Reverse Modality Effects on Spatial Knowledge Instruction: When Reading is Better than Listening

Jongpil Cheon
Steven M. Crooks
Fethi Inan
Raymond Flores
Fatih Ari
Texas Tech University

Abstract

This study investigated the reversed modality principle in spatial learning content with two different modality conditions. Participants were randomly assigned to two groups (visual text and spoken text). The findings revealed no significant differences in terms of mental effort for the instruction and assessments, the usability level, and perceived usefulness. However, the significant effects on three assessments showed that the visual text group performed better than the spoken text group. The results support a reverse modality effect. This study provides theoretical support for establishing boundaries for the modality principle as well as practical implication for instructional designers.

Theoretical Framework and Hypothesis

The modality principle in multimedia learning has received considerable research support over the last few decades (Ginns, 2005; Mayer, 2001; Sweller, van Merrienboer, & Paas, 1998). This principle states that learning from words and pictures is improved when written or on-screen text is replaced with spoken text. A theoretical rationale for this principle is provided by Baddeley’s (1992) model of working memory. According to Baddeley, working memory contains two sub-systems, one for processing visual information and another for processing verbal information. Presenting textual information visually (as on-screen text) during multimedia learning is purported to overload the visual subsystem and strain attentional resources. This occurs because of the need to temporarily hold and process text along with pictorial information (e.g., animation) in the same memory subsystem (Mayer & Moreno, 1998). However, according to the modality principle, this unimodal presentation format can be improved by employing a bimodal format, wherein textual information is presented auditorily and pictorial information is presented visually. This presentation format is purported to reduce cognitive load by using the total capacity of working memory (both visual and verbal subsystems) more efficiently (Tabbers, Martens, & van Merrienboer, 2004). The modality principle has been linked to reduced mental effort and study time during instruction and to improved performance on retention, transfer, and matching tests (Tabbers et al., 2004).

The modality principle has been linked to reduced mental effort and study time during instruction and to improved performance on retention, transfer, and matching tests (Ginns, 2005; Tabbers et al., 2004). The modality principle has also been validated across a variety of computer-based media, such as multimedia explanations, agent based computer games, and virtual reality (Moreno, 2006).

Notwithstanding the research support for the modality principle, recent research has shown that the principle may not apply to all multimedia learning situations. For example, Tabbers et al. (2004) found that a visual presentation of text was superior to a spoken presentation when learners were given control over the pacing of the instruction. There is also evidence that visual text may be superior to spoken text if the subject matter pertains to learning spatial relations (Penny, 1989).

In the current study we examined the generalizability of the modality principle to an instructional situation that prior research suggests may be conducive to a reverse modality effect. That is, where on-screen text is superior to spoken text. To create this situation the following elements were incorporated into the treatment materials: 1) learner control of instructional content, and 2) a learning task with a significant spatial component. Our hypothesis was that learners studying on-screen text would outperform those studying spoken text due to the presence of learning conditions favorable to a reverse modality effect in multimedia learning.
Method

This study investigated the effects of modality on spatial knowledge learning from a computer-based diagram and related text. Modality (visual text vs. spoken text), see Figure 1, were examined between two groups. One hundred and seventeen undergraduate students from a large southwestern university were randomly assigned to one of the two experimental conditions.

<table>
<thead>
<tr>
<th>Modality</th>
<th>Visual text</th>
<th>Spoken text</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Figure 1.</strong> Screen shots from the visual (left) and spoken text conditions</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The experimental materials consisted of a computer-presented diagram depicting 12 places of articulation in human speech. The participants accessed the text by selecting a place of articulation on the diagram with a computer mouse. In the visual text condition, descriptive sentences were presented as on-screen text; in the spoken-text condition the text was presented as narration. The participants had complete control over the pace and sequence with which they selected the hyperlinked text. They were instructed to study the diagram and the related text for 10 minutes. After 10 minutes of study, the participants were given the survey, reconstruction, labeling, and matching assessments.

The dependant measures consisted of the number of clicks on description dots, the number of clicks on replay buttons, mental effort, the usability level, the usefulness of either visual or spoken text, a diagram reconstruction test, a special labeling test, and a matching test. All dependant measures were delivered via computer. In the instruction, there are 12 different round markers in the diagram as shown in Figure 1. Once they click a marker, they can see the written text or listen to the spoken text in a pop-up box. Both types of text will stay for 25 seconds, and a review or replay button will be shown after the written text disappears or the spoken text finishes. See Figure 2. The two different numbers of clicks tracked how many times participants revisit or replay the descriptions for each place.
Subjective mental effort ratings were obtained after the experimental treatment and after each of the three assessments. These ratings measured (on a 7-point Likert scale) the amount of cognitive resources the participants perceived they invested in the instruction and the three assessments. This subjective measurement scale has been used to calculate instructional efficiency scores in many studies investigating cognitive load theory (e.g., Kalyuga & Sweller, 2005; Tuovinen & Paas, 2004). The level of usability measured how easily participants operated or navigated the interface. The usefulness of text types represented participants’ perceptions toward either visual text or spoken text. The reconstruction test presented each participant with a blank outline of the Places of Articulation diagram and a listing of the 12 places of articulation next to the diagram. Each participant was required to drag the name of each place to its correct location on the diagram. We measured the distance between correct locations and the locations participants moved. Similar to the reconstruction test, the spatial labeling test provide the unlabeled places in the articulation diagram and the list so that participants can move a description to the right place. This test measured only right or wrong places. Finally, the matching test consisted of a listing of the 12 places of articulation (e.g., Dental) and a corresponding listing of the sounds made at each place (e.g., “th” as in thunder), plus three distracters. The participants were required to match each place with its corresponding sound.

### Results and Conclusions

Independent t-Tests were conducted to compare all variables between groups. The results revealed that there were no significant differences in terms of mental effort, usability, and usefulness. However, five dependent variables had significant differences. Regarding the numbers of clicks, the numbers of clicks for the descriptions (M = 36.80, \( p < .001 \)) and for the replay button (M = 1.39, \( p = .047 \)) in the visual text group than the numbers of clicks (M = 17.39 for the description, M = .82 for the replay) in spoken text group. In addition, the visual text group had higher scores in all test scores as shown in Table 1. Note that the values of the reconstruction test represent the distance from the correct place, so lower numbers mean higher scores. In summary, we observed a reverse modality effect for all dependent measures, as hypothesized.

<table>
<thead>
<tr>
<th>Test</th>
<th>Visual Text Mean scores (SD)</th>
<th>Spoken Text Mean scores (SD)</th>
<th>Independent t-test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reconstruction</td>
<td>66.14 (51.75)</td>
<td>90.61 (52.72)</td>
<td>( t(115) = -2.533, p = .013 )</td>
</tr>
<tr>
<td>Spatial labeling</td>
<td>8.87 (2.96)</td>
<td>7.57 (3.42)</td>
<td>( t(115) = 2.197, p = .030 )</td>
</tr>
<tr>
<td>Matching</td>
<td>3.54 (2.59)</td>
<td>2.60 (1.66)</td>
<td>( t(115) = 2.294, p = .024 )</td>
</tr>
</tbody>
</table>

The results of this study support the results of the study by Crooks et al. (2009) and Tabbers et al. (2004). Modality effects are not likely to occur with diagrams and text when learners have control over the pacing and sequence of their study and spatial learning is an important instructional outcome. In fact, we observed a reverse modality effect favoring visual text over spoken text. For example, the visual text group had more control over pacing while the spoken text group had to stay for 25 seconds to listen to whole narrations. So, the visual text group had more chance to revisit and review the description. The best explanation appears to be that user control and spatial learning are conditions favorable to learning from on-screen text. These results will guide practitioners during the design and development of multimedia materials and to theorists as they seek to clarify the boundaries of the modality effect in multimedia learning.

### References


