CHAPTER OUTLINE

Modality Principle: Present Words as Speech Rather Than On-Screen Text
Limitations to the Modality Principle
Psychological Reasons for the Modality Principle
Evidence for Using Spoken Rather Than Printed Text
When the Modality Principle Applies
THE MODALITY PRINCIPLE

What’s New in This Chapter?

The Modality Principle has the most research support of any of the principles described in this book. Technical constraints on the use of audio in e-learning may lead consumers or designers of e-learning to rely on text to present content and describe visuals. However, when it’s feasible to use audio, there is considerable evidence that presenting words in audio rather than on-screen text can result in significant learning gains. In this chapter, we summarize the empirical evidence for learning gains that result from using audio rather than on-screen text to describe graphics. To moderate this guideline, we also describe a number of situations in which memory limitations require the use of text rather than audio. The psychological advantage of using audio presentation is a result of the incoming information being split across two separate cognitive channels—words in the auditory channel and pictures in the visual channel—rather than concentrating
both words and pictures in the visual channel. What is new in this chapter is an update to the evidence reported in the second edition of *e-Learning and the Science of Instruction*, including extensions of the modality principle to classroom contexts and supporting evidence from eye-tracking studies. We also have added more discussion of the boundary conditions for the modality principle—that is, the situations in which it applies most strongly. Overall, there continues to be strong and consistent support for using narration rather than on-screen text to describe graphics, especially when the presentation is complex or fast-paced and when the verbal material is familiar or short. In particular, audio narrations must be brief and clear to be effective.

**DESIGN DILEMMA: YOU DECIDE**

Now that they have agreed on the value of adding relevant visuals, as described in Chapter 4, the Excel design team has bogged down in discussions about how best to explain those graphics. Reshmi, the instructional designer, believes that providing words in text, as shown in Figure 6.1, allows learners to move at their own pace rather than have to wait for audio to play. “Besides that,” she adds, “we must meet 508 compliance to accommodate learners with hearing loss. We must provide words in text!” Matt, the project leader, also prefers using text, as file sizes will be smaller and updates will be easier. However, Michael, a graduate student in multimedia learning who is interning from the local university, disagrees strongly: “In our class last semester, the professor went on and on about the benefits of audio. You are losing a big learning opportunity if you rely on text alone!” Based on your experience or intuition, which option(s) do you select:

A. Reshmi and Matt are right. The advantages of explaining on-screen graphics with text outweigh the disadvantages.

B. Michael is right. Learning is much better when words are presented in audio narration.

C. Everyone can be accommodated by providing words in both text and audio.

D. Not sure which options are correct.
Modality Principle: Present Words as Speech Rather Than On-Screen Text

Suppose you are presenting a verbal explanation along with an animation, video, or series of still frames. Does it matter whether the words in your multimedia presentation are represented as printed text (that is, as on-screen text) or as spoken text (that is, as narration)? What do cognitive theory and research evidence have to say about the modality of words in multimedia presentations? You'll find the answer to these questions in the next few sections of this chapter.

Based on cognitive theory and research evidence, we recommend that you put words in spoken form rather than printed form whenever the graphic (animation, video, or series of static frames) is the focus of the words and both are presented simultaneously. Thus, we recommend that you avoid
e-learning courses that contain crucial multimedia presentations where all words are in printed rather than spoken form, especially when the graphic is complex, the words are familiar, and the lesson is fast-paced.

The rationale for our recommendation is that learners may experience an overload of their visual/pictorial channel when they must simultaneously process graphics and the printed words that refer to them. If their eyes must attend to the printed words, they cannot fully attend to the animation or graphics—especially when the words and pictures are presented concurrently at a rapid pace, the words are familiar, and the graphic is complex. Since being able to attend to relevant words and pictures is a crucial first step in learning, e-learning courses should be designed to minimize the chances of overloading learners’ visual/pictorial channel.

Figure 6.2 illustrates a multimedia course that effectively applies the modality principle. This section of the lesson is providing a demonstration of how to use a new online telephone management system. As the animation

---

**Figure 6.2. Audio Explains the Animated Demonstration of the Telephone System.**

Audio: While Bill is talking to Don, Julie calls with a question. Bill knows that Julie needs to talk to Sally in the Art Department and decides to transfer her while he is talking to Don.
Chapter 6: Applying the Modality Principle

Figure 6.3. Visual Described by Audio Narration.

<table>
<thead>
<tr>
<th>Audio</th>
<th>Off</th>
<th>On</th>
</tr>
</thead>
</table>

Illustrates the steps on the computer screen, the audio describes the actions of the user. Another good example is seen in Figure 6.3 from our Excel sample lesson. Audio narration describes the visual illustration of formatting an absolute cell reference in Excel. In both of these examples, the visuals are relatively complex, and therefore using audio allows the learner to focus on the visual while listening to the explanation.

Limitations to the Modality Principle

When simultaneously presenting words and the graphics explained by the words, use spoken rather than printed text as a way of reducing the demands on visual processing. We recognize that in some cases it may not be practical to implement the modality principle, because the creation of sound may involve technical demands that the learning environment cannot meet (such
as bandwidth, sound cards, headsets, and so on), or may create too much noise in the learning environment. Using sound also may add unreasonable expense or may make it more difficult to update rapidly changing information. We also recognize the recommendation is limited to those situations in which the words and graphics are simultaneously presented, and thus does not apply when words are presented without any concurrent picture or other visual input.

Additionally, there are times when the words should remain available to the learner for memory support—particularly when the words are technical, unfamiliar, not in the learner’s native language, or needed for future reference. For example, a mathematical formula may be part of an audio explanation of an animated demonstration, but because of its complexity, it should remain visible as on-screen text. Key words that identify the steps of a procedure may be presented by on-screen text and highlighted (thus used as an organizer) as each step is illustrated in the animation and discussed in the audio. Another common example involves the directions to a practice exercise. Thus, we see in Figure 6.4 (from an Excel virtual classroom session)

Figure 6.4. Practice Directions Provided in On-Screen Text in a Virtual Classroom Session.
that the instructor narration throughout most of the program is suspended when the learner comes to the practice screen. Instead, the directions to the practice remain in text in the box on the spreadsheet for reference as the learners complete the exercise.

One advantage to virtual classrooms is the use of instructor speech to describe graphics projected on the whiteboard or through application sharing. In virtual classroom sessions, participants hear the instructor either through telephone conferencing or through their computers via voice-over-IP. However, virtual classroom facilitators should be careful to place text on their slides for instructional elements such as practice directions, memory support, and technical terms.

Psychological Reasons for the Modality Principle

If the purpose of the instructional program is to present information to the learner, then it does not matter whether you present graphics with printed text or graphics with spoken text. In both cases, identical pictures and words are presented, so it does not matter whether the words are presented as printed text or spoken text. This approach to multimedia design is suggested by the information acquisition view of learning—the idea that the instructor’s job is to present information and the learner’s job is to acquire information. Following this view, the rationale for using on-screen text is that it is generally easier to produce printed text rather than spoken text and it accomplishes the same job—that is, it presents the same information.

The trouble with the information acquisition view is that it conflicts with much of the research evidence concerning how people learn. This book is based on the idea that the instructional professional’s job is not only to present information, but also to present it in a way that is consistent with how people learn. Thus, we adopt the cognitive theory of multimedia learning, in which learning depends both on the information that is presented and on the cognitive processes used by the learner during learning.

Multimedia lessons that present words as on-screen text can create a situation that conflicts with the way the human mind works. According to the cognitive theory of learning—which we use as the basis for our
recommendations—people have separate information processing channels for visual/pictorial processing and for auditory/verbal processing. When learners are given concurrent graphics and on-screen text, both must be initially processed in the visual/pictorial channel. The capacity of each channel is limited, so the graphics and their explanatory on-screen text must compete for the same limited visual attention. When the eyes are engaged with on-screen text, they cannot simultaneously be looking at the graphics; when the eyes are engaged with the graphics, they cannot be looking at the on-screen text. Thus, even though the information is presented, learners may not be able to adequately attend to all of it because their visual channels become overloaded.

In contrast, we can reduce this load on the visual channel by presenting the verbal explanation as speech. Thus, the verbal material enters the cognitive system through the ears and is processed in the auditory/verbal channel. At the same time, the graphics enter the cognitive system through the eyes and are processed in the visual/pictorial channel. In this way neither channel is overloaded but both words and pictures are processed.

The case for presenting verbal explanations of graphics as speech is summarized in Figures 6.5 and 6.6. Figure 6.5 shows how graphics and on-screen text can overwhelm the visual channel, and Figure 6.6 shows how graphics and speech can distribute the processing between the visual and auditory channels. This analysis also explains why the case for presenting words as speech only applies to situations in which words and pictures are presented.
simultaneously. As you can see in Figure 6.5, there would be no overload in the visual channel if words were presented as on-screen text in the absence of concurrent graphics that required the learner’s simultaneous attention.

Evidence for Using Spoken Rather Than Printed Text

Do students learn more deeply from graphics with speech (for example, narrated animation) than from graphics with on-screen text (for example, animation with on-screen text blocks), as suggested by cognitive theory? Researchers have examined this question in several different ways, and the results consistently support our recommendation. Let’s consider several studies that compare multimedia lessons containing animation with concurrent narration versus animation with concurrent on-screen text, in which the words in the narration and on-screen text are identical. Some of the multimedia lessons present an explanation of how lightning forms, how a car’s braking system works, or how an electric motor works (Craig, Gholson, & Driscoll, 2002; Mayer, Dow, & Mayer, 2003; Mayer & Moreno, 1998; Moreno & Mayer, 1999a). Others are embedded in an interactive game intended to teach botany (Moreno, Mayer, Spires, & Lester, 2001; Moreno & Mayer 2002b), and a final set are part of a virtual reality training episode concerning the operation of an aircraft fuel system (O’Neil, Mayer, Herl, Niemi, Olin, & Thurman, 2000).
For example, in one study (Moreno & Mayer, 1999a) students viewed an animation depicting the steps in lightning formation along with concurrent narration (Figure 6.7) or concurrent on-screen text captions (Figure 6.8). The words in the narration and the on-screen text were identical, and they were presented at the same point in the animation. On a subsequent test in which students had to solve transfer problems about lightning, the animation-with-narration group produced more than twice as many solutions to the problems as compared to the animation-with-text group, yielding an effect size greater than 1. The results are summarized in Figure 6.9. We refer to this finding as the modality effect—people learn more deeply from multimedia lessons when words explaining concurrent graphics are presented as speech rather than as on-screen text.
Figure 6.8. Screens from Lightning Lesson Explained with On-Screen Text.

From Moreno and Mayer, 1999a.

The charge results from the collision of the cloud's rising water droplets against heavier, falling pieces of ice.

The negatively charged particles fall to the bottom of the cloud, and most of the positively charged particles rise to the top.

A stepped leader of negative charges moves downward in a series of steps. It nears the ground.

A positively charged leader travels up from such objects as trees and buildings.

The two leaders generally meet about 605 feet above the ground.

Negatively charged particles then rush from the cloud to the ground along the path created by the leaders. It is not very bright.

As the leader stroke nears the ground, it induces an opposite charge, so positively charged particles from the ground rush upward along the same path.

This upward motion of the current is the return stroke. It produces the bright light that people notice as a flash of lightning.

Figure 6.9. Better Learning When Visuals Are Explained with Audio Narration.

From Marino and Mayer, 1999a.
In a more interactive environment aimed at explaining how an electric motor works, students could click on various questions and for each see a short animated answer along with narration or printed text delivered by a character named Dr. Phyz (Mayer, Dow, & Mayer, 2003). In the frame on the right side of the top screen in Figure 6.10, suppose the student clicks the question, “What happens when the motor is at the start position?” As a result, the students in the animation-with-text group see an animation along with on-screen text, as exemplified in the B frame on the bottom right side of Figure 6.10. In contrast, students in the animation-with-narration group see the same animation and hear the same words in spoken form as narration as in the A frame on the bottom left side of Figure 6.10. Students who received narration generated 29 percent more solutions on a subsequent problem-solving transfer test, yielding an effect size of .85.
A more recent study found that the modality effect applies to students in a high school setting. The students learned better from web-based biology lessons that contained illustrations and narration than for lessons containing illustrations and on-screen text (Harskamp, Mayer, & Suhre, 2007). Replicating the modality effect in a more naturalistic environment such as a high school class boosts our confidence that the guidelines derived from laboratory studies apply to real-world learning environments.

Consistent with cognitive theory, recent eye-tracking studies found that students who viewed animation with narration on lightning formation spent more time looking at the graphics than did students who received animations with on-screen text (Schmidt-Weigand, Kohnert, & Glowalla, 2010a, 2010b). When graphics were described by on-screen text, students were largely guided by the text so processing of the graphics suffered.

Also consistent with cognitive theory, researchers have found that the modality effect is stronger for less-skilled learners than for more-skilled learners (Seufert, Schutze, & Brunken, 2009).

In a review of research on modality, Mayer (2005c) identified twenty-one experimental comparisons of learning from printed text and graphics versus learning from narration and graphics, based on published research articles. The lessons included topics in mathematics, electrical engineering, environmental science, and aircraft maintenance as well as explanations of how brakes work, how lightning storms develop, and how an electric motor works. In each of the twenty-one comparisons, there was a modality effect in which students who received narration and graphics performed better on solving transfer problems than did students who received on-screen text and graphics. The median effect size was .97, which is considered a large effect. Based on the growing evidence for the modality effect, we feel confident in recommending the use of spoken rather than printed words in multimedia messages containing graphics with related descriptive words.

In a somewhat more lenient review that included both published articles and unpublished sources (such as conference papers and theses) and a variety of learning measures, Ginns (2005) found forty-three experimental tests of the modality principle. Overall, there was strong evidence for the modality
effect, yielding an average effect size of .72, which is considered moderate to large. Importantly, the positive effect of auditory modality was stronger for more complex material than for less complex material, and stronger for computer-controlled pacing than for learner-controlled pacing. Apparently, in situations that are more likely to require heavy amounts of essential cognitive processing to comprehend the material—that is, lessons with complex material or fast pacing—it is particularly important to use instructional designs that minimize the need for extraneous processing.

When the Modality Principle Applies

Does the modality principle mean that you should never use printed text? The simple answer to this question is: Of course not. We do not intend for you to use our recommendations as unbending rules that must be rigidly applied in all situations. Instead, we encourage you to apply our principles in ways that are consistent with the way that the human mind works—that is, consistent with the cognitive theory of multimedia learning rather than the information delivery theory. As noted earlier, the modality principle applies in situations in which you present graphics and their verbal commentary at the same time, and particularly, when the material is complex and presented at a rapid continuous pace. If the material is easy for the learner or the learner has control over the pacing of the material, the modality principle becomes less important.

As we noted previously, in some cases words should remain available to the learner over time—particularly, when the words are technical, unfamiliar, not in the learner’s native language, lengthy, or needed for future reference. For example, when you present technical terms, list key steps in a procedure, or are giving directions to a practice exercise, it is important to present words in writing for reference support. When the learner is not a native speaker of the language of instruction or is extremely unfamiliar with the material, it may be appropriate to present printed text. Further, if you present only printed words on the screen (without any corresponding graphic) then the modality principle does not apply. Finally, in some situations people may learn better from multimedia lessons that have a few well-placed printed words along with spoken words, as we describe in the next chapter on the redundancy principle.
What We Don’t Know About Modality

Overall, our goal in applying the modality principle is to reduce the cognitive load in the learner’s visual/pictorial channel (that is, through the eyes) by off-loading some of the cognitive processing onto the auditory/verbal channel (that is, through the ears). Some unresolved issues concern:

1. When is it helpful to put printed words on the screen with a concurrent graphic?
2. Is it helpful to put concise summaries or labels for key components on the screen as printed words?
3. When it is not feasible to provide audio, how can we eliminate any negative effects of on-screen text?
4. Do the negative effects of on-screen text decline over the course of long-term training?

DESIGN DILEMMA: RESOLVED

The Excel design team was in a quandary about use of text and audio in their course. The options presented were:

A. Reshmi and Matt are right. There are many advantages to communicating words as on-screen text.

B. Michael is right. Learning is much better when words are presented in audio narration.

C. Everyone can be accommodated by providing words in both text and audio.

D. Not sure which options are correct.

We recommend that audio narration will promote better learning on screens that include important and detailed graphics as shown in Figure 6.3. Therefore we select Option B. Although Option C might seem like a good compromise, as we will see in the next chapter, using both text and audio to explain a graphic can be problematic. Some elements in the Excel lesson should be presented as text, such as unfamiliar terms and directions for practice exercises.
In this chapter we have seen that learning is improved when graphics or animations presented in e-lessons are explained using audio narration rather than on-screen text. What would be the impact of including both text and narration? In other words, would learning be improved if narration were used to read on-screen text? We will address this issue in the next chapter.

**Suggested Readings**


