Multimedia: How to Combine Language and Visuals - By: WOLFGANG SCHNOTZ & HOLGER HORZ
Posted by Michael Lambarena on February 28 2010 14:19:41

Multimedia: How to Combine Language and Visuals

1. Challenges of Using Multimedia

In the last decade, advanced computer technology has allowed for development of information systems and learning environments that combine language with other forms of human communication in innovative ways. Language in the form of written texts, for example, can be combined not only with static pictures or graphs as in printed material, but also with animation or video. Furthermore, language can not only be used as written text, but can also be presented in its 'natural' form of auditory, spoken text. The different ways of displaying information can also be combined with further auditory input such as sound or music. Furthermore, computer-based hypermedia also allows quick access to, and flexible combination of, various sources of information.

This general development is usually associated with the use of multimedia. The term 'multimedia' is relatively unclear because it "…conjures up a variety of meanings" (Mayer, 2005, p. 5). It sometimes refers to the technological basis such as computers, networks or devices for information display. In other cases it addresses the representational format such as text, graphics or animation. Finally, it sometimes refers to the sensory modality in perceiving the presented information such as the eye or the ear. Multimedia has become a hot topic during the recent years. When computer scientists focus primarily on the technological aspects of multimedia, psychologists and educational specialists are more concerned with the representational issues and how individuals grasp the information that is represented through the use of different sensory channels and further employ this information for generating new knowledge in their minds. In fact, these topics are per se not related to new technologies because the research was started on more traditional audiovisual media, like the educational films (e.g. Heinrich, 1961). Instead, they are relevant in more traditional scenarios.

Multimedia learning and information systems generally aim at helping individuals to construct knowledge structures about a subject matter or to update their knowledge about a specific topic

(Mayer, 2005; Schnotz, 2005). Frequently, these systems also allow multiple ways of self-regulated learning or information research. It is widely assumed that self-regulated learning should lead to more elaborated and better applicable knowledge systems. When we will deal with multimedia information or learning systems in the following, we will first focus on the functions of *instructional* pictures (static and dynamic) and *instructional* text (visual or auditory). We are not concerned with decorative pictures, because these pictures are usually considered as "seductive details" that interfere with the information processing (Harp & Mayer, 1998; Clark, 2005). We will then focus on the theoretical background of multimedia learning and epitomise three theories that attempt to explain how multimedia information processing occurs from a psychological point of view. The theoretical approaches propose solutions to some crucial problems concerning multimedia design. We will report the essentials of designing texts and pictures in multimedia information or learning environments and explain the rationale behind these essentials. Finally we will discuss instructional limitations of multimedia and suggest some criteria for deciding whether the information should be presented via multimedia or via a single medium.

2. A Picture can be Worth More than 10,000 Words - Do We Still Need Texts?

A common expression claims that a picture is worth 10,000 words. Pictures have indeed various advantages as tools of communication. As pictures are not dependant on a specific language, they are less culturally specific than texts. The use of pictograms in international airports demonstrates that individuals from different backgrounds have a better understanding of pictorial displays than text. Pictures also allow reading of large amounts of information in situations when drawing inferences from texts is difficult (Kosslyn, 1994). For example it is common for user manuals of complex technical products (Camcorder, DVD etc.) to use text combined with pictures to explain the different functions of the device. If the user manuals would only use text some users would have difficulty understanding the different functions of a device. Furthermore, pictures are informationally more complete than texts: a triangle, for instance, must include all of the necessary features of a triangle, which accordingly can be read off from the figure. Text, on the contrary, can be highly selective and describe only a few features of the same figure leaving out the rest.

Pictures differ from texts in various ways. First, pictures are internally consistent. If a relation between two elements in a picture is modified (for example, by moving one element to the left), all relations to the other elements are automatically modified accordingly. Therefore, pictures cannot contain contradictions. (The famous pictures of 'impossible objects' drawn by Maurits C. Escher seem to contradict this claim, but a closer look at it reveals that these so-called pictures are in fact combinations of different elements drawn from different mutually exclusive perspectives.) A text, on the contrary, can include contradictory propositions. If it does not, that is, if it is consistent, any change in the relation between two elements requires a laborious updating of numerous other relations in order to maintain consistency.

In addition, pictures differ from texts with regard to their information access structure. In a picture, data is arranged in a two-dimensional space. Accordingly, different elements can be relatively easily related to each other and an observer can use his/her visual schemata to extract topological and spatial information with low cognitive load. In a text, on the contrary, the data is presented in a linear order. Interrelating elements from different parts of the linear structure usually requires higher effort. Figure 1 presents an example that demonstrates the relative ease of processing topological and spatial information from a picture in comparison to a verbal description in a written text.

"Consider a problem given in the following natural language statements:
We have three pulleys, two weights, and some ropes arranged as follows:

1. The first weight is suspended from the left end of a rope over Pulley A. The right end of this rope is attached to, and partially supports, the second weight.
2. Pulley A is suspended from the left end of a rope that runs over Pulley B, and under Pulley C. Pulley B is suspended from the ceiling. The right end of the rope that runs under pulley C is attached to the ceiling.
3. Pulley C is attached to the second weight, supporting it jointly with the right end of the first rope.

The pulleys and ropes are weightless; the pulleys are frictionless; and the rope segments are all vertical, except where they run over or under the pulley wheels. Find the ratio of the second to the first weight, if the system is in equilibrium."

(Larkin & Simon, 1987, p. 72)

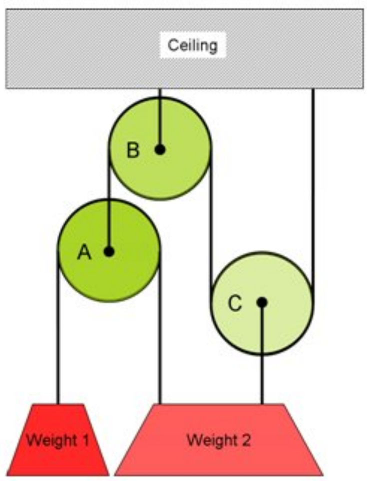

Figure 1: Following Larkin and Simon every participant of the study got the textual problem (left side above) and drew a sketch of situation somewhat like Figure 1. This demonstrates that a pictorial representation is easier to understand than a textual representation of the given problem because the construction of a mental model requires lower mental effort if a pictorial model is given. (Larkin & Simon, 1987, p. 72-73)

Pictures can facilitate information processing if they are adequately designed according to the target of communication. More specifically, the following functions of instructional pictures that can be distinguished are illustrated here by a few examples:

- Pictures allow for information that belongs together to be organised contiguously. This reduces the searching efforts when thematically related information has to be interconnected (see Figure 2a for an example).
- Pictures also allow information to be organised depending on its importance. For instance, it is possible to express the centrality of an element with regard to all other elements (see Figure 2b for an example).
- Pictures allow humans to easily perceive spatial information because we possess automated perception routines that can be applied without the cognitive load (see Figure 2c for an example).

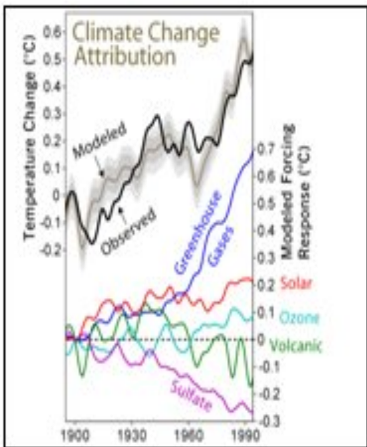

Figure 2 a: A global climate model to reconstruct the historical temperature record and the degree to which the associated temperature changes can be decomposed into various forcing factors. (Global Warming Art, 2007a).
The diagram is an example for instructional pictures contiguously grouping information which belongs together.

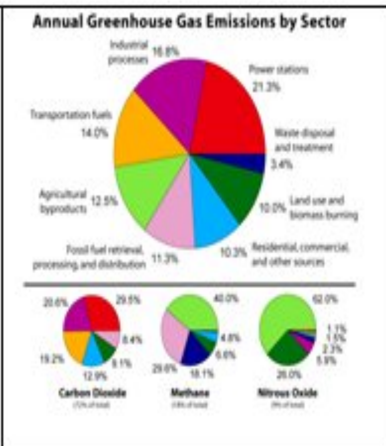

Figure 2b: The relative fraction of man-made greenhouse gases coming from each of eight categories of sources. The top panel shows the sum over all greenhouse gases, weighted by their global warming potential over the next 100 years. Lower panels show the comparable information for each of these three primary greenhouse gases, with the same coloring of sectors as used in the top chart. (Global Warming Art, 2007b).
The diagram is an example for instructional pictures grouping information depending on its importance. Here the central information is presented in the top panel of the diagram; the additional details are grouped in smaller size in the lower panel.

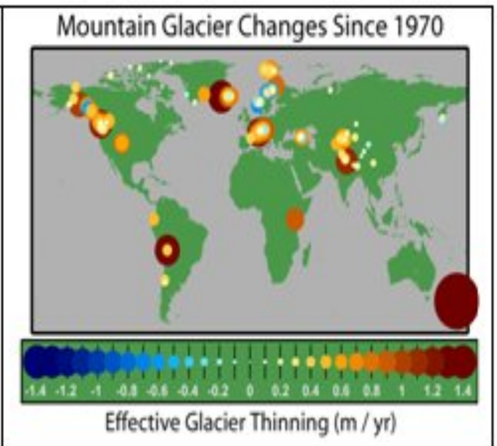

Figure 2c: The effective rate of change in glacier thickness, also known as the glaciological mass balance. The map shows the average annual rate of thinning since 1970 for 173 glaciers. Larger changes are plotted as larger circles and vice versa (Global Warming Art, 2007c).
The diagram is an example for instructional pictures facilitating information perception and understanding because of its support of automatically used perception inferences.

Picture and text are useful for different purposes in specific instructional situations whereby the advantages of pictures correspond to the disadvantages of text, and vice versa. Despite their numerous advantages, pictures cannot generally compete with text because text is a more efficient medium than pictures, in specific instructional constellations as described in the following: In a way, text is, in terms of representation, a much more powerful instrument than pictures for representing and communicating knowledge. There is no problem in formulating general negations or disjunctions by descriptions as, for example, *"Pets are not allowed"* or *"High blood pressure can be caused by nicotine or a lack of movement"*. On the contrary, pictures can show only specific negations, and they can illustrate disjunctions only through a series of pictures. In most advanced cultures, texts play a central role for the distribution of information. Perception and comprehension of pictures often need verbal guidance from a text as well, especially if individuals have only low prior knowledge on the subject matter (Bernard, 1990). Text authors can guide the processing of pictorial information more precisely than via visual cues in a picture, because individuals have less freedom while processing language than in processing pictures. A further advantage of text is that it can be processed through different sensory modalities: It can be presented visually (on paper or on a screen), auditory (via loudspeakers or headphones), or even in the tactile mode (in the form of Braille-writing). Although there is an option of displaying the silhouette and internal structure of an object for tactile perception, the possibilities are much more limited.

To summarize: There is no “general rule” that text or pictures are useful for all instructional aims of a particular kind. Accordingly, visual communication through pictures will never be able to fully replace verbal communication through text and vice versa.

3. Multimedia Learning: Knowledge Acquisition from Texts and Pictures

Numerous studies have found that students learn better from text combined with pictures than from text alone. Whereas early research has focused on the mnemonic function of pictures combined with narrative text (Levie & Lentz, 1982; Levin, Anglin, & Carney, 1987), recent studies deal also with the explanatory function of pictures. Mayer and his colleagues, for example, have found that students understand technical devices or natural phenomena better when they learn from text and pictures combined. This so-called ‘multimedia effect’ seems to be especially strong when learners have low domain-specific knowledge (Mayer, 1997, 2001; Mayer & Moreno, 2002). In other cases, however, adding pictures to a text can also have detrimental effects on learning (Sweller, van Merriënboer & Paas, 1998). So, the question arises whether multimedia learning is more successful than learning from a single medium, and why it is more successful.

Dual Coding. The advantage of combining texts with pictures was often explained through Paivio’s (1986) Dual Coding Theory (DCT). According to this theory, the human cognitive system entails two separate, but related subsystems: a verbal and an imagery system. Both systems can interact and can also be activated independently, both have a limited capacity and allow a dual coding of information. Normally verbal information is processed only in the verbal subsystem, but pictorial information is processed in both systems, in the pictorial and in the verbal subsystem of the human cognitive system (for a critique see Engelkamp, 1990). As the two subsystems are interconnected, it is possible to establish cross-referential connections between the representation of a subject matter in the verbal system and a representation of the same subject matter in the imagery system. In this way, the overall memory representation is more elaborated, which is assumed to be the basis of the positive text-pictures effects mentioned above.

Cognitive Theory of Multimedia Learning. Based on the dual coding theory of Paivio (1986), Mayer (1997, 2001) has developed a Cognitive Theory of Multimedia Learning (CTML). The theory assumes that the human cognitive system includes a verbal and pictorial (image) subsystems. Accordingly, individuals can use different representational formats to internally encode and store knowledge. Furthermore, the theory draws on ideas of Baddeley (1986) about human working memory. According to Baddeley, humans' working memory is characterized by a rather limited capacity and high decay rates. It includes a central executive and two slave systems: the phonological loop and the visual sketchpad. Similarly, Mayer assumes two sensory subsystems in working memory: a phonological system and a visual system. His first basic assumption on multimedia learning merges the notion of dual coding and the notion of two sensory subsystems. Humans are supposed to process information in working memory through two channels: a phonological-verbal channel and a visual-pictorial channel. The second basic assumption is that these two channels have a limited capacity to convey and process information. The third basic assumption is that humans are active sense-makers: They engage in active cognitive processing to construct coherent knowledge structures from both the available external information and their prior knowledge.

According to Mayer's theory, verbal selection processes synthesise a so-called propositional text base and verbal organization processes lead to a text-based mental model. Similarly to the verbal processes pictorial selection processes synthesise a picture base and pictorial organization processes lead to a picture-based mental model. Integration processes constitute referential relations between the text-based and picture-based mental models (see Figure 3a).

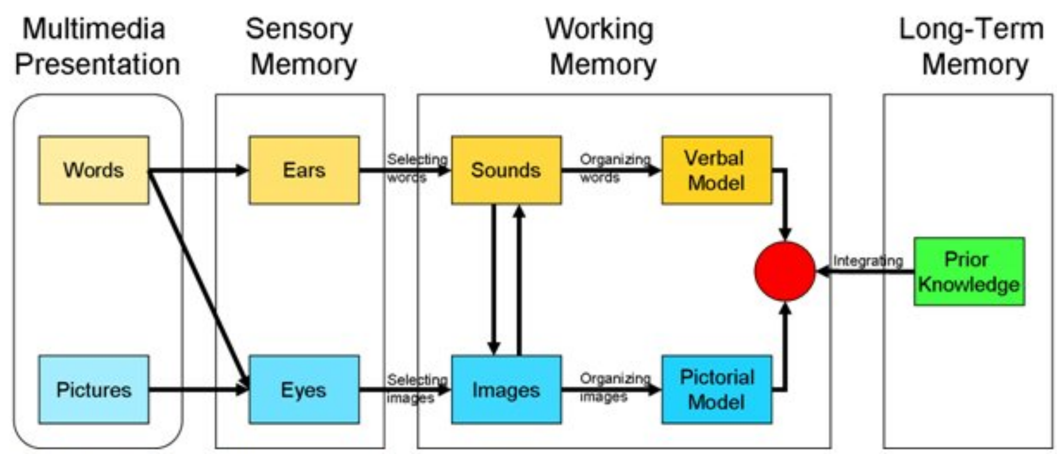

Figure 3a: Cognitive Theory of Multimedia learning (Mayer, 2005, S. 37)

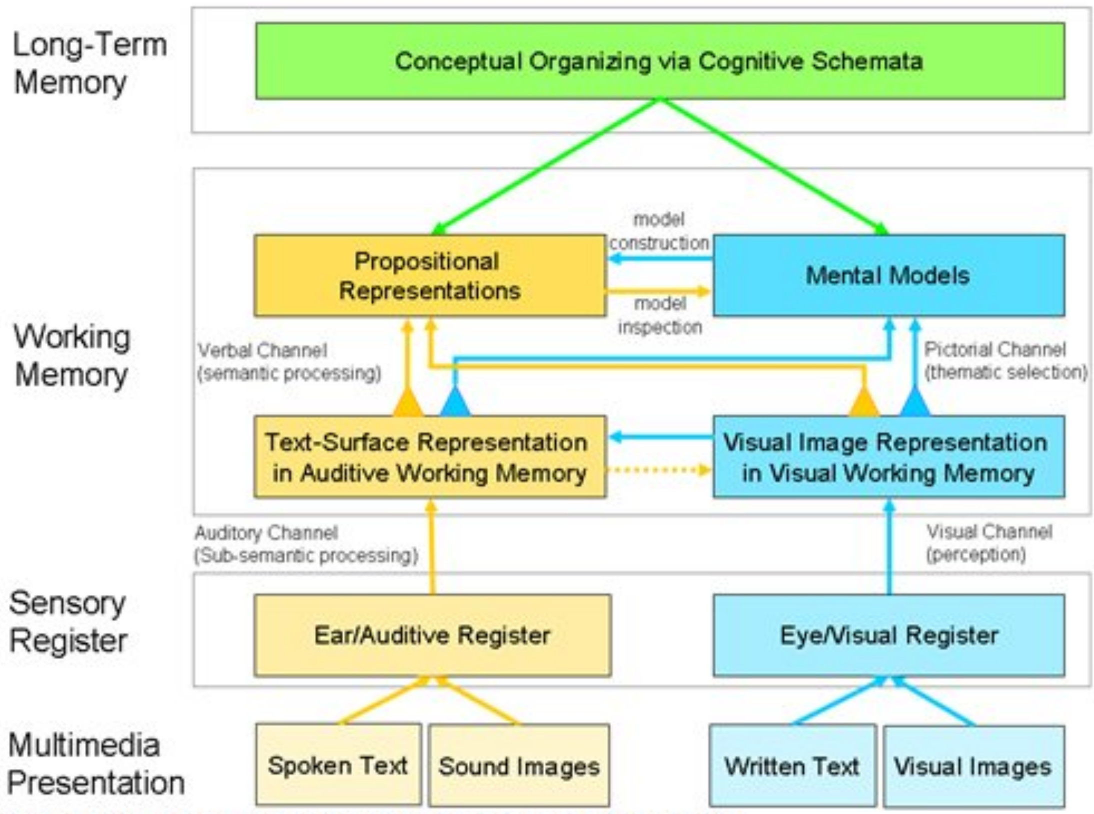

Figure 3b: Integrative model of text and picture comprehension (Schnotz, 2005)

Integrated Model of Text and Picture Comprehension. The parallelism of the text and image processing assumed in Mayer's model seems to be questionable insofar as texts and pictures use completely different principles of representation. Schnotz (2005) has therefore developed (based on previous work of Schnotz and Bannert, 2003) an integrative model of the text and picture comprehension (IMTPC). The model, which is schematically shown in Figure 3b, refers to the single or combined comprehension of written and spoken text, visual pictures and auditory pictures (i.e. sound images). It is based on the following assumptions: Text and picture comprehension take place in cognitive architecture including a working memory of limited capacity, modality specific sensory registers as information input systems and long-term memory. Verbal information (i.e. information from written texts or spoken texts) and pictorial information (i.e. information from visual pictures and from auditory pictures, or from sounds, respectively) is transmitted to working memory through the visual and auditory channels. The channels have limited capacity to process and store information. Further information processing in working memory takes place in two different representational channels: the verbal channel and the pictorial channel. Information from written or spoken text is processed in the verbal channel. Information from visual pictures or from sounds is processed in the pictorial channel. These channels also have limited capacity to process and store information. Text and picture comprehension are active processes of coherence formation. In comprehension, individuals engage in building coherent knowledge structures from the available external verbal and pictorial information and from their prior knowledge.

According to these models, the main reason for the learning enhancing function of multimedia compared to single media is that multimedia can support integrative processing of verbal and pictorial information in working memory. Integrative processing requires both verbal and pictorial information to be held simultaneously in working memory. Because the capacity of working memory is limited and decay rates are high, students learn better from words and pictures than from words alone under the condition that verbal and the pictorial information are simultaneously available in working memory. From this fundamental condition of integrative processing, various principles for the multimedia design can be derived.

4. Multimedia Design Principles

Most principles of multimedia design deal with a basic issue of comprehending a subject matter based on multiple representations: Verbal and pictorial information that should be mentally integrated by cross-referential connections has to be simultaneously available in the learner's working memory in order to allow these connections to occur. Availability of information can be reduced by split of attention, and it can be enhanced by contiguity using different sensory modalities in multimedia learning.

Minimise Split of Attention. If pictures are presented with written text, then a split of visual attention is required, because the eye has to switch repeatedly between words and pictures. Split of visual attention has negative effects on learning in different ways (Sweller, 1994; Sweller & Chandler, 1994; Kalyuga, Chandler, & Sweller, 1999). On the one hand, if time for learning is limited and if a complex picture has to be processed (which is often the case with learning from animation), then the visual channel can process only portions of the available information because of the channels limited transmission capacity. That is, either parts of the picture or parts of the next one have to be ignored. The effects of split of attention are especially strong if text is combined with animation due to the fluent nature of the visual display. The split of attention situation limits the quantity of the perceptible information. On the other hand, split of visual attention reduces the simultaneous availability of verbal and pictorial information in working memory. If the learner switches from the text to the picture or vice versa, the switch requires search processes that in turn cost time. During the time of searching for new information, old information decays from working memory. If the learner switches from text to picture, verbal information is lost. If he/she switches from picture to text, pictorial information is lost. In any case, the simultaneous availability of verbal and pictorial information in working memory is reduced by a split of visual attention, because it prevents full temporal contiguity.

Split of visual attention can be reduced if text and pictures are presented close to each other since this minimises search processes. This effect is known as the *spatial-contiguity effect* in multimedia learning. Moreno and Mayer (1999) conducted a study where the physical proximity of the on-screen text and the animation was manipulated. The overall result was that learners demonstrate better learning outcomes when on-screen text and visual materials are physically integrated rather than separated.

Should Text be Written or Spoken in Multimedia Learning? The best way to avoid split of attention in multimedia learning seems to be the use of spoken instead of written text. In this case, the full capacity of the auditory sensory channel can be devoted to the text, whereas the full capacity of the visual channel can used for the picture. Under this condition, a maximum of verbal and pictorial information can be simultaneously held in working memory. This effect is known as the *modality effect* in multimedia learning. However, the modality effect is not only due to avoidance of split of attention. Mayer and Moreno (1998) found that even when text and animations were presented sequentially, so that no split-attention was required, students learned better when the verbal information was presented as narrated text rather than visual on-screen text. The authors concluded that the positive effects of narrations combined with animations are not only due to an avoidance of split attention. They assume that using both the auditory and the visual channel in multimedia learning leads to a higher working memory's capacity involved in cognitive processing than a single channel only.

One could speculate that students learn generally better when verbal information is presented as spoken text simultaneously with pictures rather than as printed text, even under the condition of maximal spatial contiguity, because presenting pictures only with spoken text avoids any split of visual attention. From a practical point of view, one could recommend that whenever pictures are combined with text, the text should be presented as spoken rather than as written text. However, this seems problematic for various reasons. First, the studies on the modality effect presented by Mayer (2001) have all employed animated pictures, that is, under conditions when reading on-screen text is necessarily associated with a loss of fleeting pictorial information. The results might be less clear-cut if static pictures had also been used. Second, an important aspect is how much control of processing is given to the learner. Written text provides better control because stable graphemic information allows re-reading in case of comprehension difficulties whereas spoken text is fleeting. A recommendation such as 'if there is picture, present text in the auditory mode' is therefore not justified in all cases. Text should be presented as narrated text generally, if an animation is shown, but not necessarily in case of static picture. In addition the validity of the recommendation depends of the complexity and recentness of pictures' information. Additionally, reading a text offers the advantage of a higher control of processing the text. This is especially important in the case of difficult texts.

The temporal arrangement of text and pictures is of course also relevant for spoken texts, if verbal and pictorial information should be held simultaneously in working memory. Mayer and Moreno (2003) found that learners achieve better learning outcomes when verbal and visual materials are temporally synchronized rather than separated in time. This effect is known as the *temporal-contiguity effect* in multimedia learning. In a way, the previously mentioned *spatial-contiguity effect* is only a variant of the *temporal-contiguity effect*, because the lower the proximity of text and pictures, the longer the search processes associated to the switches from one source of information to the other will take, and the lower temporal contiguity of verbal and pictorial information in working memory will be. In other words: Spatial contiguity enhances temporal contiguity on the cognitive level.

Multimedia designers often tend to prefer multiple equivalent sources of information in order to allow learners to choose the source they prefer most. For example, a picture could be accompanied both by written text and auditory text. However, Mayer and Moreno (2003) also found that students learned better from "animation plus only narration" than from a combination of "animation plus narration plus simultaneously presented visual text". According to the authors' terminology, this effect is known as the *redundancy effect* in multimedia learning. Redundancy between two sources of text information combined with pictures lowers learning success.

Motivating Devices or Seductive Details? Can pictures and entertaining adjuncts in the form of sounds, or music enhance learning with multimedia? According to the theories of multimedia learning, learners process multimedia information in their visual and verbal channels. When additional verbal information is presented which does not compete against other auditory information, one could assume that a motivational effect of attractive adjuncts, like music, foster the information processing. Learning experiments showed, however, that students learned better when additional materials like music was excluded rather than included in multimedia learning environments (Moreno & Mayer, 2000). The results can be explained by the theories of Baddeley and Paivio (see above). The additional auditory information (music) interfered with the read text, because the text was encoded in the phonological loop in working memory where it interfered with the music. Therefore there was less capacity available for processing the read text and linking it with the pictorial information.

5. 'Instructional Overkill' by Multimedia?

Perhaps the most fundamental result of multimedia learning is that there are no appropriate rules of thumb concerning multimedia design. Whereas the combination of text and pictures as the core of multimedia is helpful for learning in many cases, multimedia is not the silver bullet for all problems in learning and instruction. If a text is combined with task-inappropriate visualisations, comprehension and learning can be inhibited by pictures. Furthermore, if pictures are provided, learners can use them instead of reading the text, which can also be detrimental for learning (Schnotz & Bannert, 1999). Sweller and Chandler (1991) have demonstrated that in some instances of teaching, one source of information can be sufficient enough for learners to acquire a full understanding of the learning content, whereas further sources of information simply require mental capacity without further benefit. This so-called *redundancy effect* (which is different from the Mayer's redundancy principle mentioned above) occurs when learners have higher prior knowledge and multimedia instruction would be an 'instructional overkill' for them.

The so-called *expertise reversal effect* can be considered as a special case of the redundancy effect: An instructional method which is efficient for novices, such as multimedia learning, can become inefficient if learners are more advance with regard to their expertise (Kalyuga, Ayres, Chandler, & Sweller, 2003). Sometimes, presenting only a text or only a picture (e.g. a technical diagram) leads to better learning results than multimedia instruction. Learning environments cannot be designed on the basis of mechanically applied principles. Too many factors such as prior knowledge, individual preferences or learning orientations as well as cognitive abilities interact and can have an influence on processes and outcomes of multimedia learning.

Multimedia leaning environments can facilitate and enhance learning processes. However, multimedia does not automatically result in positive effects on comprehension and learning. Moreover, in most cases the development of multimedia information costs much more material and human resources than delivering information through a single medium (Horz, Wessels & Fries, 2003). Therefore authors have to decide carefully how many resources should they invest into creating multimedia environments instead of using a single source of information in a more classical learning or information environment. The didactical aspect of this decision depends on the learning targets and the authors' knowledge about the target audience. Basically, multimedia design can only be successful if it is based on a sufficient understanding of what goes on in the mind of the learner.

Authors

Holger Horz, (horz@uni-landau.de)
University of Koblenz-Landau,
Faculty of Psychology, Landau, Germany.

Among his current research interests are:
- Instructional design of blended learning and multimedia learning
- Development of learning with text and pictures
- Learning from animation
- Support in complex learning environments

Wolfgang Schnotz, (schnotz@uni-landau.de),
Professor of General and Educational Psychology
University of Koblenz-Landau, Faculty of Psychology,
Landau, Germany.

Among his current research interests are:
- Cognitive foundations of learning from multimedia
- Visual thinking and learning
- Comprehension of text and graphics
- Learning from animation

The use of the figures in this article courtesy of the authors.

References

- Baddeley, A. D. (1986). *Working memory*. Oxford, England: Oxford University Press.

- Bernard, R. M. (1990). Using extended captions to improve learning from instructional illustrations. British *Journal of Educational Technology. 21*, 215–225.

- Clark, R. E.. (2005). Five common but questionable principles of multimedia. In R.E. Mayer (Ed.), *Cambridge Handbook of Multimedia Learning* (pp. 97-116). Cambridge: Cambridge University Press.

- Engelkamp, J. (1990). *Das menschliche Gedächtnis. Das Erinnern von Sprache, Bildern und Handlungen*. Göttingen: Hogrefe.

- Global Warming Art (2007-04-05, a). *Climate Change Attribution* [Internet; png-Image]. Available: http://www.globalwarmingart.com/wiki/Image:Greenhouse_Gas_by_Sector_png

- Global Warming Art (2007-04-05, b). *Greenhouse Gas by Sector* [Internet; png- Image]. Available: http://www.globalwarmingart.com/wiki/Image:Greenhouse_Gas_by

_Sector_png

- Global Warming Art (2007-04-05, c). *Glacier Mass Balance Map* [Internet; png- Image]. Available: http://www.globalwarmingart.com/wiki/Image:Glacier_Mass_Balance_Map.png

- Harp, S. F., & Mayer, R.E. (1998). How seductive details do their damage:A theory of cognitive interest in science learning. *Journal of Educational Psychology*, 90, 414-434.

- Heinrich, K. (1961). *Filmerleben, Filmwirkung, Filmerziehung*. Hannover: Schroedel.

- Horz, H., Wessels, A., & Fries, S. (2002). Gestaltung und zyklische Nutzung virtualisierter Präsenzlehre. In U. Rinn & J. Wedekind (Eds.), *Referenzmodelle netzbasierten Lehrens und Lernens* (pp. 71-99). Münster: Waxmann.

- Kalyuga, S., Ayres, P., Chandler, P. & Sweller, J. (2003). The Expertise Reversal Effect. Educational *Psychologist*, 38 (1), 23-31.

- Kalyuga, S., Chandler, P., & Sweller, J. (1999). Managing Split-attention and Redundancy in Multimedia Instruction. *Applied Cognitive Psychology*, 13, 351-371.

- *Kosslyn, S. M. (1994). Image and Brain*: The Resolution of the Imagery Debate. MIT Press, Cambridge, MA, 1994.

- Larkin, J. H. & Simon, H.A. (1987). Why a diagram is (sometimes) worth then ten thousand words. *Cognitive Science*, 11, 65-99.

- Levie, H.W. & Lentz, R. (1982). Effects of text illustration: A review of research. *Educational Communication and Technology Journal*, 30, 195-232.

- Levin, J.R., Anglin, G.J. & Carney, R.N. (1987). On empirically validating functions of pictures in prose. In D.M. Willows & H.A. Houghton (Eds.), *The psychology of illustration. Vol. 1: Basic research* (pp. 51-86). New York: Springer.

- Mayer, R. E. (1997). Multimedia learning: Are we asking the right questions? *Educational Psychologist*, 32, 1-19.

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

- Mayer, R. E.. (2005). Introduction to multimedia learning. In R.E. Mayer (Ed.), *Cambridge Handbook of Multimedia Learning* (pp. 1-16). Cambridge: Cambridge University Press.

- Mayer, R. E. & Moreno, R. (1998). A split-attention effect in multimedia learning: Evidence for dual processing systems in working memory. *Journal of Educational Psychology*, 90, 312-320.

- Mayer, R. E., & Moreno, R. (2002). Aids to computer-based multimedia learning. *Learning and Instruction*, Vol.12, 107-119.

- Mayer, R. & Moreno, R. (2003) Nine Ways to Reduce Cognitive Load in Multimedia Learning. *Educational Psychologist*, 38, 43-52.

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

- Moreno, R. & Mayer, R. E. (2000). A coherence effect in multimedia learning: The case for minimizing irrelevant sounds in the design of multimedia instructional messages. *Journal of Educational Psychology*, 97, 117-125.

- Paivio, A. (1986). *Mental representation: A dual coding approach*. Oxford, England: Oxford University Press.

- Schnotz, W. (2005). An Integrated Model of Text and Picture Comprehension. In R.E. Mayer (Ed.), *Cambridge Handbook of Multimedia Learning* (pp. 49-69). Cambridge: Cambridge University Press.

- Schnotz, W., & Bannert, M. (1999). Einflüsse der Visualisierungsform auf die Konstruktion mentaler Modelle beim Bild- und Textverstehen. *Zeitschrift für experimentelle Psychologie*, 46, 216-235.

- Schnotz, W., & Bannert, M. (2003). Construction and interference in learning from multiple representations. Learning and Instruction, 13, 141-156.

- Sweller, J. (1994). Cognitive load theory, learning difficulty and instructional design. *Learning and Instruction*, 4, 295-312.

- Sweller, J., & Chandler, P. (1991). Evidence for cognitive load theory. *Cognition and Instruction*, 8(4), 351-362.

- Sweller, J. and Chandler, P. (1994). Why some material is difficult to learn. *Cognition and Instruction*, 12, 185-233.

- Sweller, J., Merrienboer, J. G. v., & Paas, F. G. W. C. (1998). Cognitive architecture and instructional design. *Educational Psychological Review*, 10(3), 251-296.