

# Das Modalitätsprinzip im multimedialen Instruktionsdesign: Empirische Basis, theoretische Erklärungen und vernachlässigte Faktoren.<sup>1 2</sup>

Klaus D. Stiller  
Institut für Experimentelle Psychologie  
Universität Regensburg  
Germany  
klaus.stiller@uni-regensburg.de

**Zusammenfassung:** Das Modalitätsprinzip ist eine empirisch fundierte Gestaltungsrichtlinie für Lehrmaterialien aus Texten und Bildern. Es empfiehlt Erläuterungen zu Visualisierungen gesprochen anstatt geschrieben zu präsentieren. Die bevorzugten Erklärungen führen die empirisch festgestellten Modalitätseffekte auf eine übermäßige Belastung des visuell-räumlichen Subsystems des Arbeitsspeichers durch die gleichzeitige Verarbeitung geschriebener Texte und Bilder oder auf den Effekt der geteilten visuellen Aufmerksamkeit zwischen diesen Informationsquellen zurück. Weiterhin werden Lernbeeinträchtigungen durch die notwendigen Blickbewegungen beim Lesen und in Spezialfällen der Vorteil auditiver Texte durch die längere Verfügbarkeit von Gesprochenem im auditiven Register berücksichtigt. Erklärungen durch die zeitliche Trennung aufeinanderfolgender Informationseinheiten, sowie motivationsförderliche Eigenschaften gesprochener Texte werden kaum herangezogen. Es werden die Modalitätseffekte konstituierenden Faktoren aufgezeigt und Folgerungen für Theorie und Empirie gezogen.

**Abstract:** The modality principle is an empirically based rule for designing instructions composed of text and pictures, i.e. multimedia instructions. It is recommended to explain visualizations by narration instead of visual texts. The preferred explanations refer the empirically observable effects to a higher workload of the visual-spatial subsystem of the working memory by the processing of visual texts or to splitting visual attention between visual texts and pictures. Furthermore, explanations consider learning disruptions by inevitable eye movements while reading. Some special instances of modality effects might be referred to a longer sustain of speech in the auditory register in comparison to read information in the visual register. Temporal split-attention effects, e.g. between multimedia information units, as well as motivation enhancing features of narration as sources for multimedia effects were not satisfactorily considered so far. The contributing factors to modality effects are highlighted and conclusions concerning theory and research are drawn.

Es gibt eine Reihe von Arbeiten, welche Modalitätseffekte in der Multimediaforschung theoretisch und empirisch elaborieren und die Rahmenbedingungen seiner Gültigkeit erforschen. Nachfolgend wird zuerst das Modalitätsprinzip angeführt und seine empirische Basis charakterisiert. Standarderklärungen für Modali-

---

<sup>1</sup> Stiller, K. D. (2011). *Das Modalitätsprinzip im multimedialen Instruktionsdesign: Empirische Basis, theoretische Erklärungen und vernachlässigte Faktoren*. Verfügbar unter <http://epub.uni-regensburg.de/22508/>

<sup>2</sup> Dieser Artikel wurde in seiner Ursprungsform im Oktober 2009 fertig gestellt und seitdem inhaltlich im Wesentlichen nicht verändert. Aus diesem Grunde sind viele neuere empirische Artikel zum Modalitätseffekt nicht berücksichtigt, ebenso assoziierte theoretische Verfeinerungen. Die Intention des Artikels bestand darin, die empirische Basis zum Modalitätseffekt festzuhalten, eine bis dahin fehlende Übersicht über theoretische Erklärungsmuster bereitzustellen und die damals aktuelle und unbearbeitete Frage nach Randbedingungen des Modalitätseffekts zu stimulieren. Diese Intention fand seit 2009 bei Gutachtern diverser deutscher Journale keinen Anklang (mal mit wenig Kritik, mal mit viel Kritik, aber mit durchaus unterschiedlichen Ablehnungsgründen), vielmehr wird heutzutage eine differenziertere, bewertende und integrierendere Ausarbeitung gewünscht, wobei ich persönlich nicht sehe, in welchem deutschen Journal dies tatsächlich verständlich und entsprechend exakt angesichts der Umfangbeschränkungen von Artikeln geleistet werden kann. Nun ist der Artikel der Allgemeinheit zur Verfügung gestellt, mit der Hoffnung, dass es trotz dem Scheitern an einer vorzeigbaren Veröffentlichung dem einen oder anderen helfen mag.

tätseffekte sowie Spezialerklärungen, die nur im Einzelfall zutreffen oder ausgewählte Teilerklärungen liefern, folgen. Im Anschluss wird eine motivationale Forschungsrichtung im multimedialen Instruktionsdesign aufgegriffen, welche zur Differenzierung des Modalitätseffekts beiträgt, aber bisher weitgehend vernachlässigt wurde.

## **Das Modalitätsprinzip und seine empirische Basis**

Das Modalitätsprinzip ist eine Handlungsanweisung zur effektiven Gestaltung von Lehrmaterialien, bestehend aus Texten und Bildern. Es empfiehlt, Erläuterungen zu dynamischen und statischen Visualisierungen gesprochen anstatt geschrieben zu präsentieren (Low & Sweller, 2005; Mayer, 2005b). Diese Empfehlung wurde aus vielfältigen empirischen Belegen abgeleitet, die zeigen, dass Lernende effektiver mit gedruckten und computerpräsentierten Lehrmaterialien lernen (z.B. Animation zur Blitzentstehung, mathematische Lösungsbeispiele, computerbasierte Simulation einer biologischen Umwelt, Computerlernprogramm zum Elektromotor), wenn gesprochene anstatt geschriebene Texte verwendet werden (Ginns, 2005; Low & Sweller, 2005). Vorausgesetzt wird dabei, dass sowohl die bildhaften als auch die verbalen Informationen nicht redundant sind und deshalb von den Lernenden gemeinsam verarbeitet (d.h. integriert) werden müssen, damit sie die Lerninhalte vollends verstehen.

Die Überlegenheit gesprochener Texte zeigte sich dabei in einer Vielzahl von Indikatoren: Weniger mentale Anstrengung während des Lernens (Tabbers, 2002; Tindall-Ford et al., 1997; Van Gerven, 2002), kürzere Lösungs- oder Beantwortungszeiten von Problemlöse- und Wissensaufgaben (Jeung et al., 1997; Mousavi et al., 1995; Rinck & Glowalla, 1996), mehr Erfolg bei unterschiedlichen Wissensaufgaben wie dem Behalten und Erinnern von Faktenwissen, der Zuordnung von Bild- und Textinformationen, dem Beschriften von Bildern, dem Lerntransfer und der praktischen Anwendung des Gelernten (Kalyuga et al., 1999, 2000; Craig et al., 2002; Mayer & Moreno, 1998; Moreno & Mayer, 1999; Schmidt-Weigand, 2006; Stiller, 2007) als auch kürzere Reaktionszeiten in einer neben dem Lernen applizierten Zweitaufgabe (Brünken et al., 2002, 2004).

Ältere Studien, welche Modalitätseffekte zu Gunsten gesprochener Texte zeigten, sind bezüglich der Inhaltsbereiche, der Lernzeit, der Lernsettings und des Präsentationsablaufs beschränkt. Die Inhalte wurden hauptsächlich aus den Formal-/Idealwissenschaften, sowie den Naturwissenschaften entlehnt (Geometrie, Mathematik, Physik), die Instruktionszeiten lagen unter fünf Minuten, die Studien wurden im Labor durchgeführt und der Präsentationsablauf (Geschwindigkeit sowie Sequenz der Informationen) war vorgegeben (z.B. Craig et al., 2002; De Westelinck et al., 2005; Mayer, 2001; Mousavi et al., 1995). Seitdem bestätigten sich aber die Befunde auch für im Zeitablauf und in der Informationssequenz lernergesteuerte Lehrinheiten, welche bis zu 90 Minuten Lernzeit erforderten. Dabei wurde der Effekt in realitätsnahen Klassenräumen sowie im Labor untersucht, wobei auch nicht-wissenschaftliche Inhalte verwendet wurden (z.B. Brünken et al., 2005; Mayer, Dow et al., 2003; Moreno & Mayer, 2002; Moreno et al., 2001; O'Neil et al., 2000; Stiller, 2007; Tabbers, 2002). Dies zeigt, dass die bestätigenden Befunde von Modalitätseffekten mittlerweile aus einer Vielzahl von unterschiedlichsten Lernsituationen hervorgegangen sind.

## **Standard- und Spezialerklärungen**

### **Visueller Split-Attention Effekt**

Die bevorzugte Erklärung für Modalitätseffekte bezieht sich auf den Effekt der geteilten visuellen Aufmerksamkeit. Bei visueller Textpräsentation muss die Aufmerksamkeit zwischen Bildern und Texten aufgeteilt werden, damit die relevanten Informationen aufgenommen, adäquat organisiert und integriert werden können. Die Aufmerksamkeitsaufteilung belastet das Arbeitsgedächtnis, z.B. durch visuelle Suchprozesse der referenzierten bildhaften oder verbalen Informationen (Sweller, 1999). Damit erhöht sich die Wahrscheinlichkeit, relevante Informationen aus Texten und Bildern zu übersehen. Dies gilt besonders bei Instruktionen mit Bewegtbildern und systemgesteuerten Präsentationen ohne Lernereinfluss. Wenn relevante Informationen fehlen, erfolgt die anschließende Konstruktion adäquaten Wissens lückenhaft und unvollständig.

Neben dem visuellen Split-Attention Effekt tritt bei einer Aufteilung der Aufmerksamkeit zwischen zwei visuellen Informationsquellen auch ein zeitlicher Effekt der geteilten Aufmerksamkeit auf (Rummer et al.,

2008; Stiller, 2007). Die Auswahl relevanter Informationen aus zwei visuellen Informationsquellen kann nur sequenziell erfolgen. Wenn die Lernenden Informationen aus dem gelesenen Text ziehen, müssen diese so lange im Arbeitsgedächtnis gehalten werden, bis die korrespondierenden Bildinformationen gefunden werden, und umgekehrt. Erst dann können diese Informationen integriert werden. Diese verlängerte Aktivierung von Informationen erhöht die Belastung des Arbeitsgedächtnisses und kann dazu führen, dass relevante Informationen aus dem Arbeitsgedächtnis verloren gehen, bis die entsprechende Information aus der komplementären Informationsquelle gefunden wurde.

Demgegenüber ermöglichen gesprochene Texte eine simultane Aufnahme der Text- und Bildinformationen und schaffen somit eine optimale raum-zeitliche Kontiguität, welche eine visuelle Aufteilung der Aufmerksamkeit beseitigt und eine zeitliche Aufteilung der Aufmerksamkeit reduzieren kann. Verbale und bildhafte Informationen können zeitgleich und parallel ausgewählt sowie unmittelbar aufeinander bezogen werden. Gesprochene Texte entlasten das Arbeitsgedächtnis der Lernenden, indem sie z.B. übermäßige Such- und assoziierte Prozesse reduzieren und Informationen nicht länger als notwendig im Arbeitsgedächtnis aktiviert gehalten werden müssen, um die korrespondierenden Informationen zu verknüpfen.

Gesprochene Erläuterungen zu Bildern können weiterhin als effektive Strategie zur Förderung der Bildverarbeitung angesehen werden. Zum Einen haben Blickbewegungsstudien gezeigt, dass störende Fixationswechsel zwischen Bildern und geschriebenen Texten auftreten, und somit Bilder nicht ungestört wahrgenommen und verarbeitet werden können (Rayner, 1998; Rayner & Pollatsek, 1987, 1989; Tabbers, 2002). Zum Anderen erhöhen gesprochene Texte die Verarbeitungszeit der Bilder, während bei geschriebenen Texten mehr Zeit mit dem Lesen verbracht wird (Hasebrook, 1994; Rinck & Glowalla, 1995; Tabbers, 2002). Die zusätzliche Zeit für die Bilder wird – so die Vermutung – von den Lernenden zu einer intensiveren Bildverarbeitung genutzt. Hierdurch werden mehr relevante Bildinformationen aufgenommen, welche im Weiteren auch zu einer adäquateren Repräsentation der Inhalte führen und effektiver ans Vorwissen angeknüpft werden.

## **Überlastung des Arbeitsgedächtnisses und Verarbeitungsstörungen**

Als weiteres Argument für Modalitätseffekte wird angeführt, dass visuelle Texte und Bilder dasselbe für die Verarbeitung visueller Reize zuständige Subsystem des Arbeitsgedächtnisses (siehe Baddeley, 1997) benutzen und dieses dadurch überlastet werden könne. Dagegen würden auditive Texte im für die auditive Reizverarbeitung zuständigen Subsystem verarbeitet. Auditive Texte erschließen damit eine zusätzliche Ressource und verringern somit die Wahrscheinlichkeit einer Überlastung des Arbeitsgedächtnisses (Rumner et al., 2008). So nimmt z.B. Mayer (2005a) in seiner Cognitive Theory of Multimedia Learning an, dass visuelle Texte zumindest anfänglich denselben visuell-bildhaften Verarbeitungskanal benutzen wie Bildinformationen und daher eine Überlastung dieses Kanals möglich ist, während auditive Texte sofort in den entsprechend auditiv-verbale Kanal gelangen und damit nicht um dieselben Ressourcen konkurrieren. Ähnlich wird innerhalb der Cognitive Load Theory (Sweller, 1999) in Anlehnung an Baddeleys Konzeption des Arbeitsgedächtnisses (1997) argumentiert, dass visuelle Textinformationen mit Bildinformationen im visuell-räumlichen Speicher um Ressourcen konkurrieren, während gesprochene Texte in der artikulatorischen Wiederholungsschleife verarbeitet werden. In der Multimediaforschung wird jedoch darauf hingewiesen, dass diese Annahme in Baddeleys Theorie nicht enthalten sei und bei geübten Lesern durch den artikulatorischen Rehearsalprozess die visuellen Texte unmittelbar aus dem sensorischen Register in den phonologischen Speicher transferiert würden. Insgesamt wird dieser Annahme der Überlastung des räumlich-visuellen Arbeitsgedächtnisses wenig Gewicht zugeschrieben (z.B. Tabbers, 2002) und von Rumner et al. (2008) deutlich kritisiert.

Dennoch wird in der Literatur darauf verwiesen, dass die Verarbeitung graphemischer Informationen und die Blickbewegungssteuerung das visuell-räumliche Arbeitsgedächtnis belasten (Baddeley, 1986; Kürschner & Schnotz, 2008). Im Besonderen verweisen Kürschner und Schnotz (2008) in Bezug auf das Verstehen von Texten darauf, dass bei der Konstruktion visuell-räumlicher Inhalte aufgrund der visuellen Informationsaufnahme und der damit verbundenen Blickbewegungssteuerung eher die Gefahr einer Überlastung des visuell-räumlichen Arbeitsgedächtnisses besteht, auch wenn sowohl gesprochene als auch geschriebene Texte im phonologischen Speicher verarbeitet werden. Ferner würde die Modalität eines Textes unter bestimmten Verarbeitungsbedingungen die Repräsentation visuell-räumlicher Inhalte aufgrund unterschiedlich starker Belastung der verschiedenen Subsysteme des Arbeitsgedächtnisses beeinflussen (Kürschner & Schnotz, 2008).

Baddeley (1997) verweist darauf, dass auch physikalische Merkmale gesprochener und geschriebener Texte im Langzeitgedächtnis kodiert werden können. Bei Lesetexten können die visuellen Zeichen zumindest teilweise vom visuell-sensorischen Register in das visuell-räumliche System des Arbeitsgedächtnisses

gelangen, dort Ressourcen belegen und in der Folge den Lernprozess behindern. Auch Kürschner und Schnotz (2008) gehen in ihrem Modell des integrierten Text- und Hörverstehens davon aus, dass mentale Repräsentationen (auditiv und visuell basierte Oberflächenrepräsentationen) im Arbeitsgedächtnis durchaus modalitätsspezifische Merkmale (physikalische Merkmale der gesprochenen oder geschriebenen Texten) besitzen (vgl. auch Baddeley, 1997; Engelkamp, 1990). Diese Merkmale müssen daher in den entsprechend zuständigen Subsystemen des Arbeitsgedächtnisses verarbeitet worden sein, damit sie Teil einer mentalen Repräsentation werden können. Weiterhin merken Rummer et al. (2008) an, dass mit dem Lesen einhergehende Prozesse (z.B. Blickbewegungen und deren Steuerungen) das Bildbehalten beeinträchtigen und somit zu einer Reduktion der Lernleistung, im Vergleich zu auditiver Textdarbietung, führen können.

Insgesamt erscheint die Annahme einer Überlastung des visuell-räumlichen Arbeitsspeichers durch die inhaltlich bezogene Verarbeitung visueller Texte nicht haltbar. Eine Belastung durch die Verarbeitung der visuellen Merkmale der Texte sowie durch Blickbewegungssteuerungen aus Gründen des Split-Attention-Phänomens als auch der Notwendigkeit, Sakkaden und Fixationen beim Lesen ausführen zu müssen, scheint sich aber durchaus einzustellen. Diese Belastung wirkt sich gegebenenfalls störend auf Wissenskonstruktionsprozesse, insbesondere visuell-räumlicher Repräsentationen, aus.

### **Verfügbarkeit von Informationen im akkustisch-sensorischen Register**

Rummer et al. (2008) führen an, dass unter bestimmten Bedingungen der in der gedächtnispsychologischen Forschung bekannte Modalitätseffekt auftritt und zu einer Überlegenheit von auditiven gegenüber visuellen Texten führen kann (siehe auch Penney, 1989). Der gedächtnispsychologische Modalitätseffekt besagt, dass identisches verbales Material (in der Regel unverbundene Buchstaben, Ziffern oder Wörter) besser behalten wird, wenn es auditiv statt visuell dargeboten wird. Grundlegend dafür ist, dass auditiv-sensorische Informationen im sensorischen Register länger verfügbar sind als visuell-sensorische Informationen. Des Weiteren impliziert der Effekt, dass die letzten Items einer Liste von Items aufgrund der Nutzung von akustisch-sensorischer Informationen des sensorischen Registers länger verfügbar sind als visuell-sensorische Informationen. Entsprechend resümieren Rummer et al. (2008), dass der Modalitätseffekt unter speziellen Bedingungen, etwa wenn Texte aus jeweils nur einem Satz bestehen (siehe z.B. Moreno & Mayer, 1999; Mousavi et al., 1995), auf die Nutzung akustisch-sensorischer Informationen zurückgehen kann. Lernende haben sozusagen bei auditiver Textpräsentation mehr Zeit, relevante Informationen zu selektieren und einer weitergehenden Verarbeitung zuzuführen.

### **Unterschiede im Lese- und Hörverstehen**

Kürschner und Schnotz (2008) beschreiben prinzipielle Unterschiede beim Lese- und Hörverstehen und den mentalen Repräsentationen, welche aus modalitätsspezifischen Gründen entstehen. Durch die physikalischen Unterschiede zwischen geschriebenen und gesprochenen Texten ergeben sich auf niedrigeren Verarbeitungsebenen des Textverstehens, wie etwa bei basalen Wahrnehmungsprozessen und der Identifikation von Buchstaben und Wörtern, sowie bei den Verarbeitungsstrategien, modalitätsspezifische Unterschiede.

Im Weiteren führen sie einen Vorteil bezüglich des kurzfristigen Behaltens einfachen verbalen Lernmaterials bei auditiver Präsentation von Texten an, der auf den längeren Nachhall von diesem Textmaterial im sensorischen Register zurückgeführt wird (siehe vorhergehenden Abschnitt). Zudem ist auditiv präsentiertes Lernmaterial weniger anfällig für externe Störungen als visuell präsentiertes Lernmaterial (Baddeley, 1986). Für das langfristige Behalten einfachen verbalen Lernmaterials sind die Ergebnisse uneinheitlich, wobei je nach Aufgabenanforderung die eine oder andere Modalität zu besserem Behalten führte. Untersuchungen zu längeren und semantisch reichhaltigeren Texten weisen darauf hin, dass die Modalität der Textpräsentationen unter bestimmten Verarbeitungsbedingungen spezifische Auswirkungen auf Behaltens- und Verstehensleistungen haben kann (Kürschner & Schnotz, 2008). So wurden mehrfach Vorteile des Leseverstehens gegenüber dem Hörverstehen bei der Wiedergabe von Detailwissen nachgewiesen, während das Hörverstehen unter bestimmten Bedingungen mehr das prinzipielle Verständnis von Lehrinhalten fördern kann (Kürschner & Schnotz, 2008). Insbesondere weisen Kürschner und Schnotz (2008) darauf hin, dass beim Lesen aufgrund der visuellen Informationsaufnahme und der damit verbundenen Blickbewegungssteuerung bei der Repräsentation visuell-räumlicher Inhalte eher die Gefahr einer Überlastung des visuell-räumlichen Arbeitsgedächtnisses besteht als beim Hören.

Zusammenfassend argumentieren Kürschner und Schnotz (2008), dass bei der Nutzung des Arbeitsgedächtnisses modalitätsbedingte Unterschiede vorhanden sind. Modalitätsbedingte Unterschiede auf den hierarchisch niedrigeren Verarbeitungsebenen nehmen Einfluss darauf, welche Informationen den Konstruktionsprozessen auf den höheren Ebenen zur Verfügung stehen. So können mentale Repräsentationen unterschiedlicher inhaltlicher Ausprägungen entstehen. Unterschiede auf den hierarchisch niedrigeren Verarbeitungsebenen übernehmen somit gewissermaßen eine Filterfunktion für die Konstruktionsprozesse auf den höheren Ebenen.

Angesichts der Unterschiede bei der Verarbeitung von Lese- und Hörtexten bleibt offen, inwieweit die instruktionspsychologischen Modalitätseffekte auf bildunabhängige Unterschiede der kognitiven Verarbeitung beim Hör- und Leseverstehen beruhen und inwieweit sie auf den visuellen Split-Attention Effekt zurückzuführen sind.

## **Vernachlässigte Faktoren**

### **Temporaler Split-Attention Effekt**

Effekte der geteilten zeitlichen Aufmerksamkeit treten auf, wenn Informationen aus zeitlich getrennten Quellen aufeinander bezogen und integriert werden müssen, damit ein Verständnis der präsentierten Inhalte erzielt werden kann. Der Fokus liegt dabei nicht auf der Integration zeitlich getrennter Bild- und Textquellen (siehe z.B. temporal contiguity effect; Moreno & Mayer, 1999), sondern auf Informationseinheiten, bestehend aus Bild und Text, welche zeitlich getrennt aufeinander folgen.

Eine Untersuchung von Stiller et al. (2009) zeigte, dass bei einer moderaten Präsentationsgeschwindigkeit der Text-Bild-Einheiten zeitliche Aufmerksamkeitseffekte zum Modalitätseffekt beitragen können. Lernende wurden durch einen zeitlichen Split-Attention Effekt in ihren Lernleistungen beeinträchtigt, wenn sie die visuell präsentierten Informationen schneller verarbeiten konnten, als diese vom Lehrsystem bereitgestellt wurden. In der Untersuchung wurden Lehrpräsentationen, bestehend aus 13 Informationseinheiten zum Aufbau des Auges, entweder mit gesprochenem oder geschriebenem Text systembestimmt (zeitlich vorgegeben) präsentiert oder lernergesteuert bearbeitet. Unter Lernersteuerung mussten die Lernenden durch Klicken auf einen „Weiter-Button“ die nächste der 13 Informationseinheiten selbständig abrufen. In Übereinstimmung mit der visuellen Split-Attention-Annahme ergab sich im Vergleich der systembestimmten Präsentationen ein Modalitätseffekt zugunsten gesprochener Texte. Allerdings zeigte sich, dass die mit visuellen Texten selbstgesteuert Lernenden im Vergleich zur systembestimmten Präsentation (1) durchschnittlich 80 Sekunden weniger Zeit für die Bearbeitung der Präsentation benötigten (es bearbeiteten 76 % der selbstgesteuert Lernenden die Präsentation schneller als die schnellste vorgegebene Lernzeit der systembestimmten Präsentation) und (2) sowohl signifikant besser im Leistungstest (Faktenwissen behalten, Strukturwissen verbalisieren, Bilder beschriften) abschnitten, als auch eine signifikant niedrigere mentale Anstrengung berichteten. Wäre der gefundene Modalitätseffekt bei den systembestimmten Präsentationen mit dem visuellen Split-Attention Effekt erklärt worden, hätte das den paradoxen Schluss zur Folge, dass durch eine schnellere lernergesteuerte Bearbeitung dieser kompensiert worden ist. Die Erklärung mittels einer Reduktion von zeitlichen Split-Attention-Effekten zwischen den 13 Lehreinheiten ist hingegen widerspruchsfrei mit der schnelleren Bearbeitungszeit unter Lernersteuerung im Vergleich zur systembestimmten Präsentationszeit vereinbar.

Diese zeitlichen Aufmerksamkeitseffekte sind bisher kaum als Erklärung von Modalitätseffekten herangezogen worden. Dies könnte daran liegen, dass die Inhalte in bisher verwendeten Lehrmaterialien meist sehr schnell präsentiert wurden und die Lernenden hauptsächlich darin herausforderten, alle relevanten Informationen innerhalb der zugestandenen Lehrzeit aufzunehmen, d.h. gegen den Effekt der geteilten visuellen Aufmerksamkeit zu arbeiten. Wenn die Lernenden dann selbst die Präsentationsgeschwindigkeit der Lehrmaterialien bestimmen konnten, so führte das meist zu signifikanten Lernzeitverlängerungen (z.B. Tabbers, 2002). Insgesamt ist es aber schwer, anhand der Ergebnisse von Experimenten mit nicht offensichtlich schnellen Präsentationen zu sagen, ob sie auf einen visuellen oder zeitlichen Effekt der Aufmerksamkeit zurückgehen, wenn die zeitliche Komponente nicht mit einbezogen wird, da sich die Vorhersagen unter der Annahme eines Effekts der getrennten visuellen oder zeitlichen Aufmerksamkeit in Bezug auf Lernleistungen und mentaler Anstrengung gleichen.

## Motivationale Effekte

Motivationale Effekte im Rahmen der Modalitätseffekte wurden bisher vernachlässigt, obwohl einige Designprinzipien und auch Theorien wie die Social Agency Theory (Mayer, 2005c) dies nahelegen. Als mehr oder weniger gut bestätigte Gestaltungsprinzipien oder -effekte können angeführt werden:

(1) Stimmenprinzip (voice principle): Personen lernen erfolgreicher, wenn Texte mit einer akzentfreien, menschlichen Stimme gesprochen werden. Ungünstig erwiesen sich entsprechend maschinelle Stimmen oder Stimmen mit einem fremdsprachlichen Akzent (Atkinson et al., 2005; Mayer, Sobko et al., 2003; Nass et al., 2003; Stern et al., 2006).

(2) Sprechergeschlecht Effekt (speaker's gender effect): Empirische Ergebnisse zeigen vereinzelt, dass Lernende Frauenstimmen positiver bewerteten und mit ihnen bessere Lernleistungen erzielten (Linek, 2007). Allerdings ist die Befundlage hierzu insgesamt noch spärlich und inkongruent (Hartley et al., 1989; Karayianni & Gardiner, 2003).

(3) Personalisierungsprinzip (personalisation principle): Dieses Prinzip postuliert, Erläuterungen zu Bildern in einem personalisierten anstelle eines formalen Stils vorzunehmen. Personalisieren bedeutet dabei, umgangssprachliche Formulierungen zu verwenden. Dafür stehen prinzipiell zwei Techniken zur Verfügung: „Ich“ und „Du“ und die zugehörigen Personalpronomen anstatt formaler Konstruktionen benutzen oder Sätze hinzufügen, welche sich direkt an den Lerner wenden. Empirisch wurde gezeigt, dass Lernende erfolgreicher mit personalisierten anstatt formalen Texten lernen (Ginns & Fraser, 2010; Mayer et al., 2004; Moreno & Mayer, 2000, 2004; Stiller & Jedlicka, 2010).

Für diese Designprinzipien wird die Social Agency Theory (Mayer, 2005c) als Erklärung von Effekten herangezogen. Ausgangspunkt dabei ist die An- oder Abwesenheit sozialer Hinweisreize. Soziale Hinweisreize können in Form von personalisierten Formulierungen (Personalisierungsprinzip), der Verwendung einer menschlichen Stimme (Stimmenprinzip) oder dem hörbaren Geschlecht des Sprechers (Sprechergeschlecht Effekt) auftreten. Derartige Hinweisreize führen beim Lernenden zur Aktivierung einer sozialen Reaktion, welche in der Folge die aktive, kognitive Informationsverarbeitung des Lernenden steigert. Eine höhere aktive, kognitive Verarbeitung des Lernenden führt dann zu besseren Leistungen in Lerntests.

Bisher ist der Zusammenhang zwischen den gerade beschriebenen Prinzipien/Effekten sowie der Social Agency Theory und den Modalitätseffekten noch unklar. Dies wird im Folgenden erläutert. Der Austausch geschriebener durch gesprochene Texte kann zugleich ein Hinzufügen von sozialen Hinweisreizen in Bezug auf die vorher aufgegriffenen Designprinzipien oder -effekte bedeuten.

(1) Stimmefeffekte können als Hinweis dafür angesehen werden, dass bei Modalitätseffekten motivationale Prozesse eine Rolle spielen. Wenn – rein didaktisch gesehen – visuelle Texte eher maschinell generierten Stimmen entsprechen, welche keine Stimmlage, keine Intonation, kein Geschlecht usw. aber auch keinen Akzent haben, dann werden durch einen Textmodalitätswechsel diese Komponenten eingeführt. Insgesamt deutet die empirische Befundlage darauf hin, dass die menschliche Stimme als sozialer Hinweisreiz fungiert und damit auf die Lerner motivation wirkt.

(2) Das Sprechergeschlecht gilt als Spezialfall des Stimmprinzips. Geschriebene Texte offenbaren das Geschlecht hingegen nicht, während bei gesprochenen Texten das Geschlecht des Sprechers meist erkennbar ist. Insofern sind Effekte des Sprechergeschlechts spezielle Hinweise darauf, dass bei Modalitätseffekten motivationale Prozesse eine Rolle spielen.

(3) Eine gesprochene Textdarbietung könnte eine Form der Personalisierung beinhalten, da sich Lernende angesprochen fühlen und eine Zuwendungsreaktion erfolgt, auch wenn Texte formal gehalten werden. Durch das direkte Ansprechen der Lernenden kann die Personalisierung erhöht werden.

Stimmen variieren auf vielen Dimensionen wie z.B. Sprechgeschwindigkeit, Tonlage, Klangfarbe, usw. All diese Dimensionen werden durch den Vergleich von geschriebenen und gesprochenen Texten variiert. Im methodischen Sinne handelt es sich um eine Konfundierung von Variablen. Hier zeigt sich eine Vielzahl von sozialen Hinweisreizen, welche gemeinsam ihre Wirkung auf den Lernenden und dessen Lernprozess entfalten.

## Fazit und Ausblick

### Theoretische Implikationen

Die empirischen Ergebnisse legen nahe, welche Prozesse am Entstehen von Modalitätseffekten beteiligt sind. Folgende Komponenten im Prozess der Informationsverarbeitung sind zu beachten.

(1) Lese- und Hörverstehen: Unterschiede beim Lese- und Hörverstehen wirken sich auf die Bildung mentaler Repräsentationen aus. Ungeklärt ist dabei, welche Bedeutung die Verarbeitung modalitätsspezifischer Merkmale von Texten für die Arbeitsgedächtnisbelastung besitzt. Besonders beim Lesen treten Blickbewegungen und deren Steuerung notwendigerweise auf. Bei bestimmten Lernmaterialien können auditive Texte durch einen längeren Verbleib der sensorischen Informationen im auditiv-sensorischen Register zu Lernvorteilen führen.

(2) Visuelle Split-Attention Effekte: Wenn Bilder und geschriebene Texte räumlich getrennt dargeboten werden, wird das Arbeitsgedächtnis, z.B. durch Blickbewegungen zwischen Text und Bild und deren Steuerung oder Suchprozesse, beeinträchtigt. Zusätzlich wird das Arbeitsgedächtnis durch einen zeitlichen Split-Attention Effekt belastet. Hierbei müssen Informationen im Gedächtnis länger als nötig aktiv gehalten werden.

(3) Motivationale Effekte: Eine tiefere Verarbeitung von Informationen aufgrund motivationaler Effekte durch soziale Hinweisreize in Lernmaterialien ist durch gesprochene Texte anzunehmen.

Für eine adäquate Theorie, welche Effekte der Textmodalität beschreiben, erklären und vorhersagen kann, besteht künftig die Herausforderung darin, Elemente zu identifizieren, welche zur Komposition von Modalitätseffekten beitragen. Dies ist vor allem in motivationaler Hinsicht unerforscht, während kognitive Prozesse noch zu vervollständigen und zu präzisieren sind.

### Empirische Implikationen

Zukünftig gibt es zwei Herausforderungen. Der Ersten liegt eine grundlagenorientierte, auf den Erkenntnisgewinn über das menschliche Lehren und Lernen fokussierte Perspektive zugrunde. Hier muss experimentell und methodisch eindeutig geklärt werden, wie sich Modalitätseffekte zusammensetzen. Der Zweiten liegt eine anwendungsorientierte, auf das Praxiswissen fokussierte Perspektive zugrunde, unter welchen Bedingungen Modalitätseffekte zu beachten sind. Dabei ist zu klären, unter welchen Rahmenbedingungen Modalitätseffekte eintreten oder ausbleiben und wie dabei Modalitätseffekte durch das Einwirken auf die einzelnen Effekte konstituierenden Komponenten moderiert werden.

Empirisch wurden zu den Rahmenbedingungen des Modalitätseffekts bisher wenige Ergebnisse veröffentlicht, wobei vermutet wird, dass hier auch ein „publication bias“ bezüglich von Nicht-Effekten mitwirkt. Speziell können die Arbeiten angegeben werden, in denen bisher insgesamt oder unter bestimmten Randbedingungen kein Modalitätseffekt (Barron & Kysilka, 1993; Paechter, 1996; Rinck & Glowalla, 1996; Schmidt-Weigand, 2006, Exp. in Kap. 4) oder sogar Vorteile für geschriebene Texte (Stiller, 2007; Stiller et al., 2009; Tabbers, 2002) gefunden wurden. Spezielle Hinweise auf Randbedingungen finden sich bezüglich der Komplexität der Lehrmaterialien (Ginns, 2005; Jeung et al., 1997; Tindall-Ford et al., 1997), der Komplexität der Wissensabfrage (Guan, 2003; Leahy et al., 2003), der Steuerung einer Lehrpräsentation (Stiller et al., 2009; Stiller et al., 2011) und der Lernermerkmale Arbeitsgedächtniskapazität, Gedächtnisstrategien, Vorwissen und Einstellung zum Computer (z.B. Seufert et al., 2009; Stiller, 2007, Exp. 2). Als weitere Anwärtler für Randbedingungen werden in der Literatur diskutiert: Lernzeit, Lesegeschwindigkeit, Textlänge und Textschwierigkeit, räumliches Vorstellungsvermögen, räumlicher Informationsgehalt von Texten und die Art des Wissen (Kürschner & Schnotz, 2008), Umfang der Redundanz zwischen Text- und Bilderinformationen sowie der kommunikative Erfahrungshintergrund (Kommunikationsschwerpunkt; Rickheit et al., 1987).

Darüber hinaus existieren auch Variablenkonfundierungen beim Wechsel der Präsentationsmodalität von Texten bzw. schwer zu beurteilende Variablenkonstanthaltungen auf der einen und Merkmalsvariabilität auf der anderen Seite. Als Beispiel für Konfundierungen sind Verarbeitungshilfen durch grafische Gestaltung von Lesetexten (Ballstaedt, 1997) und die Prosodie bei Hörtexten (Imhof, 2003), wie z.B. Variationen der Sprechgeschwindigkeit, Frequenz, Intonation und den Einsatz von Pausen, zu nennen. Inwieweit dabei gesprochene und geschriebene Texte unabhängig von der Wortidentität gleich sind, ist fraglich. Schwer zu beurteilende Konstanthaltungen betreffen hauptsächlich die Sprechgeschwindigkeit der auditiven Texte im Vergleich zu einer prinzipiell möglichen Variation der Lesegeschwindigkeit durch die Lernenden selbst. In den Experimen-

ten wird die Lehrbedingung mit gesprochenen Texten als Ausgangspunkt für die Gestaltung der zu vergleichenden Lehrpräsentationen mit Lesetexten verwendet.

## Literaturverzeichnis

- Atkinson, R. K., Mayer, R. E. & Merrill, M. M. (2004). Fostering social agency in multimedia learning: Examining the impact of an animated agent's voice. *Contemporary Educational Psychology*, 30, 117-139.
- Baddeley, A. D. (1986). *Working memory*. Oxford: Clarendon Press.
- Baddeley, A. D. (1997). *Human memory. Theory and practice*. Hove, UK: Psychology Press.
- Ballstaedt, S.-P. (1997). *Wissensvermittlung. Die Gestaltung von Lernmaterial*. Weinheim: PVU.
- Barron, A. E. & Kysilka, M. L. (1993). The effectiveness of digital audio in computer-based training. *Journal of Research on Computing in Education*, 25, 277-289.
- Brünken, R., Plass, J. L. & Leutner, D. (2004). Assessment of cognitive load in multimedia learning with dual task methodology: Auditory load and modality effects. *Instructional Science*, 32, 115-132.
- Brünken, R., Seufert, T. & Zander, S. (2005). Förderung der Kohärenzbildung beim Lernen mit multiplen Repräsentationen. *Zeitschrift für Pädagogische Psychologie*, 19, 61-75.
- Brünken, R., Steinbacher, S., Plass, J. L. & Leutner, D. (2002). Assessment of cognitive load in multimedia learning using dual task methodology. *Experimental Psychology*, 49, 109-119.
- Craig, S. D., Gholson, B. & Driscoll, D. M. (2002). Animated pedagogical agents in multimedia educational environments: Effects of agent properties, picture features, and redundancy. *Journal of Educational Psychology*, 94, 428-434.
- De Westelinck, K., Valcke, M., De Craene, B. & Kirschner, P. (2005). Multimedia learning in social sciences: Limitations of external graphical representations. *Computers in Human Behavior*, 21, 555-573.
- Engelkamp, J. (1990). *Das menschliche Gedächtnis. Das Erinnern von Sprache, Bildern und Handlungen*. Göttingen: Hogrefe.
- Ginns, P. (2005). Meta-analysis of the modality effect. *Learning and Instruction*, 15, 313-331.
- Ginns, P. & Fraser, J. (2010). Personalization enhances learning anatomy terms. *Medical Teacher*, 32, 776-778.
- Guan, Y.-H. (2003). *The effects of multimedia presentations on information processing. Eye-movement analyses of text and picture integration in a multimedia-based learning scenario*. Retrieved November 12, 2009, from <http://deposit.ddb.de/cgi-bin/dokserv?idn=967067294>
- Hartley, J., Brown, C. & Michael, D. (1989). The effect of sex of speaker and listener on recall from a medical audiotape. *British Journal of Educational Technology*, 20, 191-199.
- Hasebrook, J. (1994). *Vermittlung und Erwerb von Strukturwissen: Studierhilfen für gedruckte und elektronische Lehrtexte*. Marburg: Universität Marburg.
- Imhof, M. (2003). *Zuhören – Psychologische Aspekte auditiver Informationsverarbeitung*. Göttingen: Vandenhoeck & Ruprecht.
- Jeung, H. J., Chandler, P. & Sweller, J. (1997). The role of visual indicators in dual sensory mode instruction. *Educational Psychology*, 17, 329-343.

- Kalyuga, S., Chandler, P. & Sweller, J. (1999). Managing split-attention and redundancy in multimedia instruction. *Applied Cognitive Psychology*, *13*, 351-372.
- Kalyuga, S., Chandler, P. & Sweller, J. (2000). Incorporating learner experience into the design of multimedia instruction. *Journal of Educational Psychology*, *92*, 126-136.
- Karayianni, I. & Gardiner, J. M. (2003). Transferring voice effects in recognition memory from remembering to knowing. *Memory and Cognition*, *31*, 1052-1059.
- Kürschner, C. & Schnotz, W. (2008). Das Verhältnis gesprochener und geschriebener Sprache bei der Konstruktion mentaler Repräsentationen. *Psychologische Rundschau*, *59*, 139-149.
- Leahy, W., Chandler, P., & Sweller, J. (2003). When auditory presentations should and should not be a component of multimedia instruction. *Applied Cognitive Psychology*, *17*, 401-418.
- Linek, S. B. (2007). *Geschlechtsspezifisches Design von narrativen Animationen: "Speaker/Gender Effect" und die Schema-Inkongruenz von Information* (Band 6 der Reihe Wissensprozesse und digitale Medien). Berlin: Logos.
- Low, R. & Sweller, J. (2005). The modality principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 147-158). New York, NY: Cambridge University Press.
- Mayer, R. E. (2005a). Cognitive theory of multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 31-48). New York, NY: Cambridge University Press.
- Mayer, R. E. (2005b). Principles for managing essential processing in multimedia learning: segmenting, pre-training, and modality principles. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 169-182). Cambridge: Cambridge University Press.
- Mayer, R. E. (2005c). Principles of multimedia learning based on social cues: Personalization, voice, and image principles. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 201-212). New York, NY: Cambridge University Press.
- Mayer, R. E. (2001). *Multimedia learning*. New York, NY: Cambridge University Press.
- Mayer, R. E., Dow, G. & Mayer, S. (2003). Multimedia learning in an interactive self-explaining environment: What works in the design of agent-based microworlds? *Journal of Educational Psychology*, *95*, 806-813.
- Mayer, R. E. & Moreno, R. (1998). A split-attention effect in multimedia learning: Evidence for dual processing systems in working memory. *Journal of Educational Psychology*, *90*, 312-320.
- Mayer, R. E., Sobko, K. & Mautone, P. D. (2003). Social cues in multimedia learning: Role of speaker's voice. *Journal of Educational Psychology*, *95*, 419-425.
- Mayer, R. E. Fennell, S., Farmer, L. & Campbell, J. (2004). A personalization effect in multimedia learning: Students learn better when words are in conversational style rather than formal style. *Journal of Educational Psychology*, *96*, 389-395.
- Moreno, R. & Mayer, R. E. (1999). Cognitive principles of multimedia learning: The role of modality and contiguity. *Journal of Educational Psychology*, *91*, 358-368.
- Moreno, R. & Mayer, R. E. (2000). Engaging students in active learning: The case for personalized multimedia messages. *Journal of Educational Psychology*, *92*, 724-733.
- Moreno, R. & Mayer, R. E. (2002). Learning science in virtual reality multimedia environments: Role of methods and media. *Journal of Educational Psychology*, *94*, 598-610.

- Moreno, R. & Mayer, R. E. (2004). Personalized messages that promote science learning in virtual environments. *Journal of Educational Psychology*, 96, 165-173.
- Moreno, R., Mayer, R. E., Spires, H. A. & Lester, J. C. (2001). The case for social agency in computer-based multimedia learning: Do students learn more deeply when they interact with animated pedagogical agents? *Cognition and Instruction*, 19, 177-214.
- Mousavi, S. Y., Low, R. & Sweller, J. (1995). Reducing cognitive load by mixing auditory and visual presentation modes. *Journal of Educational Psychology*, 87, 319-334.
- Nass, C., Robles, E., Heenan, C., Bienstock, H. & Treinen, M. (2003). Speech-based disclosure systems: Effects of modality, gender of prompt, and gender of user. *International Journal of Speech Technology*, 6, 113-121.
- O'Neil, H. F., Mayer, R. E., Herl, H. E., Niemi, C., Olin, K. & Thurman, R. A. (2000). Instructional strategies for virtual aviation training environments. In H. F. O'Neil & D. H. Andrew (Eds.), *Aircrew training and assessment* (pp. 105-130). Mahwah, NJ: Erlbaum.
- Paechter, M. (1996). *Auditive und visuelle Texte in Lernsoftware. Herleitung und empirische Prüfung eines didaktischen Konzepts zum Einsatz auditiver und visueller Texte in Lernsoftware*. Münster: Waxmann.
- Penney, C. G. (1989). Modality effects in delayed free recall and recognition: visual is better than auditory. *The Quarterly Journal of Experimental Psychology*, 41, 455-470.
- Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, 124, 372-422.
- Rayner, K., & Pollatsek, A. (1987). Eye movements in reading: A tutorial review. In M. Coltheart (Ed.), *Attention and performance XII: The psychology of reading* (pp. 327-362). London: Erlbaum.
- Rayner, K. & Pollatsek, A. (1989). *The psychology of reading*. Englewood Cliffs, NJ: Prentice Hall.
- Rickheit, G., Strohner, H. & Müssler, J. (1987). Modalitätsspezifische Textverarbeitung bei Personen mit unterschiedlichem Kommunikationsschwerpunkt. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 19, 65-77.
- Rinck, M. & Glowalla, U. (1996). Die multimediale Darstellung quantitativer Daten. *Zeitschrift für Psychologie*, 204, 383-399.
- Rummer, R., Schweppe, J., Scheiter, K. & Gerjets, P. (2008). Lernen mit Multimedia. Die kognitiven Grundlagen des Modalitätseffekts. *Psychologische Rundschau*, 59, 98-107.
- Schmidt-Weigand, F. (2006). *Dynamic visualizations in multimedia learning: The influence of verbal explanations on visual attention, cognitive load and learning outcome*. Retrieved December 06, 2010, from <http://geb.uni-giessen.de/geb/volltexte/2006/2699/>
- Seufert, T., Schütze, M. & Brünken, R. (2009). Memory characteristics and modality in multimedia learning: An aptitude-treatment-interaction study. *Learning and Instruction*, 19, 28-42.
- Stern, E. S., Mullenix, J. W. & Yaroslavsky, I. (2006). Persuasion and social perception of human vs. synthetic voice across person as source and computer as source conditions. *International Journal of Human-Computer Studies*, 64, 43-52.
- Stiller, K. D. (2007). *Computerised multimedia learning. Modes of text presentation and access to text*. Hamburg: Verlag Dr. Kovač.

Stiller, K. D., Freitag, A., Zinnbauer, P. & Freitag, C. (2009). How pacing of multimedia instructions can influence modality effects: A case of superiority of visual texts. *Australasian Journal of Educational Technology*, 25, 184-203. Retrieved May 13, 2010, from <http://www.ascilite.org.au/ajet/ajet25/stiller.html>

Stiller, K. D. & Jedlicka, R. (2010). A kind of expertise reversal effect: Personalisation effect can depend on domain-specific prior knowledge. *Australasian Journal of Educational Technology*, 26, 133-149. Retrieved May 13, 2010, from <http://www.ascilite.org.au/ajet/ajet26/stiller.html>

Stiller, K. D., Petzold, K. & Zinnbauer, P. (2011). Presentation time concerning system-paced multimedia instructions and the superiority of learner pacing. *Australasian Journal of Educational Technology*, 27, 693-708. Retrieved September 22, 2011, from <http://www.ascilite.org.au/ajet/ajet27/stiller.html>

Sweller, J. (1999). *Instructional design in technical areas*. Camberwell, Australia: ACER Press.

Tabbers, H. K. (2002). *The modality of text in multimedia instructions. Refining the design guidelines*. Heerlen: Educational Technology Expertise Centre, Open University of the Netherlands.

Tindall-Ford, S., Chandler, P., & Sweller, J. (1997). When to sensory modes are better than one. *Journal of Experimental Psychology: Applied*, 3, 257-287.

Van Gerven, P. W. M. (2002). *Efficient complex skills training into old age: Exploring the benefits of cognitive load theory*. Retrieved May 13, 2009, from [http://www.personeel.unimaas.nl/p.vangerven/Index\\_files/Pascal%20van%20Gerven%20-%20Dissertation.pdf](http://www.personeel.unimaas.nl/p.vangerven/Index_files/Pascal%20van%20Gerven%20-%20Dissertation.pdf)