The Modality Effect in Multimedia Learning: Theoretical and Empirical Limitations

Anne Schueler, Katharina Scheiter, University of Tuebingen, Konrad-Adenauer-Str. 40,
72072 Tuebingen, Germany,
Email: a.schueler@iwm-kmrc.de, k.scheiter@iwm-kmrc.de,
Peter Gerjets, Knowledge Media Research Center, Konrad-Adenauer-Str. 40,
72072 Tuebingen, Germany, p.gerjets@iwm-kmrc.de
Ralf Rummer, Saarland University, Postfach 151150,
66041 Saarbruecken, Germany, r.rummer@mx.uni-saarland.de

Abstract: Numerous studies in the field of multimedia learning have confirmed the modality principle, which claims that the presentation of pictures with auditory texts leads to better learning outcomes than the presentation of pictures with visual text. However, there are still some doubts concerning the theoretical explanation and the boundary conditions of the modality principle. An overview is given with regard to these two points. Furthermore, two experiments are reported, which aimed at clarifying the theoretical explanation (Experiment 1) and the boundary conditions of the modality effect (Experiment 2). Experiment 1 was conducted to test whether the modality effect occurs irrespective of the temporal contiguity of text and pictures. In Experiment 2, it was investigated, if the modality principle also holds for longer, complex texts and if the time available for information processing acts as a moderating variable.

Introduction

In the research on multimedia learning, there have been several recommendations on how to present text and pictures to learners to support a deeper understanding of the presented topic (Mayer, 2001; Sweller, van Merriënboer, & Paas, 1999). According to one of these recommendations, the so called modality principle, text should be presented spoken (auditory) instead of written (visual) to learners (Mayer, 2001). The superiority of pictures with auditory text over pictures with visual text has been confirmed in numerous empirical studies, which used different measurements and different topics. An overview of the various studies is given by Ginns (2005) in his meta-analysis. He concludes that "across a broad range of instructional materials, age groups and outcomes, students who learned from instructional materials using graphics with spoken text outperformed those who learned from a graphics with printed text" (Ginns, 2005, p. 326). The overall weighted mean effect size of effects was moderate to large, which also confirms the significance of the modality principle for the design of multimedia material. Thus, it seems as if the modality principle is one of the best confirmed principles in multimedia learning. However, there are at least two aspects, which lead to criticism. The first concerns the theoretical explanation of the principle; the second concerns its boundary conditions, because – as will be shown later – under certain conditions the modality effect could not be found. Therefore, in the current paper we want to consider these two points of criticism.

Theoretical Explanations of the Modality Effect

The modality effect is often explained against the background of the Cognitive Theory of Multimedia Learning (Mayer, 2001). According to this theory, the cognitive system consists of two channels, which are both limited in capacity. In the initial stage of processing in working memory, one of these channels is assumed to handle all auditory material, that is, auditory text, whereas the other channel is assumed to deal with all visual material, that is, visual text and pictures. Because of the limited capacity of the channels, the use of only one channel when processing visual text and pictures can lead to an overload of this channel. On the other hand, using two channels when processing auditory text and pictures is assumed to augment the working memory capacity available and therefore, yields performance. In the following, this explanation for the modality effect will be called the modality-specific explanation. In the current paper, we will not discuss the conceptual problems associated with the use of Baddeley's model of working memory (Baddeley, 1992) as a theoretical foundation for the aforementioned explanation. These problems have been elaborated elsewhere (see Rummer, Schweppe, Scheiter, & Gerjets, in press).

An alternative explanation for the modality effect might be the fact that auditory text and picture can be perceived simultaneously, whereas in the case of visual text and picture learners can only process one source of information at a given point in time. Thus, learners have to look at the picture and then to read the text or vice versa. Sweller et al. (1999) call such a situation a *split attention* situation, because the learners have to split their attention at least between two sources of information. Mayer (2001) distinguishes two types of split attention: a

spatial split attention and a temporal split attention. The corresponding principles are the *spatial contiguity principle* and the *temporal contiguity principle*. The spatial contiguity principle refers to visual text and states that students learn better, when corresponding visual text and pictures are presented spatially integrated instead of spatially separated. This is the case, because when presenting text and picture close to each other the learners do not to have to search for the corresponding pictorial and verbal information, which leads to more cognitive capacity being available for other cognitive processes. The temporal contiguity principle refers to auditory text only and states that students learn better, when corresponding auditory text and pictures are presented near in time (simultaneously) than far away in time (successively). This is the case, because according to Mayer (2001) the probability of making mental connections between the verbal and the pictorial information is enhanced, if both are represented in the cognitive system at the same time. With a successive presentation, words and pictures are not necessarily hold active in parallel in working memory; thus, the learners have to reconstruct the respective information. These reconstruction processes can impose cognitive load. Furthermore, important aspects of one source of information may be forgotten when perceiving the other source of information – thereby, resulting in lower learning outcomes.

In the case of presenting visual text and picture, full spatial contiguity and full temporal contiguity are never given. Spatial contiguity is absent, because the learners have to switch between visual text and picture, even if the text is integrated in the picture. For the same reason temporal contiguity is absent, because the learners can never perceive both sources of information simultaneously. Thus, it may be this lack of spatial and temporal contiguity that causes the modality effect. Accordingly, the modality effect might be explained solely by the fact that auditory text establishes contiguity rather than by assuming that it makes available cognitive processing resources. However, Mayer rejected this alternative explanation for the modality effect due to the results of an experiment, which will be described below.

Empirical Evidence: Modality or Lack of Contiguity?

A first study that aimed at differentiating between modality and lack of contiguity as possible explanations for the modality effect is reported by Moreno and Mayer (1999). In Experiment 1, they investigated whether the modality effect can be traced back to the fact that auditory text provides access to an additional cognitive resource (i.e., modality-specific explanation) or that learners need to split visual attention in the case of written test (i.e., spatial contiguity explanation). If the first explanation was true, a general superiority of auditory text over visual text, regardless of the proximity of the visual text to the picture, should be observable. If the second explanation was true, the modality effect should diminish if the visual text is integrated rather than spatially separated from the picture, because with integrated visual text less visual search is necessary. Moreno and Mayer found a general superiority of auditory text, thereby apparently confirming the modality-specific explanation, but, unfortunately, do not report the comparison between auditory text and integrated visual text. Thus, it is still possible that there is a smaller modality effect or even no significant difference between these two conditions, which would indicate that the modality effect does not occur irrespective of spatial contiguity (see Rummer et al., in press). In Experiment 2, Moreno and Mayer (1999) investigated whether the modality effect can be traced back to the modality-specific explanation or to the temporal contiguity explanation, that is, to the fact that only with auditory text a simultaneous processing of verbal and pictorial information is possible. If the modality-specific explanation was true, a general superiority of auditory text is expected, irrespective of the temporal contiguity of auditory text and picture. If the temporal contiguity explanation was true, a superiority of auditory text is expected only if temporal contiguity is given with the presentation of auditory text and picture, but not if no temporal contiguity is given. The authors presented text (auditory or visual) sequentially before or subsequent to a corresponding picture, or they presented text simultaneously with the picture. In case of a simultaneous presentation, only auditory text should yield a temporal contiguous presentation. In case of a sequential presentation, there should be no temporal contiguity, neither in the condition with auditory nor in the condition with visual text. Thus, "if the advantage of narration over on-screen text resides in a modality principle, then the advantage for auditory-visual presentations should not disappear when they are made sequential" (Moreno & Mayer, 1999, p. 360). In other words: A modality effect for the sequential groups would indicate that there is something beyond temporal contiguity, namely, enhanced resources made available by using auditory text. The results seem to support the modality-specific explanation: Learners with auditory text and picture showed superior performance than learners with visual text and picture, both with simultaneous and sequential presentation. However, a closer look at the results shows that there is one problem with the data: The authors found no temporal contiguity effect for the two auditory groups, which means that learners with sequential presentation were as good as learners with simultaneous presentation. One explanation might be the chosen material: The authors presented only a short text segment to the learners, then presented the corresponding part of the animation, then presented the next text segment and so on. With this kind of presentation usually no temporal contiguity effect is found (see for example Mayer, Moreno, Boire & Vagge, 1999). Mayer (2001) explains the absence of a temporal contiguity effect with short auditory text segments as follows: The mental integration of text and picture is not affected,

because related information is presented with a very short time-lag. This means that the information is still available in working memory, when the related information enters. Therefore, in the case of short text segments, temporal contiguity might be given for learners with auditory text, even if text and picture are presented sequentially. Thus, the chosen material was inappropriate to distinguish between the temporal contiguity explanation and the modality-specific explanation, because in both auditory groups temporal contiguity was established, which leads in turn to better performance of the auditory groups. However, with longer text segments also the auditory group with sequential presentation should suffer from a lack of temporal contiguity, because like learners with visual text they will then have to reconstruct the respective information. If both the auditory as well as the visual condition suffer from a lack of temporal contiguity in similar ways, according to the temporal contiguity explanation no difference is expected between the two sequential groups with auditory and visual text. There are two studies, which seem to confirm the assumption that with longer texts and sequential presentation the modality effect disappears: Tiene (2000) compared simultaneous with sequential presentation and varied the modality of text presentation. He found a modality effect with simultaneous presentation, but not with sequential presentation. Unfortunately, the data for temporal contiguity between the two auditory groups are not reported. Baggett and Ehrenfeucht (1983) showed that in the case of sequential presentation no modality effect occurred, that is, learners who read the text after or before they saw a movie, scored even better in a post test than learners who heard the text. Additionally, they found a temporal contiguity effect for auditory text, which confirms the assumption that temporal contiguity is the crucial aspect.

Experiment 1

The purpose of this experiment was to confirm the assumption that the modality effect is a result of absent temporal contiguity with visual text and not the result of enhanced capacity with auditory text. By varying on the one hand the contiguity of presentation and on the other hand the text modality, it is possible to distinguish between the temporal contiguity and modality-specific explanation: According to the modality-specific explanation, the modality effect should occur irrespective of whether the presentation is simultaneous or sequential. According to the temporal contiguity explanation a modality effect is expected only with simultaneous presentation, but not with sequential presentation. To test these assumptions, an experiment was conducted, where students learned about the development of tornados.

Method

Participants and Design

Sixty-eight participants with different educational backgrounds participated in the study. Four of them had to be excluded from the analysis because of extreme values in prior knowledge. The participants could receive credit points for attendance. The remaining sixty-four participants (M = 24.06 years, SD = 6.43 years; 44 female vs. 20 male) were assigned randomly to one of four conditions, which resulted from a 2 x 2 Design, with text modality (auditory vs. visual text) and contiguity of presentation (simultaneous vs. sequential) as independent variables.

Materials

The material consisted of paper-pencil material and a computerized learning environment. The paper-pencil material comprised a personal questionnaire, a questionnaire measuring prior knowledge, and a questionnaire which measured learning outcomes by recall (verbal and pictorial) and transfer items.

The personal questionnaire required the participants to state data about their age, gender, profession, their last grade in geography and physics and their interest in weather phenomena. The prior knowledge test consisted of six multiple choice items, where the correct answer had to be chosen from four alternatives. For every correct answer they got one point. The recall and transfer tests contained multiple choice items, open questions, and pictorial items. Verbal recall was measured by 13 multiple choice questions, where the learners had to choose the correct answer from four possible answers, and three open questions (e.g., “What is the difference between a funnel cloud and a tornado?”). For each correct answer the learners were assigned one point. Pictorial recall was measured by four pictorial items: The learners were given a picture, in which they had to draw specific aspects concerning the development of a tornado. Two independent raters evaluated the drawings with regard to several aspects; three points were the maximal score per picture. Disagreements were resolved by consensus. Transfer was measured by two multiple choice items and two open questions. For each correct answer the learners received one point. The time for responding to the tests was unlimited.

The computerized learning environment consisted of eight static pictures and related German texts explaining the development of a tornado. Each text passage consisted of approximately three sentences per picture. In the simultaneous conditions with auditory text, the learners heard the text when seeing the picture. In the simultaneous condition with visual text, the text was presented below the picture (see Figure 1, part B). In the sequential condition with auditory text, the learners first heard the auditory text while the screen remained blank. Then the picture was presented. In the sequential condition with visual text, first the text was presented

on a screen, than the picture was presented on a new screen (see Figure 1, part C). The presentation was system-paced and the available time was determined by the length of the narration. In the simultaneous groups the presentation lasted three minutes whereas in the sequential groups it lasted six minutes.

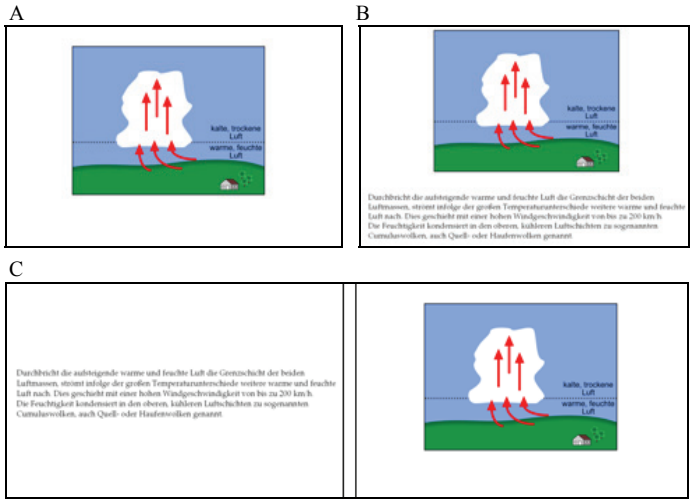

Figure 1. Frames selected from the presentation. A: Simultaneous presentation with auditory text. B: Simultanous presentation with visual text. C: Sequential presentation with visual text.

Procedure

Participants were tested individually. First, participants were given a short written instruction about the experiment. Second, they completed the personal questionnaire. Third, they completed the multiple choice test which assessed prior knowledge. Fourth, they learned about tornados in one of the four experimental conditions. Sixth, they answered the questionnaire, which measured recall and transfer performance. The experiment lasted about thirty minutes.

Results

Two-factor analysis of variances were conducted for the independent variables with modality and temporal contiguity as the between subject factors. Table 1 shows the means and standard deviations for each group. Because the control variables did not affect learning outcome, no analyses of covariance had to be conducted.

Table 1: Means and standard deviations (in parentheses) for the percentage performance in verbal recall, pictorial recall, and transfer as a function of contiguity of presentation and text modality.

	Simultaneous presentation		Sequential presentation	
	Auditory text	Visual Text	Auditory Text	Visual Text
Verbal Recall	71.88 (15.31)	72.27 (10.94)	70.83 (16.14)	77.73 (17.53)
Pictorial Recall	70.10 (20.16)	54.17 (21.94)	61.39 (17.07)	71.35 (11.06)
Transfer	45.59 (29.63)	45.31 (26.17)	45.59 (29.63)	48.44 (17.00)

For recall of verbal information, neither the difference between the visual groups and the auditory groups ($F < 1$) nor the difference between the simultaneous groups and the sequential groups was significant ($F < 1$). Also the interaction between the two factors was not significant ($F < 1$). For recall of pictorial information no differences were found, neither between the visual and the auditory groups ($F < 1$) nor between the simultaneous and the sequential groups ($F < 1$). However the interaction between the two factors was significant ($F(1,60) = 8.18$, $p = .01$, partial $\eta^2 = 0.12$). A Bonferroni posthoc analysis yielded a significant modality effect in the simultaneous condition ($F(1,60) = 6.39$, $p = .01$, partial $\eta^2 = 0.10$), whereas in the sequential condition no modality effect appeared ($F(1,60) = 2.35$, $p = .13$, partial $\eta^2 = 0.04$). Interestingly, within the visual group learners with sequential presentation outperformed those with simultaneous presentation ($F(1,60) = 7.22$, $p = .01$, partial $\eta^2 = 0.11$), that is, an inverted temporal contiguity effect was found, whereas between the auditory groups no difference was found ($F(1,60) = 1.85$, $p = .18$, partial $\eta^2 = 0.03$) For transfer, the ANOVAs showed no difference between the visual and the auditory groups ($F < 1$) or between the simultaneous and the sequential groups ($F < 1$), and also the interaction between the two factors was not significant ($F < 1$).

Discussion

To summarize, there was no main effect of modality. This contradicts the modality-specific explanation, according to which a general modality effect should have occurred for a simultaneous as well as for a sequential presentation. Thus, the results indicate that there is no general advantage of using an auditory modality. Results from major interest are the interactions. According to the temporal contiguity explanation, a modality effect in the simultaneous condition and no modality effect in the sequential condition were expected. This prediction could be confirmed for pictorial recall, that is, with simultaneous presentation learners with auditory text scored higher than learners with visual text, whereas with sequential presentation there was no difference between the two groups. At first view, it seems as if the temporal contiguity explanation could be confirmed. However, there was no effect of temporal contiguity within the auditory groups, despite the fact that the text segments had been rather long. This indicates that temporal contiguity did not influence the performance within the auditory simultaneous group. Accordingly, the superiority of the auditory simultaneous group over the visual simultaneous group cannot be explained with temporal contiguity. Therefore, it seems more reasonable to refer to the spatial contiguity principle for explaining the result: Whereas with simultaneous presentation the learners with auditory text did not need to split their attention, learners with visual text had to read the text and to look at the picture and to relate them. As this requires visual search processes, which the group with auditory text did not have to undertake, the modality effect occurred. Within the sequential groups no spatial split was given, which resulted in the same performance for learners with visual and auditory text. Due to the fact that with verbal recall and transfer no interaction could be found, it seems as if within the simultaneous groups, learners with visual text allocated more attention to the text than to the picture. This leads to the following explanation of the found results: Although processing mainly the text, learners with visual text and simultaneous presentation tried to refer to the picture. This led to worse processing of the picture, because less time was allocated to it, and furthermore, it led to enhanced cognitive load because of visual search processes. This assumption might explain, why with regard to pictorial items a modality effect appeared, but not with regard to verbal items. It could also explain the found inverted temporal contiguity effect for visual text with regard to pictorial recall, because with sequential presentation enough time could be allocated to the picture, resulting in better performance. That is, despite the potential drawbacks associated with a sequential processing of text and pictures, learners with written text actually benefited from a sequential presentation and the additional processing time, because it allowed for a more thorough processing of the picture and potentially the text as well. On the other hand, learners with auditory text did not show any better learning outcomes, when more time was available. Thus, one boundary condition for the modality effect might be the available time for processing the material. Why this might be the case will be discussed in the next section.

Time as a Boundary Condition of the Modality Effect

Besides the theoretical problems of the explanation of the modality effect, it remains the question under what conditions the modality effect appears. Although the modality principle has been replicated in a number of laboratory studies there are some doubts, if the recommendation to use auditory text with picture is as universally valid as has been suggested. In fact, there are studies which showed that the modality effect did not arise under certain conditions. For example, with increasing expertise of the learner, with very complex pictorial material or with self pacing (for an overview see Ginns, 2005) the auditory text was shown to be no longer superior to visual text.

Further doubts concerning the validity of modality principle come from text comprehension research, where visual and auditory texts *without* pictures are compared. From this research area, it is even known that visual text may lead to better recall of the presented information than auditory text (e.g., Furnham, 2001). This superiority of visual over auditory text is explained by the permanence of visual text. Whereas auditory text is transient and the information is only available at one point in time, with visual text several processing strategies can be implemented. First, the reader can slow down the reading rate (Byrne & Curtis, 2000), second, s/he can read through sections of the text several times (e.g., Furnham, de Siena, & Gunter, 2002) and third, s/he can skip text passages which are easy or not relevant to the topic (Bazerman, 1985). Thus, with visual text there are several processing strategies which are not given with auditory text. These strategies can explain the superiority of visual text as it was found in text comprehension research. Especially with more complex texts this strategies should be helpful, because the learners can adapt the processing of the text to their own cognitive prerequisites. However, under some conditions the superiority of visual text over auditory text was not found. First, when presenting very easy texts there was no difference between auditory and visual text (Müsseler, Rickheit, & Strohner, 1985). Second, when presenting the visual text in single sentences instead of longer text passages there was also no difference between auditory and visual text (Kintsch, Kozminsky, Streby, McKoon, & Keenan, 1975). Third, with high speed of presentation even a superiority of auditory text was found (Sannomiya, 1981). It is important to note that obviously all of these moderating variables have an influence on whether text processing strategies can and need to be applied.

Regarding the studies conducted in multimedia research it appears that often short and not very complex texts have been used and that the time for presentation was constrained by the duration of the auditory

condition. For example, Moreno and Mayer (1999) presented one or two visual sentences together with the corresponding parts of the animation and the available time was determined by the length of the auditory condition. Only a few studies presented longer text materials, for example, the studies of Baggett and Ehrenfeucht (1983) and Tiene (2000). As just mentioned, in these studies no modality effect was found under sequential presentation (Tiene, 2000) or the modality effect was even inverted (Baggett & Ehrenfeucht, 1983). Beside the fact that temporal contiguity was missing, it might be that time plays a crucial role: Because the learners had only to perceive the text and not to process the picture simultaneously, with visual text they had more time to apply text processing strategies which in turn leads to better performance when reading the text than when listening to the text. Similar results were found by Tabbers (2002): When the learners had control over the pace of instruction, learners with visual text and picture showed higher transfer performance than learners with auditory text and picture. Thus there is some evidence that with more time available and longer texts the modality effect cannot be confirmed.

Because of the empirical findings concerning text comprehension research and studies, in which longer texts were used, it is assumed that the modality principle is not as valid as numerous studies seem to suggest. Instead, no modality effect is expected, when presenting the learners with longer, complex texts, if they have the possibility to control the pace of presentation. On the other hand, a modality effect is expected, if the presentation is system-paced, that is, restricted in time. This is the case, because there is a smaller change that text processing strategies can be applied to visual text. Furthermore, spatial split attention may be more harmful, since pictorial and verbal information cannot be reviewed.

Experiment 2

The purpose of this experiment was to examine, how with longer, complex texts, text modality and pacing affect learning outcome. The material consisted of animations and texts that conveyed information about the reproduction of cells, namely mitosis.

Participants and Design

Eighty-one students from the University of Tuebingen, Germany, participated in the study. They received ten euros as compensation. Seven participants had to be excluded, because they studied a biological major. The remaining seventy-four participants (M = 24.09 years, SD = 4.33 years; 42 female, 29 male (with regard to gender, data for three persons are missing)) were assigned randomly to one of four conditions within a 2 x 2 Design, with text modality (auditory vs. visual text) and pacing (system-paced vs. learner-paced) as independent variables.

Material

The material consisted of paper-pencil material and a computerized learning environment, in which the independent variables were varied. The paper-pencil material consisted of a demographics questionnaire, a prior knowledge test and a recall and transfer test.

The questionnaire required the participants to state data about their major, the age, the gender, the number of semesters, the last school grade in German and the last school grade in mathematics. It also contained two questions which measured experience with biology and one question which measured time spend at the computer per week. Verbal recall was measured by one open question and 12 multiple choice questions. The open question asked the learners to write down all they could remember about the process of mitosis. The time was limited to ten minutes. Two independent scorers determined the recall score. Disagreements were resolved by consensus. To answer the multiple choice recall questions the learners had to choose the right answer from four alternatives. There was no time limitation. For each correct answer learners got one point. Transfer was measured by nine open questions. Each item was typed on a separate sheet. The transfer items asked the learners for example: “Suppose that when the nuclear envelope fragments, the pieces of the nuclear envelope are ejected from the cell. Would that affect the process of mitosis?”. As with recall, two independent scorers determined the transfer score, disagreements were resolved by consensus. The time for responding to the transfer items was three minutes per item.

The computerized learning environment consisted of six animations and related German texts, which explained the process of mitosis. Each text passage consisted of approximately six sentences per picture. In the conditions with auditory text the learners heard the text while the animation was presented on the screen. In the conditions with visual text the text was presented on the computer screen below the animation. Self pacing was operationalized by providing a “back” and a “forward” button for each of the six animations. When clicking the “back” button, the animation and the corresponding texts were displayed again. When clicking the “forward” button the next animation together with text was presented. In the system-paced condition the next animation started automatically after some seconds, thus the learners had no opportunity to interact with the environment. Examples of the material are given in Figure 2.

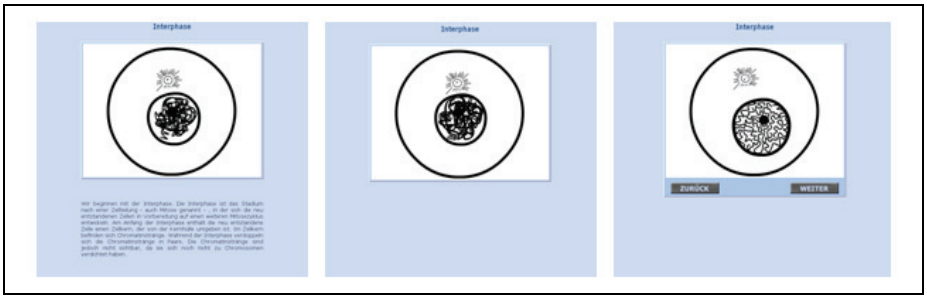

Figure 2. Frames selected from the learning environment. Left: animation with visual text. Middle: animation with auditory text. Right: navigation buttons in the learner-paced condition.

Procedure

Participants were tested individually. First, participants were given a short written instruction about the experiment. Second, they completed the personal questionnaire. Third, the learning environment was presented. Fourth, participants had to answer the recall items and seven transfer items. Fifth, they answered the multiple choice questions. Sixth, they were given the final two transfer questions. The experiment lasted approximately one hour.

Results

Two-factor analysis of variances were conducted, with modality and pacing as the between subject factors. Table 2 shows the means and standard deviations for each group. Because the control variables did not affect learning outcome, no analyses of covariance had to be conducted.

Table 2: Means and standard deviations (in parentheses) for the percentage performance in recall and transfer as a function of text modality and pacing.

	System-paced		Learner-paced	
	Auditory text	Visual Text	Auditory Text	Visual Text
Recall (multiple choice)	53.51 (17.42)	47.06 (19.53)	43.23 (12.63)	47.69 (22.65)
Recall (open question)	28.03 (13.26)	21.93 (12.41)	29.39 (14.08)	25.73 (13.56)
Transfer	41.88 (19.23)	34.72 (15.52)	38.83 (17.69)	36.69 (17.99)

For the multiple choice recall items, the ANOVA showed neither a significant difference between the visual and the auditory groups ($F < 1$), nor a significant difference between the system-paced and the learner-paced groups ($F(1,66) = 1.18$, $p = .28$, partial $\eta^2 = 0.02$). Furthermore, the interaction between the two factors was not significant ($F(1,66) = 1.51$, $p = .22$, partial $\eta^2 = 0.02$). For recall measured by the open item no significant main effects appeared, neither between the visual and the auditory groups ($F(1,70) = 2.47$, $p = .12$, partial $\eta^2 = 0.03$) nor between the system-paced and the learner-paced groups ($F < 1$). The interaction between the two factors was also not significant ($F < 1$) Also for transfer the ANOVA showed no significant difference between the visual and the auditory groups ($F(1,70) = 1.27$, $p = .26$, partial $\eta^2 = 0.02$) or between the system-paced and the learner-paced groups ($F < 1$), nor the interaction was significant ($F < 1$).

Discussion

The results of the study confirm the assumption that the modality principle is not as valid as it seems to be, because neither a main effect for modality nor an interaction between modality and pacing could be observed. This indicates that also with system-paced presentation no modality effect occurred. These results are somewhat disappointing, as we expected that with a system-paced presentation learners with visual text and picture had no opportunity to apply sophisticated reading strategies and also suffer from split attention, which should lead to worse performance in comparison with auditory text. The absent interaction between text modality and pacing and the absent main effect for pacing indicate that neither for learners with auditory text nor learners with visual text pacing had a significant influence. Thus, there are different explanations for these findings: First, with longer and complex texts the modality principle even with a system-paced presentation does not hold, because the auditory text is difficult to process due to its transientness and therefore leads to an reduction in performance similar to the spatial split attention with visual text. Second, even if learners with visual text and self pacing have the *opportunity* to apply text processing strategies, it does not necessarily mean that they will really use it. Third, it might also be that the material does not fit the criteria for multimedia material; rather, text and picture may have been redundant, so that one source of information is sufficient to

answer the test items (see Sweller et al., 1998). Currently, we are investigating the last aspect by comparing two groups with pictures with two groups without pictures.

General Discussion

To summarize, two questions arise with regard to the modality effect: What causes the modality effect and under what conditions does the modality effect appear? The first question is still difficult to answer. We reported a study in which we tried to differentiate between the temporal contiguity and the modality-specific explanation. The results do not confirm any of the two explanations. An alternative explanation might be absent spatial contiguity with visual text and picture. More research is needed with regard to this aspect. The second question asked under what conditions a modality effect arises. We argued that with longer and complex texts and more time on task available the modality effect should disappear, because the opportunity to apply text processing strategies may counteract other potential drawbacks of visual text. This assumption currently could not be confirmed. Moreover, also with restricted time on task (i.e., system paced condition) no modality effect was found. It remains unclear, if the modality effect really does not appear with longer and complex texts or if the material has been inappropriate to find a modality effect.

The results are somewhat disappointing, because they give no clear response to our questions. However, they demonstrate that a simple principle like "always present pictures with auditory text" is not valid. Therefore, more research is needed, if we want to shed light on the theoretical explanation and the boundary conditions of the modality effect.

References

Baddeley, A. D. (1992). Working memory. *Science, 255*, 556-559.

Baggett, P., & Ehrenfeucht, A. (1983). Encoding and retaining information in the visuals and verbals of an educational movie. *Educational Communications and Technology Journal, 31*, 23-32.

Bazerman, C. (1985). Physicists reading physics: Schema laden purposes and purpose laden schema. *Written Communication, 2*, 3-24.

Byrne, M., & Curtis, R. (2000). Designing health communication: Testing the explanations for the impact of communication medium on effectiveness. *British Journal of Health Psychology, 5*, 189- 199.

Furnham, A. (2001). Remembering stories as a function of the medium of presentation. *Psychological Reports, 89*, 483-486.

Furnham, A., de Siena, S., & Gunter, B. (2002). Children's and adults recall of children's news stories in both print and audio-visual presentation modalities. *Applied Cognitive Psychology, 16*, 191-210.

Ginns, P. (2005). Meta-analysis of the modality effect. *Learning & Instruction, 15*, 313-331.

Kintsch, W., Kozminsky, E., Streby, J. W., McKoon, G., & Keenan, J. (1975). Comprehension and recall of text as a function of content variables. *Journal of Verbal Learning and Verbal Behavior, 14*, 196-214.

Mayer, R. E. (2001). *Multimedia learning*. Cambridge: Cambridge University Press.

Mayer, R. E., Moreno, R., Boire, M., & Vagge, S. (1999). Maximizing constructivist learning from multimedia communications by minimizing cognitive load. *Journal of Educational Psychology, 91*, 638-643.

Moreno, R., & Mayer, R. E. (1999). Cognitive principles of multimedia learning: The role of modality and contiguity. *Journal of Educational Psychology, 91*, 358-368.

Müsseler, J., Rickheit, G., & Strohner, H. (1985). Influences of modality, text difficulty and processing control on inferences in text processing. In G. Rickheit & H. Strohner (Eds.), *Inferences in text processing* (S. 247-265). Amsterdam: North-Holland.

Rummer, R., Schweppe, J., Scheiter, K., & Gerjets, P. (in press). Lernen mit Multimedia: Die kognitiven Grundlagen des Modalitätseffekts. *Psychologische Rundschau.*

Sannomiya, M. (1981). The modality effect on text processing as a function of presentation rate. *Japanese Journal of Psychonomic Science, 3,* 17-20.

Tabbers, H. K. (2002). *The modality of text in multimedia instructions - refining the design guidelines.* Doctoral Dissertation, Open University of the Netherlands, Heerlen.

Tiene, D. (2000). Sensory mode and information load: Examining the effects of timing on multisensory processing. *International Journal of Instructional Media, 27*, 183- 197.

Sweller, J., van Merriënboer, J. J. G., & Paas, F. G. W. C. (1998). Cognitive architecture and instructional design. *Educational Psychology Review, 10*, 251-296.

Acknowledgments

The research reported in this paper was supported by the Deutsche Forschungsgemeinschaft as part of the research group "Orchestrating Computer-supported Processes of Learning and Instruction".