



## When using sound with a text or picture is not beneficial for learning

Slava Kalyuga

*The University of New South Wales*

Conventional wisdom tells us that two modalities (visual and auditory) are better than one modality in any instructional message. This paper describes two cases where combining audio explanations with visual instructions has had negative rather than positive or neutral effects. The results were explained as a consequence of working memory overload. Some guiding principles in the design of multimedia instruction are suggested.

As multimedia becomes a commonplace instructional tool, it is subject to more thorough evaluation. Artistic approaches to multimedia design give way to user needs centered engineering approaches. Cost effectiveness also represents a significant concern considering the expenses involved in producing many multimedia presentations. Some common sense beliefs surrounding multimedia are being scientifically tested. One of such belief is about the effectiveness of multiple modalities in instruction.

It is usually taken for granted that instruction employing more than one modality (eg, visual and auditory) is better than equivalent single modality formats. For example, why should adding sound to a text or picture do any harm under any circumstances? However, the value of multiple representations of information has been questioned in some recent publications evaluating the benefits of multimedia instruction (eg, Hegarty, Quilici, Narayanan, Holmquist, & Moreno, 1999; Najjar, 1996; Tergan, 1997). In some cases described in those papers, redundant multimedia did not show the expected positive effects on learning. Surprisingly, there are definable conditions when the addition of an audio explanation to visual instructions has negative rather than positive or neutral effects. Those conditions occur when processing an auditory supplement is likely to impose an excessive working memory load.

Instructional designers should be aware of such conditions to prevent their occurrence in various instructional situations and designs.

There are other, equally definable conditions under which using both auditory and visual modalities is highly beneficial, because the use of both modalities increases the capacity of working memory to handle the information. This paper considers some specific conditions (involving concurrent processing of units of information from several sources) when using an audio explanation with visual instructions would have negative effects on learning, due to working (or short-term) memory overload. Such conditions might occur with various instructional materials (including web based), design models and instructional strategies that contain dual mode (audiovisual) presentations.

### **Cognitive load theory**

We can process only a few elements (chunks) of information in working memory at any one time (Miller, 1956). Too many elements of information may overwhelm working memory, decreasing the effectiveness of instruction. Cognitive load theory (see Sweller, 1999, for a recent summary) places a primary emphasis on working memory capacity limitations as a factor in instructional design. It suggests that information presented to learners should be structured to eliminate any avoidable load on working memory. The theory is based on the basic assumption that a person has a limited processing capacity, and that proper allocation of cognitive resources is critical to learning. Any increase in resources required for various processes not directly related to learning (e.g. integration of information separated over distance or time, or processing redundant information) inevitably decreases resources available for learning. Studies described in this paper provide evidence for some of the consequences derived from these assumptions.

### **Instructional modality effect**

Modern views of working memory suggest that it consists of separate processors for auditory and visual information (Baddeley, 1992; Paivio, 1990; Penney, 1989). The amount of information that can be processed using both auditory and visual channels might exceed the processing capacity of a single channel. Thus, limited working memory may be effectively expanded by using more than one sensory modality making learning easier. For example, a visual diagram accompanied by an auditory text can be more efficient than the equivalent diagram with visually presented (written) text. To understand the instruction, the

learners must mentally integrate the diagram and its associated text. When presented entirely in visual form, the act of mental integration is cognitively demanding because the attention of the learner is split due to search and switching between the diagram and text. In such a situation, increasing effective working memory by presenting the text in auditory form might produce a positive effect on learning.

Thus, using a dual-mode instructional format in which separate sources of information (otherwise requiring mental integration) are presented with text in auditory form, might be beneficial due to cognitive load reduction. For example, it was observed that a visually presented geometry diagram combined with auditory statements enhanced learning compared to conventional, visual only presentations (Mousavi, Low and Sweller, 1995). As another example, an audio text accompanying a visual wiring diagram was superior to purely visually based instructions (Tindall-Ford, Chandler, & Sweller, 1997). Mayer and his associates (Mayer, 1997) have conducted a number of experiments demonstrating the superiority of audio/visual instructions. These studies demonstrated that in many situations, visual textual explanations may be replaced by equivalent auditory explanations with learning enhanced due to an increase of effective working memory capacity (instructional modality effect). These beneficial effects of using audio/visual presentations only occur under conditions where the two or more components of a purely visual presentation are unintelligible in isolation and must be mentally integrated before they can be understood. The following two sections describe situations when dual-mode instructional formats might not be beneficial for learning.

#### **Case 1: When equivalent auditory and visual explanations are presented concurrently**

In practice, some multimedia instructional materials use auditory explanations concurrently with the same visually presented text. From a cognitive load perspective, such concurrent duplication of information using different modes of presentation increases the risk of overloading some of the sensory channels and might have a negative learning effect. Elimination of a redundant source of information might be beneficial for learning in this situation. This effect was observed in a study designed to compare three computer-based instructional formats on the theory of soldering (using a fusion diagram). The three formats were: Diagram with

Visual text, Diagram with Audio text, and Diagram with Visual text plus Audio text (Kalyuga, Chandler & Sweller, 1999). At the time the study was conducted, participants (novice apprentices) had not yet acquired any substantial experience of soldering. Figure 1 represents a section of the Diagram with Visual text format (in the other two formats, visual textual explanations were replaced or supplemented by equivalent auditory explanations).

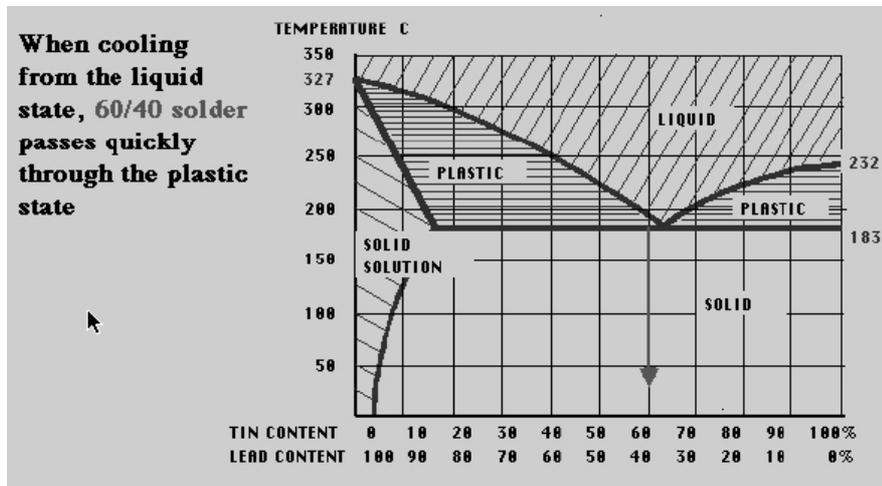


Figure 1: A section of the Diagram with Visual text instructional format for the Fusion diagram.

Means for subjective ratings of instructional difficulty (considered to be a measure of cognitive load) and test performance scores on multiple choice tasks are displayed in Figure 2. The results of the study indicated that the Diagram with Audio text group demonstrated a lower subjective rating of cognitive load and higher test performance than both the Diagram with Visual text group and the Diagram with Visual text plus Audio text group.

The instructional modality effect was replicated in this study (the Diagram with Audio text group outperformed the Diagram with Visual text group). In addition, the Diagram with Audio text group outperformed the Diagram with Visual text plus Audio text group. The inclusion of redundant, visually presented text simultaneously with an identical auditory presentation, which is common with many standard multimedia

packages, imposed an additional unnecessary cognitive load which interfered with learning (an example of a redundancy effect).

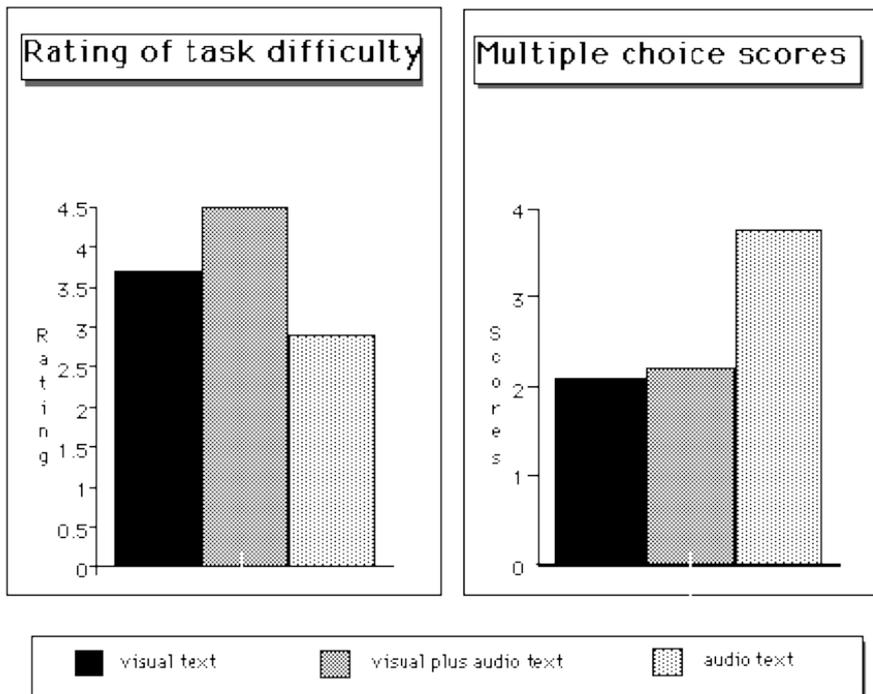


Figure 2: Charts of means for the data of experiment with the Fusion diagram instructions.

Thus, from the point of view of cognitive load theory, concurrent duplication of the same information using different modes of presentation increases the risk of overloading working memory capacity and might have a negative effect on learning. Relating corresponding elements of visual and auditory content in working memory consumes additional cognitive resources. In this case, elimination of a redundant visual source of information was beneficial.

Audio and visual explanations in the above mentioned study were presented to learners simultaneously. The negative effect on learning might not be the case when the same information is presented in different modes but not simultaneously (e.g., one mode after another, with some delay). In this case, cognitive resources might not be diverted to

establishing relations between corresponding visual and auditory elements occupying working memory at the same time. If, for example, the visual text is presented after the auditory text has been fully articulated, although either the auditory or visual text is still redundant, visual and auditory explanations must not be mentally integrated in working memory at the same time. Working memory capacity is not wasted on establishing connections between corresponding elements of visual and auditory components and precise coordination the two sensory modes. Working memory resources, otherwise used for such coordination, will be available for learning.

Thus, a non-concurrent duplication of information using different modes of presentation might not increase the risk of overloading working memory capacity and should not have negative learning consequences. If complete elimination of a redundant visual source of information is not possible or desirable for some reasons, a delayed non-concurrent presentation of this source might be beneficial for learning. It is useful to make a distinction between redundancy and revision of previously learned material. Revision is not a "redundant" activity that will interfere with learning because revision will not increase working memory load. Redundancy occurs when learners must unnecessarily translate and coordinate multiple sources of information processed simultaneously. That activity is mentally demanding and for learners who can fully understand one source of information, concurrently presenting them with other sources generates an extraneous cognitive load. Delayed presentation of a redundant source of information may effectively transform it into a form of revision that does not incur additional working memory load.

### **Case 2: When an instructional format is not matched to learner experience**

In the previous case, the diagram was not intelligible in isolation for novice learners and required additional textual explanations. To reduce cognitive load, the additional information had to be presented in auditory form concurrently with the diagram. But if an isolated diagram is sufficiently intelligible to a learner (for example, because of extensive experience in a domain), how would additional explanations (in auditory or written form) influence learning?

Studies with single modality visual instructions in electrical engineering (Kalyuga, Chandler & Sweller, 1998) indicated that low-knowledge trainees benefited from additional text based information included with diagrams of electrical circuits. High knowledge electrical trainees showed

a preference for an instructional package which consisted of the electrical circuit diagram only. Eliminating redundant text was the best way to reduce cognitive load in this situation.

Similarly, the auditory explanations may also be redundant when presented to more experienced learners. If an instructional presentation forces learners to unnecessarily attend to the auditory explanations continuously without the possibility of skipping or ignoring them, learning might be inhibited because of cognitive overload. To confirm these assertions, alterations in relative performance between different instructional conditions were observed as learners' level of experience increased (Kalyuga, Chandler, & Sweller, 2000).

Experimental materials were instructions in using cutting speed nomograms. Such nomograms indicate a proper number of revolutions per minute for drilling or turning operations and are used to set up drilling machines or lathes. The learners were given practice over a sufficient period of time to allow a substantial development of experience in this specific area. Computer based intensive training sessions were designed to practice learner skills in the domain.

Different versions of cutting speed nomograms were used at different stages of the experiment. The Diagram with Audio text format used at the first stage (before training sessions began) is represented in Figure 3. Only the headings of the sequential steps (e.g. Step 1. Select the cutting speed; Step 2. Select the diagonal line) were displayed in shaded rectangular areas to be clicked on by the learners. When a learner clicked on a step area, corresponding auditory commentaries were delivered to the learner via headphones (for example, for Step 1, "From the table, select the cutting speed range for a given material, in this case, bronze"; for Step 2, "At the right upper corner of the diagram, select the diagonal line that corresponds to the lowest available cutting speed within the suggested range for bronze", etc.). The auditory information was coordinated with screen based animations and highlights of the appropriate elements of the nomogram. The Diagram only format contained the nomogram without the step headings, textual explanations and statements. No highlights of elements of the nomogram or animations were used in this format.

The results demonstrated that after the learners became more experienced in the domain (Stage 2) due to intensive training sessions, the initial

relative advantage of the audio text at Stage 1 disappeared while the effectiveness of the diagram alone condition increased. There were no significant differences between the formats at Stage 2. Interaction effects indicated that the highest rate of learning was for a diagram only format.

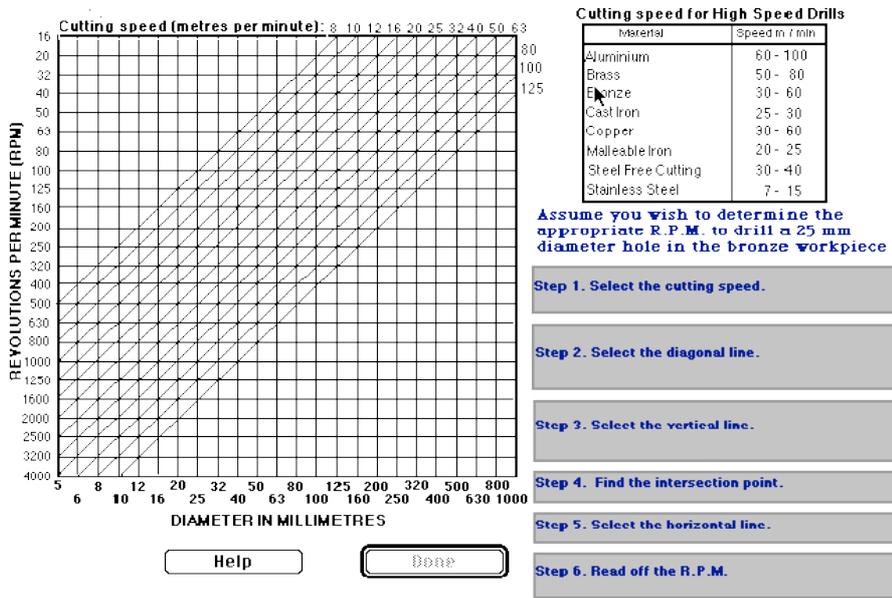


Figure 3: A section of the Diagram with Audio text instructional format for the Cutting speed nomogram.

After additional intensive training and under strictly controlled learning conditions (auditory explanations started immediately after displaying the instruction and consecutive steps followed each other without interruptions; both formats were displayed for the same 45 seconds that were necessary to articulate aloud all the textual explanations in the audio text format), substantial differences between the conditions were eventually obtained (Stage 3), providing evidence of a redundancy effect. With experienced learners, the inclusion of audio text that was difficult to ignore interfered with learning. Students found the diagram alone materials easier to process and performed at a higher level on the subsequent test. Subjective rating measures confirmed that the cognitive load profile of these two conditions was the reverse of that obtained at the first stage.

The cumulative nature of the results is illustrated in Figure 4. The diagrams on the left side of the figure indicate that performance on the multiple-choice test by the novices was very poor when presented with Diagram-only instructions compared to Diagram with Audio text instructions.

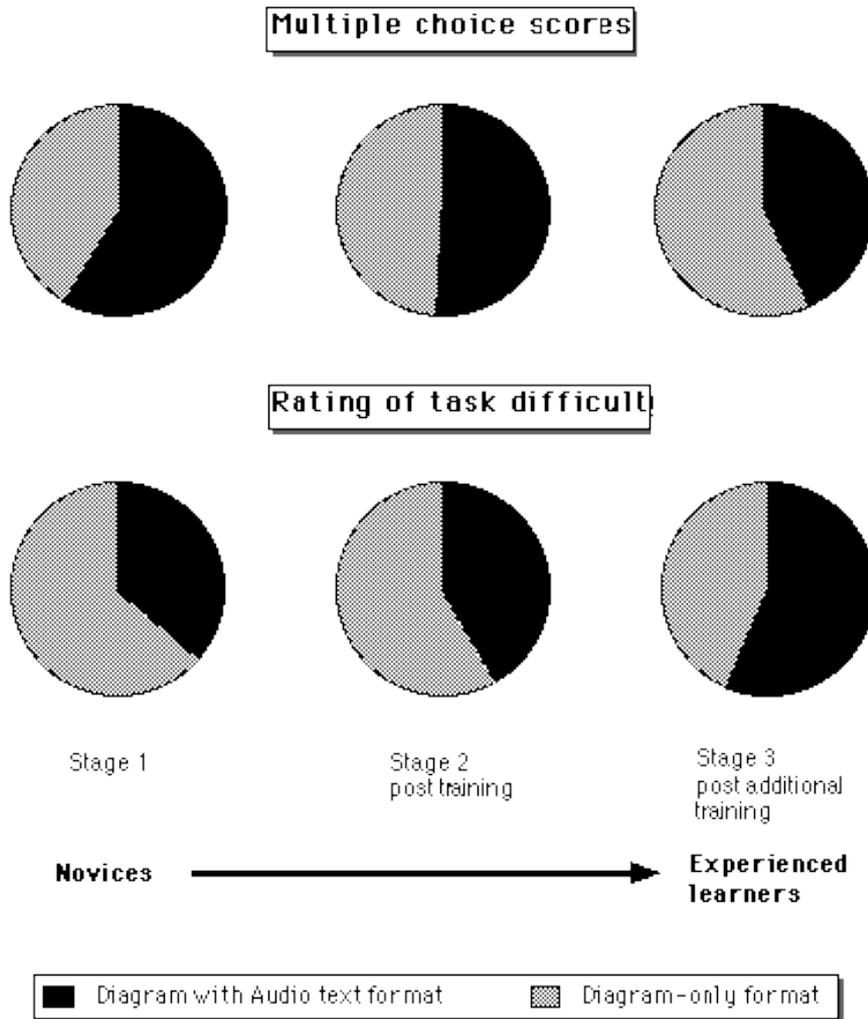


Figure 4: Comparative relations between means on the Diagram with Audio text and Diagram-only formats with increasing experience.

Furthermore, as can be seen from the subjective rating scale scores, these learners reported that the diagram-only instructions were more difficult to understand than the Diagram with Audio text instructions. As these learners became more experienced through Stage 2 and on to the substantial practice obtained by the same students prior to the tests of Stage 3, the relative effectiveness of the Diagram-only and Diagram with Audio text conditions reversed with the Diagram-only condition proving more effective and, based on subjective ratings, imposing a reduced cognitive load.

Thus, different instructional formats resulted in differential learning rates depending on the learners' experience. This is an important factor determining the effectiveness of dual-modality presentations which is not as beneficial for more experienced learners.

## **Conclusion**

Human cognitive capacity is limited: we can process only a very limited amount of information at any one time. Instructional presentations may be ineffective if they ignore limitations of the human information-processing system and force learners to process several interdependent sources of information simultaneously causing a heavy working memory load. Cognitive load considerations can provide designers with guidance in efficient structuring of instructional presentations involving more than one modality. Of course, effectiveness of multimedia instruction depends on many factors in addition to those affecting working memory load. Nevertheless, failure to take into account working memory considerations might override any positive attainments of implementing various (and frequently costly) technological innovations. Limited working memory may be effectively expanded by using more than one sensory modality, and instructional formats in which separate sources of information are presented in alternate, auditory or visual, forms might be more efficient than equivalent single modality formats. Such dual modality presentation techniques are frequently used in traditional instructional practice. For example, students may prefer listening to oral explanations of new, complex, diagram based materials (e.g. when studying geometry or engineering) rather than reading such explanations in textbooks.

In practice, however, auditory explanations are often used simultaneously with the same visually presented text. Such concurrent duplication of the same information using different modes of presentation increases the risk of overloading working memory capacity and might have a negative effect

on learning. Unnecessarily relating corresponding elements of visual and auditory content of working memory consumes additional cognitive resources. In such a situation, elimination of a redundant visual source of information might be beneficial for learning. Moreover, the auditory explanations may also become redundant when presented to more experienced learners. If an instructional presentation forces these learners to attend to the auditory explanations continuously without the possibility of skipping or ignoring them, learning might be inhibited.

The redundancy that might overload working memory generally occurs under conditions where different sources of concurrently presented information are intelligible in isolation and where each source provides similar information but in a different form. Attending to unnecessary information requires cognitive resources that consequently are unavailable for learning. If, for example, a diagram is sufficiently self-contained and intelligible in isolation, then any accompanying text (in written or auditory form) explaining the diagram which provides no additional information may be redundant and should be omitted. Redundancy occurs when learners must unnecessarily translate and coordinate multiple sources of information presented simultaneously (such as a diagram and text that redescribes the information in the diagram). That activity is mentally demanding and for learners who can fully understand one source of information, concurrently presenting them with other sources generates an extraneous working memory load.

Thus, audiovisual instructional presentations might not be efficient if they do not eliminate any avoidable load on working memory. Generally, when dealing with diagrams and text: (a) Units of textual explanations should be presented in auditory rather than written form; (b) The same units of textual explanations should not be presented concurrently in both auditory and written form (if both auditory and written text are required, written materials should be delayed and presented after auditory explanations were fully articulated); (c) When presented in auditory form, textual explanations should be easily turned off or otherwise ignored by more experienced learners.

## References

- Baddeley, A. (1992). Working memory. *Science*, 255, 556-559.
- Hegarty, M., Quilici, J., Narayanan, N.H., Holmquist, S., & Moreno, R. (1999). Multimedia instruction: Lessons from evaluation of a theory-based design. *Journal of Educational Multimedia and Hypermedia*, 8, 119-150.

- Kalyuga, S., Chandler, P., & Sweller, J. (1998). Levels of expertise and instructional design. *Human Factors*, 40, 1-17.
- Kalyuga, S., Chandler, P., & Sweller, J. (1999). Managing split-attention and redundancy in multimedia instruction. *Applied Cognitive Psychology*, 13, 351-371.
- Kalyuga, S., Chandler, P., & Sweller, J. (2000). Incorporating learner experience into the design of multimedia instruction. *Journal of Educational Psychology*, 92, 126-136.
- Mayer, R.E. (1997). Multimedia learning: Are we asking the right questions? *Educational Psychologist*, 32, 1-19.
- Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 63, 81-97.
- Mousavi, S., Low, R., & Sweller, J. (1995). Reducing cognitive load by mixing auditory and visual presentation modes. *Journal of Educational Psychology*, 87, 319-334.
- Najjar, L. (1996). Multimedia information and learning. *Journal of Educational Multimedia and Hypermedia*, 5, 129-150.
- Paivio, A. (1990). *Mental representations: A dual-coding approach*. New York: Oxford University Press.
- Penney, C.G. (1989). Modality effects and the structure of short term verbal memory. *Memory and Cognition*, 17, 398-422.
- Sweller, J. (1999). *Instructional Design*. Melbourne: ACER.
- Tergan, S. (1997). Misleading theoretical assumptions in hypertext/hypermedia research. *Journal of Educational Multimedia and Hypermedia*, 6, 257-283.
- Tindall-Ford, S., Chandler, P., & Sweller, J. (1997). When two sensory modes are better than one. *Journal of Experimental Psychology: Applied*, 3(4), 257-287.

<p>Slava Kalyuga School of Education, The University of New South Wales, NSW 2052, Australia. S.Kalyuga@unsw.edu.au</p>
---