Integration findet also bereits auf Ebene der Perzeption statt.

„Added Value“ durch Crossmodale Interaktionseffekte – Konvergenz oder Feedback?

Added value works reciprocally. Sound show us the image differently than what the image shows alone, and the image likewise makes us hear sound differently than if the sound were ringing out in the dark.¹⁵ (Chion)

Wir haben also gesehen, dass multimodale Wahrnehmung zu „überadditiven“ neuronalen Reaktionen führt. Aber inwiefern ist bei der multimodalen Wahrnehmung das Ganze nun auch qualitativ mehr als die Summe seiner Teile? Inzwischen mehrten sich durch EEG-Untersuchungen diesbezüglich die Hinweise, dass das bisher angenommene Modell einer konvergenten Verschaltung der Informationsströme von den sensorischen zu den heteromodalen Arealen (Abb. 1) des Cortex nicht ausreicht, um die beobachtbaren *crossmodalen* Rückprojektions-Phänomene zu erklären.¹⁶ Damit sind solche Situationen gemeint, in denen die experimentell dargebotenen Reize einer Modalität bereits nach

13 S. Shimojo/L. Shams: „Sensory modalities are not separate modalities: plasticity and interactions“. In: *Current Opinion in Neurobiology* (2001), Nr. 11, S. 505–509.

14 Ebd.

15 Chion: *Audio-Vision. Sound on Screen*. S. 21

16 Jon Driver/Charles J. Spence: „Multisensory Perception. Beyond modularity and convergence“. In: *Current Biology* (2000), Nr. 10, S. R731–R735.

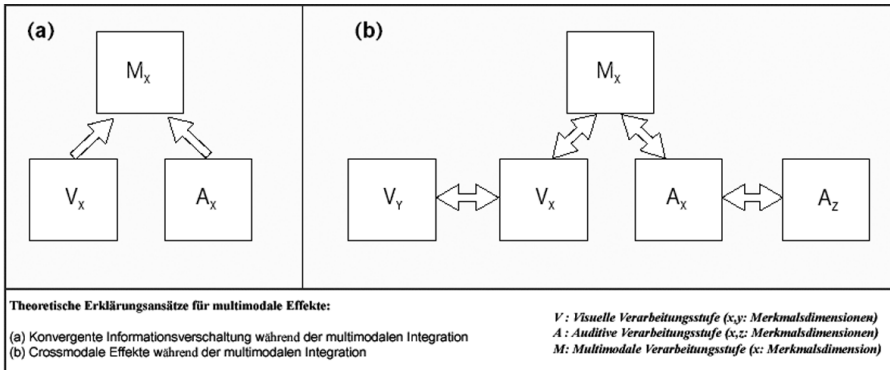


Abb. 1: Multimodale Integration – Konvergenz oder Feedback?

sehr kurzer Zeit die kognitive Verarbeitung bei der Wahrnehmung einer anderen Modalität qualitativ oder quantitativ verändern.¹⁷

Offenbar dienen heteromodale Areale des Gehirns als eine Art „Brücke“, die intersensorische Bindungsprobleme löst.¹⁸ Auf das Beispiel des TV-Moderators bezogen lautet die Frage also: Handelt es sich beim Bauchrednereffekt um eine semantisch glaubwürdige Illusion oder nehme ich auf Ebene der frühen Wahrnehmung die Szene tatsächlich so wahr, als spräche er? Es mehren sich inzwischen die Nachweise, dass in vielen Situationen tatsächlich letzteres eintritt. Auszugsweise seien hier einige Befunde genannt: Visuell dargestellte stumme Lippenbewegungen aktivieren frühe auditive Verarbeitungsstufen,¹⁹ taktile und auditive Zielreize sind in der Lage, den primären visuellen Kortex in Erwartung visueller Reize vorzuaktivieren²⁰ und insgesamt die Reizverarbeitung zu modulieren.

Der semantische Gehalt einer der zwei jeweils beteiligten Modalitäten ist dabei für die crossmodalen Effekte auf Ebene der Wahrnehmung und Aufmerksamkeit offenbar kaum entscheidend, dürfte aber später im Zeitverlauf der kognitiven Verarbeitung natürlich durchaus eine Rolle spielen, speziell bei der Sprachwahrnehmung, wie Experimente der britischen Neurokognitionsforscherin Gemma Calvert und Kollegen nahe legen.²¹ Sie identifizierten per fMRI²²

17 Vgl. z. B. E. Macaluso/C. D. Frith/J. Driver: „Modulation of human visual cortex by crossmodal spatial attention“. In: *Science* (2000), Nr. 289, S. 1206–1208.

18 Gemma A. Calvert/Michael J. Brammer/Susan D. Iversen: „Crossmodal identification“. In: *Trends in Cognitive Sciences* (1998), Nr. 2 (7), S. 247–253.

19 G. A. Calvert et al.: „Activation of auditory cortex during silent lipreading“. In: *Science* (1997), Nr. 276, S. 593–596.

20 Shimojo/Shams: „Sensory modalities are not separate modalities: plasticity and interactions“.

21 Calvert/Campbell/Brammer: „Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex“.

inzwischen eine Region im linken superior temporalen Sulcus, in welcher die Kongruenz dargebotener Sprache mit gleichzeitig visuell dargestellten Lippenbewegungen die intersensorischen Effekte modulierte. Dafür war hier die raum-zeitliche Überlappung weniger kritisch. Es wird deutlich: Multimodale Wahrnehmung ist durch ein interaktives Zusammenspiel von kulturell erlerntem „Weltwissen“ und frühen perzeptuellen Integrationsprinzipien geprägt, wie auch Chion bereits dargelegt hat: „A music cue ore voiceover that is culturally perceived as not „in“ the setting will not set the image to vibrating. And yet, the phenomenon still has a noncultural basis.“²³

Festzuhalten bleibt in diesem Zusammenhang, dass in vielen experimentellen Versuchsanordnungen auch sogenannte orthogonale/akzessorische, also nicht-informative und vom Zielreiz arbiträre Reize in einer anderen Modalität, per EEG messbare Aufmerksamkeitsverlagerungen, Erhöhungen der Reaktionszeit, optische Illusionen und Pop-Out-Effekte hervorriefen.²⁴ Interessanterweise kann dabei auch illusionär die Wahrnehmungsschwelle einer Modalität unterschritten werden, wie im Double-Flash-Phänomen, wo vor und nach einem Doppelblitz platzierte Pieptöne diese überhaupt erst als solche wahrnehmbar machen, da der Zeitunterschied eigentlich zu gering wäre, um ein wahrnehmbares Blitzen zu erzeugen.²⁵

Ganz ähnliche „faster-than-eye“-Phänomene²⁶ kennen wir auch aus dem Spielfilm, Chion nennt dafür in *Audio-Vision* einige konkrete Beispiele: So würden die komplizierten Bewegungsabläufe der Samurai und Kung-Fu-Kämpfer erst durch kookkurrenente auditive Unterstützung nachvollziehbar und würden Bewegungssillusionen wie das Öffnen der Türen im ersten STAR WARS-Spielfilm rein über Klang erzeugt.

Gibt es eine „Dominanzregel“ der audiovisuellen Wahrnehmung?

So, overall, in a first contact with an audiovisual message, the eye is more spatially adept, and the ear more temporally adept.²⁷ (Chion)

Betrachtet man die verschiedenen neuropsychologischen Experimente genauer, so fällt auf, dass häufig jeweils eine Modalität die andere dominiert, in dem Sinne, das die erstere ‚authentischere‘ bzw. exaktere Ergebnisse bezüglich

22 D.i.: Functional magnetic resonance imaging (dt: funktionale Kernspintomographie).

23 Chion: *Audio-Vision. Sound on Screen.* S. 20.

24 Vgl. z. B. Shinsuke Shimojo et al.: „Beyond perceptual modality: Auditory effects on visual perception“. In: *Acoustical Science and Technology* (2001), Nr. 22 (2), S. 61–67.

25 L. Shams/Y. Kamitani/S. Shimojo: „What you see is what you hear“. In: *Nature* (2000), Nr. 408, S. 788.

26 Chion: *Audio-Vision. Sound on Screen.* S. 12 ff.

27 Ebd. S. 11.

handlungsrelevanter Details liefert und damit die zeitliche oder räumliche Wahrnehmung der anderen verzerrt bzw. verändert. Dies scheint je nach experimentellem Paradigma und teilweise auch zwischen Versuchspersonen unterschiedlich zu sein. Man versucht dementsprechend nun experimentell zu ermitteln, welche Antezedenten dazu führen, dass bestimmte Modalitäten bei der Objekterkennung, Handlungsplanung und Interpretation ‚multimodaler Objekte‘ situativ eine Führungsrolle spielen, mit anderen Worten, wann Audio die visuelle Wahrnehmung ‚manipuliert‘, wie bei der „Double-Flash Illusion“, und wann umgekehrt.

Nach den aktuellen Befunden ist dies offenbar abhängig von a) dem antizipierten Informationsgehalt der jeweiligen Modalität betreffs des aktuellen Handlungskontextes²⁸, b) der perzeptuellen Salienz²⁹ der jeweiligen Stimuli³⁰ und c) dem situativen kognitiven Stil im Zusammenspiel mit den persönlichen kognitiven Fähigkeiten der Versuchspersonen³¹. Das bedeutet, wie bereits von Hans-Christian Schmidt³² im Kontext von Aufmerksamkeit und Filmmusik theoretisch formuliert: Wenn mir bei der medialen Wahrnehmung und Bedeutungskonstruktion in Bezug auf die aktuelle Handlung semantische Informationen fehlen, ziehe ich, um entstehende informationelle Unentschiedenheit zu kompensieren, diejenige Modalität zur Hilfe hinzu, die ich (vielleicht wegen meines situativen kognitiven Stils) bisher nur im Hintergrund und als arbiträr wahrgenommen hatte, welche sich jetzt aber am effektivsten erweist, die vorhandene Unsicherheit zu reduzieren.

Generell scheinen die von Chion diesbezüglich bereits theoretisch postulierten Grundprinzipien für Film³³ aber auch bei Multimedia wirksam zu sein, solange ‚typisch‘ filmische Narrationsverhältnisse herrschen: Visuelles dominiert also in der Raum- und Objektwahrnehmung, Auditives in der Zeitwahrnehmung und der emotionalen Prägung der Szene, beides wird allerdings grundsätzlich durch Sprachwahrnehmung, egal in welcher Modalität, dominiert. Insofern scheint nicht nur der Film³⁴, sondern auch Multimedia, soweit es narrative Strukturen in den meisten kommerziellen Lernprodukten betrifft, ein der Tendenz nach „verbozentrisches“³⁵ Medium zu sein. Dies spiegelt sich dann auch in einem Großteil der em-

28 Shimojo et al.: „Beyond perceptual modality: Auditory effects on visual perception“.

29 Kognitionspsychologischer Terminus für „hervorstechende“, prägnante Reizeigenschaften.

30 Shimojo/Shams: „Sensory modalities are not separate modalities: plasticity and interactions“.

31 Alexandra Fort et al.: „Early auditory-visual interactions in human cortex during nonredundant target identification“. In: *Cognitive Brain Research* (2002), Nr. 14, S. 20–30.

32 Hans-Christian Schmidt: „Spiel mir das Lied... Entwicklungslinien der Filmmusik“. In: *Universitas* (1988), Nr. 43 (502), S. 407–421.

33 Chion: *Audio-Vision. Sound on Screen*. S. 11 ff.

34 Ebd. S. 6.

35 Vgl. ebd. S. 5 f.

pirischen Arbeiten und Modellvorstellungen zu Multimedia als Lernwerkzeug wider, der Thematik, mit der sich die CRIMP-Gruppe im Kern befasst³⁶; sie wird auch die folgenden Ausführungen bestimmen.

Einige medienpädagogische Autoren weisen aber zu recht daraufhin, dass Multimedia als Instrument der Wissensvermittlung noch in den Kinderschuhen steckt und hier noch Änderungen eintreten könnten und sollten,³⁷ gerade weil dadurch bisher wegen ihres kognitiven Stiles benachteiligte Lerner davon profitieren dürften. So existieren schließlich unterschiedliche Lernertypen, wie Visualisierer, Prozessorientierte Typen, Domänenspezifische Lerner etc. mit einem unterschiedlichen Bedarf an modalitätsspezifischen Repräsentationsstrategien. Eines der erklärten Ziele der Forschungsgruppe CRIMP war und ist es deshalb, den Einfluss dieser Variablen im Zusammenspiel mit den geschilderten crossmodalen und Dominanzprinzipien während der Wahrnehmung multimedialer Narration zu untersuchen.³⁸ Diesen Ansatz in Bezug auf Audio werden wir im Folgenden darstellen.

Auf der Suche nach fundierten Evaluationskriterien für Audioeinsatz in Multimedia

Stellen sie sich vor, Sie seien mit der Erstellung eines Konzeptes für eine Lernsoftware-Produktion beschäftigt. Sämtliche einzusetzenden Gestaltungsmittel wollen mit Bedacht auf die zu erwartenden Kosten sorgfältig ausgewählt sein. Als Produzent sind Sie gleichzeitig eingebettet in einen Kontext von Sachzwängen der Produktion und Erwartungshaltungen von Seiten des Marktes. Welchen Stellenwert nimmt in Ihrer Konzeption die Audio-Ebene ein und nach welchen Kriterien entscheiden Sie über die Verwendung von Atmo, Musik, Speech und Spezialeffekten? Sicher, Sie könnten einen Profi engagieren, der ihnen hochästhetische, zeit- und zielgruppengemäße Musik komponiert und eine stimmungsvolle Klangatmosphäre schafft. Letztlich ist der mögliche Erfolg dann evtl. vom möglichst breiten Treffen (oder bei Misserfolg: vom Verfehlen) des Geschmacks des Ziel-Publikums abhängig. Ob die Audio-Ebene der Darbietung deswegen im Sinne der neben der Vermarktung zweiten Zielsetzung ‚Lernen‘ erfolgreich ist, sei hiermit mal dahingestellt. Auf der anderen Seite

36 Thomas Huk/Matthias Steinke/Christian Floto: „Helping Teachers Developing Computer Animations for Improving Learning in Science Education“. *SITE (Society for information technology & teacher education) Annual Conference*, Albuquerque, New Mexico (USA) 2003.

37 Michael D. Sankey: „Multiple Representation in Multimedia Material. An Issue of Literacy“. In: Mishra/Sharma (Hg.): *Interactive Multimedia in Education and Training*. London (UK) 2003.

38 Matthias Steinke/Thomas Huk/Christian Floto: „The process of learning with hypermedia systems: Linking learner characteristics, software design and log files“. *ED-Media Conference 2003*, Honolulu, Hawaii (USA) 2003.

könnten Sie auch pädagogisch an die Sache herangehen und nach ergonomisch-psychologischen Gesichtspunkten eine didaktisch sinnvolle Audio-Ebene herstellen, die zu einer optimalen Lernsituation beiträgt – dann aber von einem Großteil des Publikums als langweilig und fade empfunden wird.

Auf eine dieser beiden Arten werden offenbar aktuelle didaktisch inspirierte Softwareprodukte mit Digital Audio ausgestattet, die Nachteile sind mehr als offensichtlich. Unser Ansinnen ist es, in Bezug auf multimediale Narrationen als einem neben Hypertext, Simulation und Spiel bedeutenden Bestandteil multimedialer Wissensvermittlung mit diesen Dualismen aufzuräumen und eine sachliche, an funktionalen Prinzipien orientierte Diskussion über Audiowirkungen anzustrengen, die sich auch ästhetischen und unterhaltenden Produkten, wie sie mehr und mehr den Lernsoftware-Markt bestimmen, nicht aus falschen Vorbehalten heraus verschließt. Durch die Untersuchung der empirischen und theoretischen Arbeiten auf diesem Gebiet wurde hierbei deutlich: Viele medienpädagogische Modelle haben sich überhaupt noch nicht ernsthaft mit der Vielfalt *auditiver* Wirkungen auseinandergesetzt.

Existieren überhaupt adäquate Wirkmodelle für das Lernen mit Multimedia?

Visual and auditory perception are much more disparate natures than one might think. The reason we are only dimly aware of this is that these two perceptions mutually influence each other in the audiovisual contract, lending each other their respective properties by contamination and projection.³⁹ (Chion)

Innerhalb der großen Anzahl von empirischen Arbeiten und Theorien zur Evaluation multimedialer Bildungsprodukte geht es dementsprechend mehrheitlich um Visualisierungsstrategien und das Verhältnis von Bild und Sprache, nur wenige Arbeiten befassen sich aber bisher mit nicht-sprachlichen auditiven Gestaltungsmitteln. Wenn überhaupt, so werden meist einfach die Prinzipien visueller Gestaltung auf den auditiven Bereich übertragen, was unserer Auffassung nach einfach zu kurz gefasst ist, da ein Großteil auditiver Narrationsstrategien im Gegensatz zum Visuellen nicht-referentieller Natur ist.

Für das richtige Verständnis und damit den Aufbau von Wissensstrukturen während eines multimedialen Lehrfilmes bedarf es laut Modellen der empirischen Medienpädagogik einer kohärenten Handlungsstruktur, einer nicht zu starken Auslastung des Arbeitsgedächtnisses und möglichst weniger Ablenkungseffekte. Eine Anzahl einflussreicher instruktionspsychologischer Theo-

39 Chion: *Audio-Vision. Sound on Screen*. S. 9.

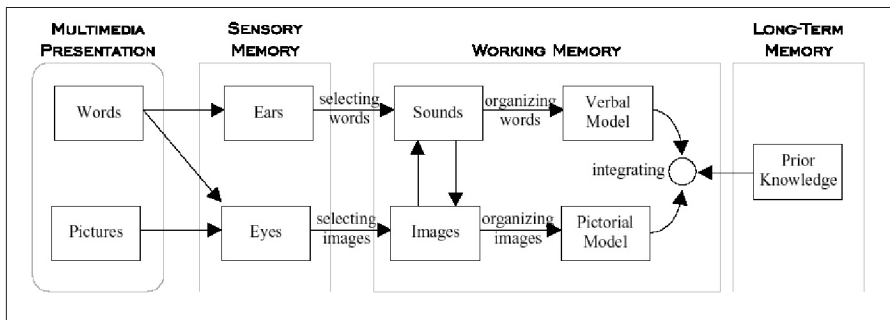


Abb. 2: Modell der Dualen Codierung (Mayer, 2001)

rien beschäftigt sich nun schon seit geraumer Zeit mit der Frage, wann und unter welchen Aspekten multimodale Darbietungen in diesem Sinne förderlich und wann ‚schädlich‘ für den Lernprozess mit multimedialen Filmen und Animationen sein können. Die wichtigsten sind die „Cognitive Theory of Multimedia Learning“⁴⁰ und die „Cognitive-Load-Theory“⁴¹. Beide Theorien gehen (u. a.) von der Grundannahme aus, dass ein semantischer Bezug, genauer die Kongruenz zwischen „Objekten“ zweier unterschiedlicher Modalitäten, ein entscheidender Faktor im Hinblick auf gut optimiertes multimediales Lernmaterial sei (vgl. z. B. Mayers Modell der dualen Codierung für Multimedia, Abb. 2).

Diese Überlegung⁴² erscheint sinnvoll, da bereits miteinander zu multimedialen Gestalten und Handlungen integrierte Bedeutungsträger logischerweise nicht zu einer Aufmerksamkeitsablenkung und/oder Überladung des Arbeitsgedächtnisses führen können, da sie vom Gehirn eben als identische Schemata (also als ein einziges Meta-Objekt) interpretiert werden.⁴³ Unimodale Teilaspekte einer auf diese Art erzeugten multimedialen narrativen ‚Gestalt‘ werden also zu bloßer Textur und sind damit keine potentiellen Störquellen bei der Kohärenzbildung mehr. Gleichzeitig werden durch diese Art der Wissensrepräsentation natürlich eine größere Anzahl verwertbarer „Cues“⁴⁴ für den Informationsabruf generiert und den individuellen kognitiven Stilen unterschiedlicher Lerner wird entgegengekommen.

40 Richard E. Mayer: *Multimedia Learning*. Cambridge 2001.

41 P. Chandler/J. Sweller: „Cognitive Load Theory and the format of instruction“. In: *Cognition and Instruction* (1991), Nr. 8 (4), S. 293–332.

42 Theoretischer Vorläufer ist das Modell der „Dualen Codierung“ von Paivio (Paivio: *Mental representations: A dual coding approach*).

43 E. Pollock/P. Chandler/J. Sweller: „Assimilating complex information“. In: *Learning and Instruction* (2002), Nr. 12 (1), S. 61–86.

44 Perzeptuelle „Anker“, die beim Wissensabruf helfen.

Die Audioebene bietet hier in Bezug auf Sprachdarbietung den besonderen Vorteil, dass sich der Rezipient visuell ganz auf die Bildinhalte konzentrieren kann, nicht zwischen visuellem Text und Animation oder Abbildungen hin und her springen muss und dadurch ein sogenannter „Split-Attention-Effect“⁴⁵ vermieden wird.

In dieser Hinsicht werden von einigen Autoren als multicodal „redundant“⁴⁶ bzw. „kohärent“ bezeichnete Darstellungsweisen desselben Sachverhaltes (z. B. Bild + auditiver Erklärungstext) daher als positiv und lernförderlich betrachtet, was auch bereits in einigen Fällen empirisch bestätigt werden konnte.⁴⁷ Oberflächlich betrachtete nondiegetische auditive Gestaltungselemente wie Musik oder Geräuschkulissen (Atmo) werden dagegen eher als „irrelevant“ abgelehnt.⁴⁸

Offensichtlich sind also diese Lernmodelle nicht für das Lernen mit Multimedia bzw. multimodalen Narrationen, sondern für das Lernen mit Bildern und Texten konzipiert.

Audiovisuelle Projektion und Synchronisation – Klang und Musik als Bedeutungsträger

Ich verstehe unter Sprache die im Vers zusammengefühten Wörter und unter Melodik das, was seine Wirkung ganz im sinnlichen entfaltet.⁴⁹ (Aristoteles: *Poetik*)

So scheint das Herstellen von multimodalen „Redundanzen“⁵⁰ der entscheidende didaktische Faktor bei Multimedia zu sein. Oder etwa doch nicht? Aktuelle Ergebnisse der Forschung der CRIMP-Gruppe⁵¹ legen nahe, dass, ähnlich wie bereits von Jeung, Chandler und Sweller 1997 gefunden,⁵² die vielbeschwo-

45 Vgl. Nitzan Ben-Shaul: „Split attention problems in interactive moving audiovisual texts“. *MelbourneDAC, 5th International Digital Arts and Culture Conference*, Melbourne (AUS) 2003.

46 Der Genauigkeit wegen sollte hier lieber der Ausdruck „kongruent“ verwendet werden.

47 Vgl. z. B. Richard E. Mayer/Roxana Moreno: „A Split-Attention Effect in Multimedia Learning: Evidence for Dual Processing System in Working Memory“. In: *Journal of Educational Psychology* (1998), Nr. 90 (2), S. 312–320.

48 Roxana Moreno/Richard E. Mayer: „A Coherence Effect in Multimedia Learning: The Case for Minimizing Irrelevant Sounds in the Design of Multimedia Instructional Messages“. In: *Journal of Educational Psychology* (2000), Nr. 92 (1), S. 117–125.

49 Aristoteles: *Poetik*. Griechisch/Deutsch. Stuttgart 1982, S. 19.

50 M. J. Bishop/Ward Mitchell Cates: „Theoretical foundations for sound’s use in multimedia instruction to enhance learning“. In: *Educational technology research and development* (2001), Nr. 49 (3), S. 5–22.

51 Thomas Huk/Matthias Steinke/Christian Floto: „The educational value of cues in computer animations and its dependence on individual learner abilities.“ *Ed-Media Conference 2003*, Honolulu, Hawaii (USA) 2003.

nen didaktischen Vorteile bimodaler Darstellung bei hoher Komplexität der visuellen Darstellung wieder verschwinden können, speziell, wenn die Lerner (z. B. aufgrund mangelnden räumlichen Vorstellungsvermögens oder Vorwissens) nicht in der Lage sind, den auditorisch dargebotenen Text auf komplexe dreidimensionale, realitätsnahe Abbildungen zu beziehen. In solchen Fällen können deiktische Elemente wie Pfeile,⁵³ aber auch optisches/akustisches „Highlighting“⁵⁴ oder auditorische Signale eine wertvolle Hilfe sein. Diese „kohärenzbildenden Maßnahmen“ sind in der Lage, das offensichtlich innerhalb der Narrationsstruktur aufgrund mangelnder semantischer Integrationsfähigkeiten der Lerner entstandene Bindungsproblem zwischen sprachlicher Erklärung und visueller Simulation zu lösen; der angestrebte „überadditive“ Effekt multimodaler Darstellung wird dadurch restauriert.

Dies ist nur *ein* Beispiel für Fälle, in denen offenbar doch in Bezug auf die sprachliche und visuelle Narration semantisch nicht-redundante Cues in der Lage sind, durch zeitliche (auditorisch) und/oder räumliche (visuell) Koinzidenz, Bindungsprobleme zu lösen und damit der Narration zur Kohärenzbildung und dem Rezipienten zum Verständnis zu verhelfen. Es liegt daher nahe vorzuschlagen, die Cognitive Load Theory und die Cognitive Multimedia Learning Theory um Ansätze zu erweitern, die sich mit der Optimierung nicht-repräsentationaler, impliziter Aspekte der multimedialen Narration und Wissensgenerierung beschäftigen. Die Forderung einiger Autoren, ‚irrelevantes‘ auditorisches Material wie Musik und Klang⁵⁵ möglichst bei multimodalen Darstellungen grundsätzlich zu vermeiden, da es Ablenkung oder Überforderung erzeugen würde,⁵⁶ ist zwar in einigen Fällen sicherlich richtig, in der Allgemeinheit und Schärfe der Formulierung allerdings zurückzuweisen, da mit nachrichtentechnischer bzw. logozentrischer Interpretation der Wirkungsweise der semiotischen Maschine Computer und ihrer inzwischen filmrealistischen Ausdrucksfähigkeit unserer Auffassung nach nicht beizukommen ist.

52 H. J. Jeung/P. Chandler/J. Sweller.: „The role of visual indicators in dual sensory mode instruction“. In: *Educational Psychology* (1997), Nr. 17 (3), S. 329–343.

53 Huk/Steinke/Floto: „The educational value of cues in computer animations and its dependence on individual learner abilities.“

54 Vgl. Douglas G. Olsen: „Creating the Contrast: The Influence of Silence and Background Music on Recall and Attribute Importance“. In: *Journal of Advertising* (1995), Nr. 24 (4), S. 29–44.

55 Gemeint sind wohl „akusmatische“ Klänge im Sinne von Pierre Schaeffer (vgl. Chion: *Audio-Vision. Sound on Screen*).

56 Vgl. Moreno/Mayer: „A Coherence Effect in Multimedia Learning: The Case for Minimizing Irrelevant Sounds in the Design of Multimedia Instructional Messages“.

Audio als Metapher: Die vielfältigen Relationen zwischen Klang und Bild

In short, we classify sounds in relation to what we see in the image, and this classification is constantly subject to revision, depending on changes in what we see. Thus, we can define most cinema ‚as a place of images plus sounds‘, with sound ‚being that which seeks its place.‘⁵⁷

Offensichtlich ist die gegenseitige Beziehung zwischen Bild und Ton, der ‚audiovisuelle Kontrakt‘, bei der narrativen Kohärenzbildung nicht so einfach mit informationstheoretischen Begrifflichkeiten zu fassen, wie es uns Begriffsverwendungen wie ‚redundant‘ und ‚irrelevant‘ glauben machen wollen. Es lohnt deshalb, ähnlich wie Aristoteles in Bezug auf die Metaphernlehre der Sprache in seiner *Poetik* (s.o.!), einen allgemeinen Blick auf die unterschiedlichen möglichen, sinnvollen semiotischen Beziehungen zwischen Klang und Bild zu werfen und dabei unter Zuhilfenahme der eingangs zitierten neurokognitiven Befunde und Chions Axiomen zur Audio-Vision festzustellen, auf welchen unterschiedlichen Wegen dabei bimodal Bedeutung erzeugt wird.

Die vielfältig in multimedialen Lernprodukten implementierten Beziehungen zwischen Ton und Bild lassen sich grob in drei Gruppen von Bindungsstrategien unterteilen; Chion spricht hier im kinematographischen Zusammenhang von drei Arten des Hörens, die wiederum von allen drei unterschiedlichen auditiven Ausdrucksmitteln Sprache, Geräusch und Musik, selbst-

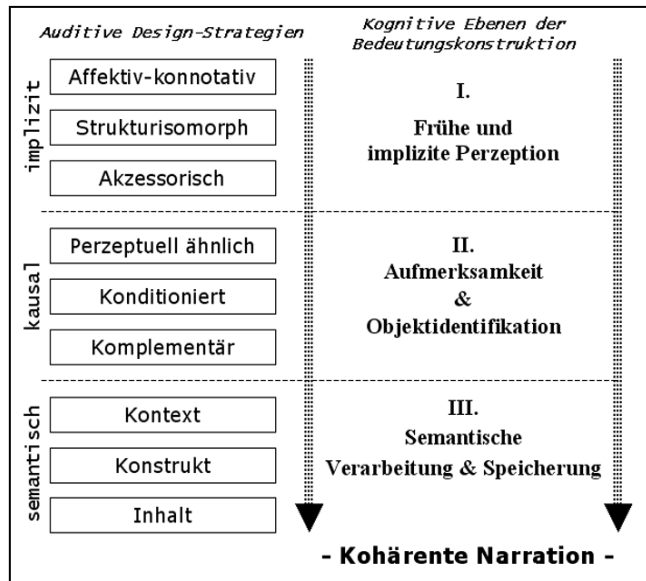


Abb. 3: Multi-Level-Forschungsansatz zur Evaluation von Audio in multimedialen Lernprodukten.

57 Chion: *Audio-Vision. Sound on Screen.* S. 68.

verständlich auch in Doppel- oder Mehrfachrollen bestritten werden können (Abb. 3, im Folgenden erläutert).

Eine erste Klasse von audiovisuellen Relationen sind die semantischen. Ihr Beitrag zum Instruktionsdesign wurde bereits ausführlich in einem Modell von Bishop und Cates gewürdigt und analysiert.⁵⁸ Die beiden Medienpädagogen unterscheiden bei ihrer Betrachtung semantischer audiovisueller Verknüpfungen zwischen Inhalts-, Konstrukt- und Kontextredundanzen. Erstere wären die informationell tatsächlich redundanten sprachlichen Bindungen, die durch die Kombination identischer geschriebener und gesprochener Sprache entstehen. Sogenannte Konstruktredundanzen meinen wiederum eher kongruent-komplementäre bzw. polyseme Verhältnisse zwischen Klang und Bild, so wie zum Beispiel zwischen einer Abbildung und deren sprachlicher Erklärung. Kontextredundanter Audioeinsatz erfüllt schließlich Situierungsfunktionen, indem er einen kongruenten auditiven Hintergrund für die dargestellte Szene liefert und kontextbezogene Abruf-Cues herstellt. Hierzu gehören primär Atmo-Sounds, die einen authentischen klanglichen Hintergrund liefern, und im selben Sinne natürlich auch kulturell-musikalische Codes, die verständniserleichternde Schemata evozieren.⁵⁹

Eine zweite Klasse von Relationen wurde vor allem von Chion beschrieben.⁶⁰ Sogenannte kausale Verknüpfungen erwachsen aus kulturell erlernten Zuschreibungsverhältnissen zwischen Ton und Bild, aber auch auf perzeptueller Ebene durch Gleichzeitigkeit und sind deswegen entlang eines Kontinuums von komplementärer Objektdarstellung, über konditionierte Relationen bis hin zu purer perzeptueller Ähnlichkeit angesiedelt. Besonders letztere Bindungen ‚funktionieren‘ lediglich aufgrund von zeitlicher Koinzidenz, aber eben nicht aufgrund von semantischer Logik.

Die dritte Klasse von Relationen unterscheidet sich grundsätzlich von den anderen beiden, da die hierbei eingesetzten auditiven Gestaltungselemente eben nicht auf sprachlich explizierbare Elemente der Narrationshandlung verweisen, sondern in einem impliziten, non-repräsentationalen Verhältnis zur dargestellten Handlung stehen. Dazu gehören handlungsstrukturierende Gestaltungsmittel, die Bewegungen strukturisomorph illustrieren, klangliche Interpunktio-nen und das musikalische oder klangliche Generieren von gespannten Erwartungshaltungen. Dabei nimmt gerade Musik eine zentrale Rolle ein. Aber auch punktuelle Geräusche, auditive ‚Signals‘, haben sich dabei in der multimedialen Instruktion als wertvoll erwiesen. Solche deiktischen Elemente sind, eingedenk der eingangs dargestellten multimodalen Integrationsprinzipien, hochgradig

58 Bishop/Cates: „Theoretical foundations for sound’s use in multimedia instruction to enhance learning“.

59 Claudia Bullerjahn: *Grundlagen der Wirkung von Filmmusik*. Augsburg 2001. S. 218.

60 Chion: *Audio-Vision. Sound on Screen*.

zeitkritisch. Im Gegensatz dazu können kulturunabhängige affektiv-konnotative Färbungen der Narration über Audio auch zeitlich etwas ungenauer positioniert werden, da Stimmungseffekte, wie durch das Excitation-Transfer-Paradigma bekannt, eher diffus und großflächig Wirkung zeitigen.

Im Ganzen (Abb. 3) sollte, auch im Abgleich mit den eingangs vorgestellten neurokognitiven Befunden, deutlich geworden sein, dass Audiotheoretiker wie Chion offenbar schon länger auf der richtigen Spur waren, als sie prinzipiell in Zweifel zogen., dass ein rein semantisches Verhältnis zwischen Audio und Vision das einzig sinnvoll ‚bedeutungsbildende‘ Verhältnis sein sollte. Mehr noch: An der Betrachtung von Audioeinsatz wird so offensichtlich, was für non-referentielle visuelle Darstellungselemente natürlich ebenfalls schon immer galt. Es existieren einfach auf verschiedenen kognitiven Verarbeitungsstufen unterschiedliche Bindungsstrategien für multimodale Reize und damit auch unterschiedliche Arten, instruktiv Bedeutung zu erzeugen. Diese Tatsache sollte nach Einschätzung unserer Arbeitsgruppe von einschlägigen Design- und Evaluations-Modellen endlich angemessen gewürdigt werden.

Ein Multi-Level-Forschungs- und Analysemodell für die Audio-Evaluation

Egal, ob es nun um Redundanz, Koinzidenz, oder Kookkurrenz geht: Multimodale Bedeutung wird, darauf haben schon Bishop und Cates in ihrem Modell für Audio in der Instruktion⁶¹ und auch Timo Saari hingewiesen,⁶² nicht nur auf unterschiedliche Arten, sondern, nach allem was wir wissen, offenbar im menschlichen Gehirn auf mindestens drei Ebenen konstituiert (Abb. 3): Auf der Ebene der frühen, präattentiven Perzeption, auf der Ebene der Aufmerksamkeit und Objektidentifikation und auf der Ebene der semantischen Verarbeitung und Speicherung.

Auf der ersten Ebene spielen Mechanismen der Kookkurrenz in Zeit und Raum die entscheidende Rolle, also besonders die Prinzipien der kausalen Bedeutungserzeugung über Audio. Auf der zweiten Ebene üben die Salienz⁶³ der dargebotenen Reize, also noch genauer zu untersuchende multimodale Figur-Grund-Trennungsmechanismen⁶⁴, aber auch kognitiver Stil und kognitive

61 Bishop/Cates: „Theoretical foundations for sound’s use in multimedia instruction to enhance learning“.

62 Timo Saari: *Mind-Based Media and Communication Technologies. How the Form of Symbolical Information Influences Felt Meaning*. University of Tampere, Tampere (SU/FIN) 2001. Dissertation.

63 Salienz = Hervorstechen, Prägnanz.

64 Der Auffassung einiger medienpädagogischer Autoren, das Aufmerksamkeitsablenkung durch den „Verbrauch“ von unimodalen, sensorischen Aufmerksamkeits-„Ressourcen“ entsteht, setzen wir die Auffassung von Neumann (Odmar Neumann: „Die Hypothese begrenzter Kapazität und die Funktionen der Aufmerksamkeit“. In: Neumann (Hg.): *Perspektiven der Kognitionspsychologie*. Berlin 1985, S. 185–229), Pashler (Harold E. Pashler: *The Psychology of*

Fähigkeiten eine bestimmende Funktion aus. Erst auf der dritten Ebene kommen die bereits bekannten Prinzipien der semantischen Kohärenz ins Spiel, für die didaktische Modelle existieren, welche, wie wir gesehen haben, vor allem auf Kongruenz als sinnstiftendem Prinzip abheben.

Im Sinne der gezielten Aufmerksamkeitssteuerung stellt funktional adäquates Lernmaterial nun, unserem Ansatz nach, die globale Kohärenz der Narration auf allen drei beschriebenen Systemebenen lokal her. Folglich sind daher Kohärenz und lernmaterialbezogene Ablenkungseffekte („Serependity“) nur zwei Seiten derselben Medaille. Eine kohärente multimodale Wahrnehmung entsteht⁶⁵ durch eine kohärente Verwendung multimodaler Gestaltungsmittel in Bezug auf die Narration, jedoch auf allen drei beschriebenen Systemebenen. Dabei verstehen wir globale mediale Text-Kohärenz als möglichst große Überlappung zwischen intendierter Bedeutung des Designers und konstruierter Bedeutung des Rezipienten, nicht wie in strukturalistischen Ansätzen als objektiv ‚richtiges‘ Textverständnis.

Für die theoretische Analyse multimedialer und filmischer Produkte bedeutet dies, gezielter als bisher nach „Added-Value“-Phänomenen multimodalen Zusammenspiels zu forschen; dies sollte natürlich auch anhand der Kombination bildbezogener und audiobezogener, und vor allem dabei zeitlich exakter Sequenzgrafiken geschehen. Der empirische Forschungsansatz besteht parallel dazu darin, innerhalb des beschriebenen Multi-Level-Modells Audioeinsatz szenariobasiert experimentell daraufhin zu überprüfen, ob er seine ‚Funktion‘, der Bereitstellung eines somit effektiven ‚multimodalen Konstruktionsangebotes‘ durch seine ‚Form‘ in einem ausreichenden Maße erfüllt oder ihr zuwiderläuft. Zweitens, um die reine Audioebene zu verlassen, ist in Anlehnung an die Ausführungen zu neurokognitiven crossmodalen Studien zu ermitteln, in welchem Zusammenhang kognitive Fähigkeiten bzw. situativ ermittelbare Kognitionsstile der Rezipienten mit unterschiedlich gewichteten multimedialen Darstellungsformen in Bezug auf Qualität, Prägnanz und Quantität Effekte zeitigen.

Attention. Cambridge 1998) und anderen Kognitionsforschern entgegen, dass Aufmerksamkeit nach aktuellen neurokognitiven Studien vor allem ein *objektbezogenes, supramodales* Prinzip der volitionalen Handlungssteuerung sein dürfte. Solange also „akzessorische“ Reize nicht perzeptuell oder semantisch salienter als die aktuell dominierende Handlung und Modalität werden, dürfte dies keine Ablenkungseffekte bewirken. Damit sollte auch klar sein, dass im intentionalen kognitiven Stil „Lesen und Betrachten“ die Aufmerksamkeit keinesfalls unwillkürlich durch zurückhaltende, leise „empathische“ Hintergrundmusik oder „Atmo“ abgelenkt werden kann, was die Auffassung von z. B. Klaus-Ernst Behne (Klaus-Ernst Behne: „Musik-Erleben: Abnutzung durch Überangebot“. In: *Media Perspektiven* (2001), Nr. 3, S. 142–148.) einer „Nichtwirksamkeit von Hintergrundmusik“ (im Sinne direkt messbarer nachteiliger oder positiver Effekte) indirekt zu bestätigen scheint.

65 Siehe auch: Mayer: *Multimedia Learning*.

Dabei steht die multimediale Narration als Mittel der Wissensrepräsentation ausdrücklich im Mittelpunkt. Viele Lernforscher sind analog dazu inzwischen der Auffassung, dass fast alles Lernen über Erzählungen funktioniert,⁶⁶ da Menschen gewohnt sind, in solchen ihre Umwelt und ihr Selbstbild zu konstruieren.⁶⁷ Das Deuten der Welt über Kausalität und Teleologien scheint ein Grundbedürfnis der menschlichen Weltinterpretation zu sein. Klang linearisiert und vektorisiert, wie wir wissen, Erfahrung, bringt überhaupt erst Zeitlichkeit in filmische Erzählung hinein, wie Chion schreibt. Insofern ist Audio also ein wichtiges, vielleicht *das* bedeutende Werkzeug bei der Erzeugung narrativer Wissensstrukturen und damit gesamtgesellschaftlicher Wissensteilhaber über Medien.

Resultierender methodischer Ansatz für empirische Forschungsarbeit

In der methodischen Praxis bedeuten diese Überlegungen, determinierende Faktoren bei der multimedialen Kohärenzbildung an typischen Szenarien im Experiment und gleichzeitig zur Erhöhung der externen Validität natürlich auch in Feldstudien und zwar auf allen drei im Modell dargestellten Ebenen zu untersuchen (Abb. 3). Das besagt methodisch, die Wirkungen impliziten Audioeinsatzes vor allem mit Hilfe der qualitativen Inhaltsanalyse und Rezipientenaussagen nach dem Free-Recall-Paradigma zu erfassen. Die Untersuchungen von Claudia Bullerjahn im Hinblick auf die Wirkung von Filmmusik scheinen hier wegweisend.⁶⁸

Im Zentrum stehen des Weiteren bei einem solchen Ansatz, neben der quantitativen und qualitativen Untersuchung und Manipulation des Lehrmaterials in Bezug auf das multimodale Zusammenspiel zwischen Audio und Vision, besonders Fragestellungen nach dem Einfluss von Lernerdispositionen, wie kognitivem Stil, dem räumlichen Vorstellungsvermögen, musikalischer Vorbildung und Mediennutzungsfrequenz.

Auch die Untersuchung von unmittelbaren Perzeptionswirkungen anhand ereigniskorrelierter Potentiale⁶⁹ an ausgewählten, typischen Szenarien kann hier sicherlich einen entscheidenden Beitrag für das Verständnis der Prozesse auf der ersten System-Ebene darstellen. Einen zusätzlich wichtigen Stellenwert in Bezug auf den Einfluss von ästhetischen Qualitäten und Unterhaltungsqualitäten im Zusammenspiel mit multimedialem Lernen sollten dabei auch weiterhin Un-

66 Vgl. hierzu z. B. Gary A. Berg: „Cognitive development through narrative: Computer interface design for educational purposes“. In: *Journal of Educational Multimedia and Hypermedia* (2000), Nr. 9 (1), S. 3–17.

67 Kay Young/Jeffrey L. Saver: „The Neurology of Narrative“. In: *SubStance* (2001), Nr. 94/95, S. 72–84.

68 Claudia Bullerjahn: „Film music and cinematic narration: a report on an empirical investigation into the effects of film music“. *3rd International Conference for Music Perception and Cognition*, Lüttich 1994.

69 EEG-Messungen lokaler Spannungsveränderungen an der Schädeloberfläche, bezogen auf den Zeitpunkt der Reizdarbietung.

tersuchungen der Nutzer-Akzeptanz, aber auch der Stabilität und Konsolidierung von Wissensformationen anhand unterschiedlicher Messzeitpunkte und wiederholter Messungen einnehmen. Dies gilt ebenfalls im Rahmen von daran unmittelbar anschließenden quantitativen Feldstudien, wie die CRIMP-Gruppe sie bereits in Bezug auf den Nutzen dreidimensionaler Animationen an Schulklassen durchgeführt hat,⁷⁰ welche dann anhand anwendungsbezogener Szenarien den Nutzen der im Experiment ermittelten heuristischen Prinzipien zahlenmäßig und im realen Anwendungsfall belegen können.

Fazit: Multimediale Erzählung als multimodales Gestaltproblem

Soweit zu dem von unserer Arbeitsgruppe vertretenen Forschungsrahmenmodell zur Analyse von Audioeinsatz im Kontext multimedialer instruktiver Narrationen. Es liegt auf der Hand, dass auf ganz ähnliche Art Audio auch im nicht-pädagogischen, unterhaltenden oder künstlerischen Kontext z. B. im Medium Film betrachtet werden kann, wie ja auch die Ähnlichkeit des Ansatzes zu den Theorien von Chion nahe legt. In einem solchen Zusammenhang wäre es sicherlich auch hochinteressant, ähnlich wie bei CRIMP, experimentelle Manipulationen an der auditiven Oberflächenstruktur der filmischen Narration vorzunehmen, wie schon in einigen Experimenten zur Filmmusikwirkung geschehen.⁷¹

Lydia Plowman hat bereits daraufhin gewiesen, dass die aktuell beobachtbaren medienpädagogischen Diskurse über die Ausdrucksfähigkeiten von Multimedia Debatten in der historischen Phase des frühen Films stark ähneln.⁷² Es gibt aber noch weitere Berührungspunkte zur Filmwissenschaft: Im Grundansatz entspricht die hier geforderte Herangehensweise sehr stark der bereits von einigen Autoren⁷³ eingeforderten Fusion kognitiver Lernmodelle mit kognitiven Ansätzen zur narrativen Filmanalyse (z. B. Bordwell/Branigan/Carroll/Ohler), versucht allerdings dabei, diese intensiv in Bezug auf nicht-repräsentationale Narrationselemente, speziell Audio, auszubauen.⁷⁴ Gleichzeitig werden im Gegensatz zur Analyse populärkultureller oder künstlerischer Filme in unserem Modell ‚Verständlichkeit‘ und ‚Kohärenz‘ als funktionale Primärziele definiert. In der Umkehrung kann dieser Ansatz deshalb natürlich auch

70 Thomas Huk/Christian Floto: „Criteria of Evaluation of Audiovisuals in Multimedia Production (CRIMP)“. *EDEN Second Research Workshop*, Hildesheim 2002.

71 Vgl. z. B. Bullerjahn: *Grundlagen der Wirkung von Filmmusik*.

72 Lydia Plowman: ‚The ‚primitive mode of representation‘ and the evolution of interactive multimedia“. In: *Journal of Educational Multimedia and Hypermedia* (1994), Nr. 3 (3/4), S. 275–293.

73 Vgl. z. B. Ben-Shaul: ‚Split attention problems in interactive moving audiovisual texts“.

74 Vgl. bezüglich der Aktualität dieses Ansatzes: Iris van Rooij/Raoul M. Bongers/W. Pim F. G. Haselager: ‚A non-representational approach to imagined action“. In: *Cognitive Science* (2002), Nr. 26, S. 345–375.

dazu dienen, innerhalb künstlerischer Produktionen Ambiguierungs- oder ‚Verschlüsselungsstrategien‘ bzw. das Erzeugen von intrinsischer Motivation, Unterhaltung und Suspense zu beschreiben.

Im Kern wendet sich das Modell damit, wie schon bei einigen anderen Autoren innerhalb der multimedialen Instruktionstheorie geschehen,⁷⁵ vom logozentrischen, strukturalistischen Ansatz der propositionalen Repräsentation als adäquater Beschreibung mentaler Modelle bei der multimedialen Bedeutungserzeugung ab, der Narration als Prinzip der Wissensvermittlung zu und nimmt damit eine funktionalistische Perspektive in der Evaluation medialer Lernprodukte ein. Multimediale Narration als spezifisch ästhetisch-didaktische Ausdrucksform und die Design-Techniken, welche zu ihrem adäquaten Verständnis durch den Rezipienten beitragen, stehen dabei im Mittelpunkt der Analyse.

Damit folgt unser Ansatz nicht nur Forderungen nach tatsächlich konstruktivistischen Lernmodellen, sondern lehnt sich auch argumentativ an die Prager Schule des linguistischen Funktionalismus an: Diese geht, wie schon Aristoteles in seiner *Poetik* (s.o.), von der *Funktion* der Äußerung aus, um ihre *Form* zu beschreiben. Ganz analog spekuliert das Modell nicht wahllos, wie bisherige Multimedia-Theorieansätze, über quantitativ beschränkte Gehirnfunktionen als Ursache für den Misserfolg didaktischer Interventionen, sondern betrachtet diejenigen lokalen auditiven Gestaltungsmittel als situationsadäquat und individuell angemessen, die mit Hilfe ihrer semantischen Bedeutung *und* ihrer Form beim richtigen Verständnis der Gesamtnarration im Sinne der intendierten globalen Funktion des medialen Textes nachweisbar helfen; das Modell hat also vom Ansatz her auch eine Verwandtschaft mit der Gestalttheorie. Die Analyse der Effekte des frühen perzeptuellen crossmodalen Zusammenspiels von Audio und Vision im interaktiven Zusammenspiel – und eben nicht getrennt auf zwei Ebenen – kann hierbei eine wertvolle Bereicherung bieten, zeigt sie doch basale konstruktivistische Prinzipien erfolgreicher filmischer und multimedialer Narrationen auf.

75 Vgl. Berg: „Cognitive development through narrative: Computer interface design for educational purposes“.