Does a Lack of Contiguity with Visual Text Cause the Modality Effect in Multimedia Learning?

Anne Schueler (a.schueler@iwm-kmrc.de)
Department of Applied Cognitive Psychology and Media Psychology, Konrad-Adenauer-Strasse 40, 72072 Tuebingen, Germany

Katharina Scheiter (k.scheiter@iwm-kmrc.de)
Department of Applied Cognitive Psychology and Media Psychology, Konrad-Adenauer-Strasse 40, 72072 Tuebingen, Germany

Peter Gerjets (p.gerjets@iwm-kmrc.de)
Multimedia and Hypermedia Research Unit, Knowledge Media Research Center, Konrad-Adenauer-Strasse 40, 72072 Tuebingen, Germany

Ralf Rummer (r.rummer@mx.uni-saarland.de)
Department of Social Psychology, Postfach 151150, 66041 Saarbruecken, Germany

Abstract

In the reported study it was examined whether the modality effect in multimedia learning is caused either by more cognitive resources being available when studying pictures accompanied by auditory text or by the need to split attention between written text and pictures. Ninety-eight students learned about the development of tornados. They heard or read the verbal explanation concurrently with, before, or after the corresponding picture. Learning was measured by verbal recall, pictorial recall, and transfer. For pictorial recall a modality effect was found with simultaneous presentation but not with sequential presentation. This indicates that the modality effect is caused by a split of attention with visual text and picture and not by enhanced cognitive resources with auditory text and picture.

Keywords: multimedia; modality effect; temporal contiguity; spatial contiguity;

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, 2005). According to one of these recommendations, the so called modality principle, verbal information should be presented as spoken (auditory) instead of written (visual) text 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. 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, it is yet unclear how to explain the modality effect. In the next section two alternative explanations will be sketched.

Theoretical Explanations of the Modality Effect

The modality effect is often explained against the background of the Cognitive Theory of Multimedia Learning (Mayer, 2005). 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, improves performance. In the following, this explanation for the modality effect will be called the cognitive resources 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 Gerjets, Rummer, Scheiter, & Schweppe, 2007).

An alternative explanation for the modality effect is based on 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, that is, no contiguity between text and picture is given. Mayer (2001) distinguishes two types of split attention that need to be avoided in multimedia design: 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 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) rather 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 a high level of 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. It is important to note that according to temporal contiguity principle the absolute amount of time is not the crucial point, that is, also sequential presentation provides more time to learn, it results in worse learning performance.

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 additional 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: Enhanced Cognitive Resources or Lack of Contiguity?

A first study that aimed at differentiating between enhanced cognitive resources 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., cognitive resources explanation) or that learners need to split visual attention in the case of written text (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 cognitive resources 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 Gerjets et al., 2007).

In Experiment 2, Moreno and Mayer (1999) investigated whether the modality effect can be traced back to the cognitive resources explanation or to the temporal contiguity explanation, that is, to the fact that with auditory text a simultaneous processing of verbal and pictorial information is possible but not with visual text. If the cognitive resources 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 cognitive resources 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 for this lack of a temporal contiguity effect 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 cognitive resources 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.

**Method**

**Participants and Design.** Ninety-eight participants with different educational backgrounds participated in the study for either payment or course credit. Six of the participants had to be excluded from the analysis because of extreme values in prior knowledge. The remaining ninety-two participants ($M = 24.67$ years, $SD = 6.39$ years; 61 female and 31 male) were randomly assigned to one of six conditions, which resulted from a 2 x 3 design, with text modality (auditory vs. visual text) and temporal contiguity of presentation (simultaneous vs. text-picture presentation vs. picture-text presentation) as independent variables. We also varied the picture and text sequence to control for possible influences of the sequence. The empirical evidence for an influence of the sequence is still inconsistent. Whereas some studies found an influence of the sequence of presentation (e.g. in favor of picture-text presentation: Verdi & Johnson, 1997), other studies did not (e.g., Moreno & Mayer, 1999). The participants of the six groups did not differ with regard to prior knowledge ($M = 45.42$ percent, $SD = 18.10$ percent).

**Materials.** The material consisted of a computerized learning environment and paper-pencil material.

The computerized learning environment consisted of eight static pictures and related 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 A). In the auditory text-picture condition, the learners heard the auditory text first while the screen remained blank. Then the picture was presented. In the visual text-picture condition, the visual text was presented on a screen, than the picture was presented on a new screen (see Figure 1, part B).

**Table 1: Predictions of the two explanations with regard to modality effect**

<table>
<thead>
<tr>
<th></th>
<th>simultaneous</th>
<th>sequential</th>
</tr>
</thead>
<tbody>
<tr>
<td>enhanced resources auditory &gt; visual</td>
<td>auditory &gt; visual</td>
<td></td>
</tr>
<tr>
<td>temporal contiguity auditory &gt; visual</td>
<td>auditory = visual</td>
<td></td>
</tr>
</tbody>
</table>

By manipulating the sequence of text and picture on the one hand and the text modality on the other hand it is possible to differentiate between the temporal contiguity explanation and the cognitive resources explanation. Table 1 gives an overview over the made predictions.

**Figure 1: Frames selected from the presentation.**
In the picture-auditory text condition, the learners first saw the picture, than heard the auditory text while the screen was blank. In the picture-visual text condition, first the picture was presented, than the text was presented on the screen. Overall, eight screens were presented in the simultaneous conditions, whereas sixteen screens were presented in the sequential conditions. The presentations were 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 (picture-text and text-picture) it lasted six minutes.

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 chosen from the later recall test (e.g., “What is the top speed of a tornado?”). The answer alternatives were: “50-450 km/h”, “100-600 km/h”, “400-900 km/h”, “more than 1000 km/h”). 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, which depended on the specific item; 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 (e.g. “Why is the season important for the development of a tornado?”). For each correct answer the learners received one point. The time for responding to the tests was unlimited.

**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 six experimental conditions. Fifth, they answered the questionnaire, which measured recall and transfer performance. A single experimental session 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 2 shows the means for each group. Because the control variables did not affect learning outcome, no analyses of covariance were conducted.

<table>
<thead>
<tr>
<th></th>
<th>simultaneous</th>
<th>text-picture</th>
<th>picture-text</th>
</tr>
</thead>
<tbody>
<tr>
<td>vR</td>
<td>71.88</td>
<td>70.83</td>
<td>75.96</td>
</tr>
<tr>
<td></td>
<td>(15.31)</td>
<td>(16.14)</td>
<td>(16.90)</td>
</tr>
<tr>
<td>pR</td>
<td>70.10</td>
<td>61.39</td>
<td>71.79</td>
</tr>
<tr>
<td></td>
<td>(20.16)</td>
<td>(11.06)</td>
<td>(19.11)</td>
</tr>
<tr>
<td>T</td>
<td>45.59</td>
<td>48.44</td>
<td>51.92</td>
</tr>
<tr>
<td></td>
<td>(29.63)</td>
<td>(22.09)</td>
<td>(24.86)</td>
</tr>
</tbody>
</table>

Note. vR = verbal Recall, pR = pictorial Recall, T = Transfer, aud = auditory, vis = visual.

For recall of verbal information, neither the difference between the visual groups and the auditory groups ($F$(1,83) = 1.73, $p = .19$) nor the influence of temporal contiguity was significant ($F$(2,83) = 1.39, $p = .26$). There was no interaction between the two factors ($F < 1$).

For recall of pictorial information there neither was a main effect for text modality ($F < 1$) or temporal contiguity ($F$(2,85) = 1.68, $p = .19$). However the interaction between the two factors was significant ($F$(2,85) = 3.98, $p = .02$). A Bonferroni posthoc analyses yielded a significant modality effect in the simultaneous condition ($F$(1,85) = 6.19, $p = .02$), whereas in the sequential conditions no modality effect appeared (text-picture presentation: $F$(1,85) = 2.28, $p = .14$; picture-text presentation: $F < 1$). Interestingly, within the auditory groups there was no difference between learners with simultaneous and sequential presentation ($F$(2,85) = 1.35, $p = .26$), whereas within the visual text groups an influence of the temporal contiguity was found ($F$(2,85) = 4.24, $p = .02$). In particular, learners with simultaneous presentation performed worse than learners with text-picture presentation ($p = .03$) and slightly worse than learners with picture-text presentation ($p = .06$). Between the picture-text and the text-picture presentation no difference was found (n.s.). These results indicate that with visual presentation an inverted temporal contiguity effect was found, whereas with auditory presentation no temporal contiguity effect occurred.

For transfer, the ANOVAs showed no difference between the visual and the auditory groups or between the groups with different presentation sequence or an interaction (all $F$s $< 1$).

**Discussion**

The purpose of the reported study was to differentiate between two possible explanations for the modality effect in multimedia learning, the cognitive resources explanation and the temporal contiguity explanation. The cognitive resources explanation assumes that auditory text makes available additional cognitive resources, which in turn improves performance compared to learning with visual text...
and picture. The temporal contiguity explanation assumes that the lack of contiguity with visual text and picture can account for the modality effect. With regard to this explanation a modality effect should occur if contiguity is given with auditory text and picture, but should not occur if contiguity is absent also with auditory text and picture. Thus, there should be an interaction between text modality and contiguity of presentation. To test which of the two explanations is valid we varied the text modality (auditory vs. visual) and the contiguity of presentation (simultaneous vs. text-picture vs. picture-text).

The results indicated that there was no general modality effect for any of the learning outcome measures. The absence of a general modality effect contradicts the cognitive resources explanation, according to which a 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 auditory text as it is assumed in the Cognitive Theory of Multimedia Learning (Mayer, 2001).

For pictorial recall an interaction between the text modality and the contiguity of presentation was demonstrated. The found pattern confirmed the temporal contiguity explanation: In the case of a simultaneous presentation – that is, with contiguity of auditory text and picture being given – learners with auditory text scored higher than learners with visual text, whereas with sequential presentation – that is, no contiguity of auditory text and picture – there was no difference between the learners with auditory and visual text. At first view, it seems as if the temporal contiguity explanation could be confirmed. However, we could not replicate the temporal contiguity within the auditory groups, despite the fact that the text segments had been rather long. Accordingly, the superiority of the auditory simultaneous group over the visual simultaneous group cannot be explained with temporal contiguity.

Thus, how might the interaction for pictorial recall be explained? In our view, an alternative explanation might be the spatial contiguity principle, which refers to the presentation of visual text and picture and states that students learn better when corresponding visual text and pictures are presented spatially integrated instead of spatially separated. With regard to the applied learning environment, spatial contiguity was absent in the simultaneous presentation of visual text and picture: learners with visual text had to read the text and to look at the picture and to relate them to each other, whereas with auditory text and picture such a split of attention was not necessary. Thus, the modality effect occurred, because within the visual text group with simultaneous presentation visual search processes were required, which the group with auditory text did not have to undertake. Within the sequential groups no spatial split was given, which resulted in the same performance for learners either with visual or auditory text. As noted before, the study of Moreno and Mayer (1999) tried to differentiate between the cognitive resources explanation and the spatial contiguity explanation by varying the text modality and the proximity of the visual text to the picture. Because they did not report the comparison between the auditory text and integrated visual text, it remained unclear if spatial contiguity can account for the modality effect. Thus, further research is needed to examine this possible explanation.

Due to the fact that for verbal recall and transfer no interaction between text modality and temporal contiguity could be found, the assumed spatial split of attention with simultaneous presentation can be specified: it seems as if learners with visual text and simultaneous presentation allocated more attention to the text than to the picture. This assumption is supported by eye tracking studies, which found that in multimedia learning environments learners first read the text and then look at the picture (see for example, Hegarty, 1992). This might explain why the learners had fewer difficulties in answering the verbal items, which primarily relied on text information than in answering the pictorial items. With sequential presentation, where there was more time to process the information, they did not have to split their attention and could rely on both, picture and visual text. Thus, with sequential presentation also the visual text group could process the pictorial information adequately, resulting in no modality effect for pictorial recall with sequential presentation. This indicates that for visual text the time for processing might be crucial.

It should be noted that this assumption at first view contradicts the temporal contiguity principle which claims that simultaneous presentation leads to better learning outcomes, because the connections between pictorial and verbal information can be made more easily. But, as just noted, the temporal contiguity assumption refers to auditory text, not to visual text. Thus, the results confirm the assumption that the temporal contiguity principle is not valid when presenting visual text and picture, as it could also be shown by Michas and Berry (2000). Rather, we even found an inverted temporal contiguity effect for pictorial recall with visual text, that is, learners with sequential presentation showed better pictorial recall than learners with simultaneous presentation. This confirms on the one hand that learners with simultaneous presentation and visual text did not allocate much attention to the pictures, resulting in worse performance with regard to the sequential presentation. On the other hand, this may indicate that with enough time visual text might even support the processing of pictures, because it leads to the construction of a more stable text based mental model due to the possibility to apply text processing strategies like slowing down the reading rate (Byrne & Curtis, 2000), reading through sections of the text several times (e.g., Furnham, de Siena, & Gunter, 2002) or skipping text passages which are easy or not relevant to the topic (Bazerman, 1985).

The assumptions made concerning the patterns of split attention lead to the following explanation of the found results: Learners with visual text and simultaneous presentation allocated their attention towards reading the
text, perhaps trying to refer to the picture if it seemed necessary. This led to worse processing of the picture, because less time was allocated to it. This assumption might explain why there was a modality effect for pictorial recall but not for verbal recall. It might also explain the found inverted temporal contiguity effect for visual text, 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 worse learning outcomes, when picture and text were presented sequentially, which contradicts the temporal contiguity principle and leads to problems with the interpretation of modality effect in terms of lack of temporal contiguity.

To summarize, we wanted to clarify the theoretical basis of the modality effect. Therefore, we conducted a study in which we tried to differentiate between the temporal contiguity and the cognitive resources 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.

Furthermore, it should be noted that for no of the dependent variables the modality effect or the temporal contiguity effect (for auditory text) could be replicated. That demonstrates that simple principles like “always present pictures with auditory text” or “always present auditory text and picture simultaneously” are not valid. To shed light on the conditions under which the principles may be effective or even not it is important to study the nature of the human mind more precisely. To do so, you should refer to methods of cognitive psychology (e.g., by applying dual task experiments) and you should consider moderating variables like text length or picture complexity, which might lead to different processing strategies of the presented material. It is also important to go beyond basic research by using more applied learning materials in more applied learning scenarios (e.g. classroom environments) to validate the results and to take in consideration other possible moderating variables. The results also show that research is still needed to clarify the theoretical basis of the principles. The assumptions made by Mayer (2005) concerning the architecture of the human mind are not sufficient to explain the found results. Thus, a theoretical differentiation of how the human mind works with regard to the processing of multimedia materials is necessary.

Acknowledgments
The research reported in this paper was supported by the Deutsche Forschungsgemeinschaft as part of the collaborative research project “Multimeda Lernen und die kognitiven Grundlagen des Modalitätseffektes” (PIs: Ralf Rummer, Katharina Scheiter, Peter Gerjets).

References