UNIPED

UNIVERSITETSFORLAGET

PEER-REVIEWED

ÅRGANG 40, NR. 2-2017, S. 129–139
ISSN ONLINE: 1893-8981
DOI: 10.18261/ISSN.1893-8981-2017-02-03

Investigating Learners' Viewing Behaviour in Watching a Designed Instructional Video

Jonas R. Persson
Førsteamanuensis, Institutt for lærerutdanning, Norges teknisk-naturvitenskapelige universitet, NTNU
jonas.persson@ntnu.no

Eirik Wattengård
Multimedieteknolog, Avdeling for utdanningskvalitet, Norges teknisk-naturvitenskapelige universitet, NTNU
eirik.wattengard@ntnu.no

Elisabeth Egholm Jacobsen
Førsteamanuensis, Institutt for kjemi, Norges teknisk-naturvitenskapelige universitet, NTNU
elisabeth.e.jacobsen@ntnu.no

ABSTRACT

The production and use of video in education has increased during recent years. However, most videos are not properly designed to enhance attention and learning. In this article, we report on the design of an instructional video based on existing multimedia learning principles as well as on film and video theory, and the result of an eye tracker study of students watching it. The eye tracker technology enables us to study objectively the effects of the design elements. We show that the design principles used in the production help viewers to focus on important aspects in the video.

Keywords
video, cognitive science, eye tracker, viewing behaviour

INTRODUCTION

Instructional video in higher education is becoming increasingly popular and is commonly used for educational purposes today. For a video to be maximally effective as a learning object, it is important to design the video to optimise the learning process. To design a video, one must have some knowledge of both the viewing behaviour (for example where

one looks, i.e. where one's attention is) as well as knowledge of different aspects of how the human sensory system works. Mayer (2003) developed a Cognitive Theory of Multimedia Learning (CTML) model which can be used for guidance. By using the principles outlined in Mayer (2006) we designed an instructional video (NTNU Openvideo, 2015) on how to operate an analytic balance, aimed at first year chemistry students as a complementary instruction to written or oral instructions, for use before or during laboratory sessions. In this study, we are interested in how students (viewers) deploy their attention in the video. By using an eye tracking system, we were able to study where and for how long the viewer focused on different objects in the video.

The video was motivated by the problems of instructing all the students in a laboratory group (about 32 students) in the specially furnished measuring lab. As the measurements are intended to achieve high precision, the room is quite small, as only a few persons are supposed to operate the balances at one time. To give oral instructions is therefore a tedious and repetitive task for the instructors, and comes with the risk of miscommunicating or misunderstanding. In our project, students can watch the video before or during laboratory sessions without additional instructions, thus improving the level of instruction in both time and quality.

Studies with eye-tracker technology on students' attention while watching instructional videos are a relatively new field, which has to our knowledge not been done in chemistry. In this article, we describe the basis for our design of an instructional video in chemistry as well as directly observing the effects. The objective of this study is to investigate where the attention is in the video and the duration of attention for different objects, and to derive the effects of different designs in the video.

COGNITIVE THEORY OF MULTIMEDIA MODEL

Teaching in lectures and laboratories is mainly instructor-centred, with the instructor controlling both pace and content. This can cause some students to miss important aspects as the instruction goes too fast, or the student is not able to observe properly. A solution to this is to make use of instructional videos which might be more student-centred and self-paced.

Richard E. Mayer (2006) put forward the Cognitive Theory of Multimedia Learning (CTML) model, where he stated that multimedia instruction refers to presentations involving words and pictures that are intended to foster learning. Figure 1 presents a cognitive model of multimedia learning intended to mimic the human information-processing system. We have three memory stores, the sensory memory, working memory and long-term memory. The information in the multimedia presentation (as words and pictures) enters the sensory memory through eyes and ears, dividing the information into the visual (printed text and pictures) and auditory (spoken word and other sounds) channels. The visual information can be investigated with the use of eye tracker technology, as this enables us to observe where the eyes, that is the attention, are directed. The sensory memory can only hold the information for a very short period, so knowing where the viewers' attention is at any time while watching the video is paramount in order to measure the design effects used in the production. The main work of learning takes place in the working mem-

ory, which is used for holding and manipulating knowledge in active consciousness. The information is processed in the working memory and organised into the models constructed with the information in the working memory. The dual channels exchange information in the working memory, something that can be demonstrated by the word "dog"; when you read the word, a visual image of a dog can appear in your mind, in the same way that an image can invoke a sound image.

The long-term memory holds the learners' knowledge, but in order to actively use that information, it has to be brought into the working memory, where new information can be integrated with it.

The CTML model is based on three basic assumptions: dual channels (auditory and visual), limited-capacity and active processing.

We have two channels where information can be brought into the working memory, one based on visual stimuli and one on auditory stimuli. Note that words have a dual representation, both as a picture and as spoken words, something that is shown by the exchange between sounds and images in the working memory. As the working memory has a finite capacity, the amount of information that can be processed in each channel is limited. To make sense of the information in the working memory, the learner has to engage in active processing of the information. This consists of different processes such as selecting images and sounds, organising images and sounds and finally integrating the information to the knowledge in the long-term memory.

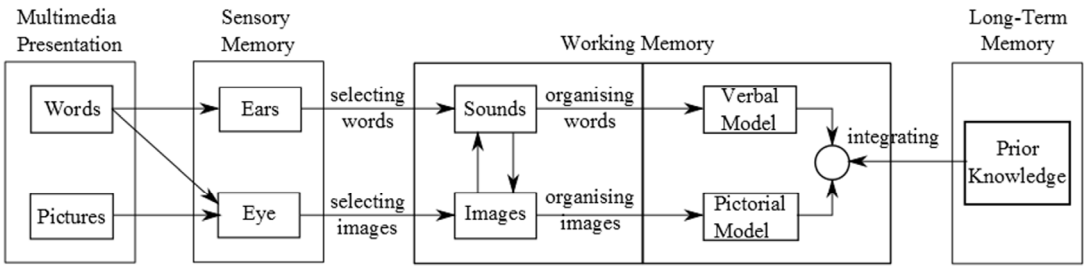

Figure 1: Cognitive Theory of Multimedia Learning. Adopted from Mayer (2006).

The assumptions give three different aspects of multimedia design that should be addressed. The amount of information that can be processed is limited, which is why it is important to exclude extraneous information. Organising information is essential, so the information should be presented in a suitably organised form. In addition, it is important to foster the generative processing, mainly through motivation in a more social context.

The consequences of the CTML model give rise to 12 principles of multimedia design as formulated by Mayer (2006) (Table 1), where the first five deal basically with reducing extraneous processing/information. The last four foster generative processing, and the middle three concern managing essential processing. These principles are the general basis for our design of instructional videos. However, we have kept the speaker in view as a social cue, and for increasing the personal connection (personalisation principle).

Table 1: Principles of Multimedia Learning (Mayer, 2006)

Coherence principle:	People learn better when extraneous words, pictures, and sounds are excluded rather than included.
Signalling principle:	People learn better when cues that highlight the organisation of essential material are added.
Redundancy principle:	People learn better from graphs and narration, than from graphics, narration, and on-screen text.
Spatial contiguity principle:	People learn better when corresponding words and pictures are presented near rather than far from each other on the page or screen.
Temporal contiguity principle:	People learn better when corresponding words and pictures are presented simultaneously rather than successively.
Segmenting principle	People learn better when a multimedia lesson is presented in user-paced segments rather than as a continuous unit.
Pre-training principle:	People learn better from a multimedia lesson when they know the names and characteristics of the main concepts.
Modality principle:	People learn better from graphics and narration than from animation and on-screen text.
Multimedia principle	People learn better from words and pictures than from words alone.
Personalisation principle	People learn better from multimedia lessons when the words are in conversational style rather than formal style.
Voice principle	People learn better when words are spoken in a standard-accented human voice than in a machine voice or foreign-accented human voice.
Image principle	People do not necessarily learn better from a multimedia lesson when the speaker's image is added to the screen.

INSTRUCTIONAL VIDEO DESIGN

The rules for communicating through moving images have developed rapidly over the past century, from the one-shot precursors to the story film like "Workers Leaving the Factory" by the Lumière Brothers from 1895 and early fiction short films like Georges Méliès' "Cinderella" from 1899, his first work with more than one shot (Thompson & Bordwell, 2000, p. 14); to today's complex narrative works created from and available on multiple digital and analogue platforms, such as Peter Jackson's *Lord of the Rings* trilogy from 2001–2003. Digital media have made the moving visual narrative language available to almost anybody almost anywhere (Thompson & Bordwell, 2000, p. 730). Thus, the language of cinema is in this regard a new and somewhat unfinished language compared to other forms of communication and the conventions in other visual arts. Like linguistic forms of communication, the language of cinema [we use this term loosely to include video with film and television] is dynamic and continuously developing, but the basic conventions are now well estab-

lished, at least in part by its accessibility to wide audiences through these multiple platforms of new media (Manovich, 2001 p. 1); and the foundation for communication through moving images is strong and clear.

The main target audience for instructional/educational videos is students attending an institution for learning – a school, college or university – most of whom are young enough to be a part of the post-MTV media consciousness. This audience became media savvy at a young age and is comfortable relating to audio-visual presentations (Brooks et al., 2000, pp. 5–6). Hence, the established visual languages speak clearly to them, and it is natural to use the rules and conventions of contemporary cinema and television as basic principles when communicating to them with the use of video.

Richard E. Mayer's Principles of Multimedia Learning (Mayer, 2006) (Table 1) provides a set of supplementary pointers for how to approach the production of an instructional/educational video, as a further specification of some of the elements in the basic language of moving images.

The "Analytical Balance" video aims to instruct the viewers in the correct procedure of using an analytical balance, a specifically constructed scale used in chemistry to find exact mass of chemicals down to accuracy in, e.g. tenths of milligrams. The importance of exact measurements in chemical research requires that such a scale is used in a very precise procedure, and this video demonstrates each step of the procedure in correct order accompanied by spoken instructions.

For the "Analytical Balance" video, the main production aspects focussed on eliminating extraneous visual information and simplifying the visual expression so that there would be a clear focus on the series of actions needed to do a correct weighing session. The script was prepared by the video producer in close cooperation with an experienced chemistry professor and a didactically oriented scholar.

We chose to shoot the video on location in a proper analytical balance laboratory in order to show the correct setting for the balance. This meant we had to clear the room of all possible distractors and block the shots to fall within the axes of action determined by the presenter, the balance and the camera (since the presenter addressed the camera directly), and still get unambiguous shots of the actions.

We also chose to have the on-camera presenter communicate verbally directly to the viewers, as opposed to adding a voiceover. We did this for two reasons: first, the presenter could then physically demonstrate the use of the balance and verbally explain the demonstration simultaneously, which provided synchronized information through the dual channels (auditory and visual) to maximise the cognitive effect on the viewers; second, the presence of the presenter in the video – made necessary by the need for a visual demonstration of the balance – could create a distractor if the presenter's face or verbal contribution had been omitted completely. The presenter's full dual channel presentation fulfilled several of Mayer's principles, most notably the Coherence and Temporal Contiguity principles, but it also provided a social cue – a face for the viewers to relate to.

For the actual production, we ensured that the technical quality of the video was as good as possible, with sharp images and good sound. The video is a combination of on-camera monologue featuring the presenter's face in an establishing shot and close-up shots of details. These are juxtaposed according to the narrative, and provide for the close-ups to

show details in actions and equipment in sync with the verbal presentation. We also ensured clarity in the presentation by directing the presenter's on-camera performance so that her visually isolated on-screen actions (close-ups of details) were simple and calculated and her verbal presentation intelligible and accurate – following the strictly written script to the point – yet informal enough not to risk alienating the viewers. All these production elements were adapted to focus the video presentation as well as possible on the necessary content, as per the personalisation, voice and image principles.

The video was finalised by the video producer in cooperation with the chemistry professor. The final result is a video that instructs as clearly and concentratedly as possible on the use of an analytical balance, for the use of chemistry students nationwide.

EYE TRACKING

Eye movements can reveal information about underlying cognitive processes (Just & Carpender, 1984). The working hypothesis is that there exists a strong correlation between where one is looking and what one is thinking about, the so called "eye-mind" hypothesis (Just & Carpender, 1984). Eye tracking technology (Holmqvist et al., 2011) is an excellent tool to observe eye movements. This means that an eye movement recording, using eye tracking equipment, can provide a dynamic trace of where a viewer's attention is directed in relation to a visual display. Measuring different aspects of eye movement, such as duration and sequence of fixations, indicates an extensive processing (Rayner, 1998). It was suggested by Rayner (1998) that eye movement parameters such as number of fixations, fixation duration and total inspection time are relevant to learning.

In this study we want to examine which objects in a video get the students' attention, which is where they look, and the total time these objects are observed. The video design is based on Mayer's multimedia learning principles (Table 1) and contemporary television and cinema conventions, and we expect viewers' attention to be on the relevant objects in the video.

EXPERIMENTAL SET-UP

Participants' eye movements were recorded with an integrated Tobii X2-60 eye tracker (Tobii, 2015) with a 17" display. The eye tracker apparatus is located below the display. The camera recorded the participants' movements while watching the video. The eye tracker data were collected and analysed using Tobii Studio software (Tobii, 2015). The Tobii X2-60 tracks eye movements with a sampling frequency of 60 Hz and an angular resolution of 0.25°.[1] With the viewer placed about 60 cm from the screen, this provided a sufficient accuracy for an analysis of different objects; in this case the resolution is better than 3 mm on the screen. Due to limitations in the eye tracking software, it was not possible for the participants to pause or rewind the video. This will give rise to an unnatural situation when

1. The sampling frequency gives information on the quality of the data in general terms. A higher frequency gives better data on fast eye-movements. In this case are we more interested in relatively slow movements and duration of fixations. The sampling frequency and angular resolution is given as information in order to compare different studies.

watching a video as a learning object, provided the student normally pauses or actively searches for information in the video. However, the situation is not entirely unlike a typical teaching situation where the students are not able to ask questions.

Participant sample

The 28 participants involved in this study were all first year engineering students taking a compulsory chemistry course. The course, given during the spring semester, consists of lectures as well as laboratory work. All participants, 14 men and 14 women, volunteered for the study and were given a small monetary compensation for participating. The number of students participating made it necessary to compile data for four weeks, which is why some students had already done some experimental work and knew how to operate an analytical balance.

Procedure

Each participant was tested individually with a set procedure:

a. An introduction to the test procedure and eye tracker technique;
b. Calibration of the Tobii X2-60 to the participant's gaze;
c. Filling out a questionnaire on demographics and study habits;
d. Watching the instructional video;
e. Questionnaire on how the participant experienced the video;
f. Specific questions on the subject of the video;
g. Reviewing of the video with eye tracker markings (optional).

Participants were given the option of reviewing the video with eye tracker markings and were asked to comment. Most of the participants volunteered for this option, and notes of these sessions were taken.

The study was conducted at the Norwegian University of Science and Technology in Trondheim between January and March 2015.

DATA ANALYSIS

In this study we were interested in the viewers' attention to different objects in the video and how much time they focussed on them. We divided the objects into two categories: attractors and distractors, where attractors are defined as objects related to the presentation in a positive way. Distractors are objects that take attention away from the attractors in the presentation. As an example a distractor might be a dark spot on the wall, or part of the equipment not referred to in the presentation at that specific time in the video. This means that an object will be an attractor in one instance and a distractor in another. The presenter, being in view almost 50% of the length of the video, is considered an attractor at all times, as she is in view at the same time as she is talking.

Different objects were assigned Areas Of Interest (AOI) in Tobii Studio, making it possible to analyse the visit duration of each individual AOI. The AOIs were slightly larger than the objects in order to take the uncertainty of the gaze into account.

Figure 2: Examples on Areas Of Interest during a scene. The presenter is considered an attractor in this scene while the balance is considered a distractor, as it is not addressed in the narration at this moment.

Figure 3: Areas of interest defined as attractors during a narration where instructions on how to operate the balance are given.

RESULTS

The video analysed had a total duration of 290 seconds, with the presenter in view for 130.4 seconds. The presenter (head) played an important role in conveying the message, as could be seen in the visit duration of the presenter AOI, being on average 97.1 seconds or 74.4%

of the time the presenter is in view (33.5% of the total length of the video). This is hardly surprising as we removed most of the possible distractors, and made sure the presenter was addressing the viewer by looking into the camera. We also found that the viewers focused mainly on the eyes or mouth.

The distractors, being defined as objects not related to the presentation, took the attention of the viewer on average 100 seconds or 34.6 % of the total length of the video. However, this time includes the time the gaze was not recorded in AOIs and this also includes time when the gaze was not detected by the eye tracker, which was typically 15% of the total time. However, this does not mean that the viewer lost attention on the presentation, but rather that the gaze was at an object not related to the presentation in that instance, or that the eye tracker didn't find the gaze. The auditory channel was unaffected and therefore we may assume that the viewers were still paying attention.

The gaze was directed to objects when mentioned and/or displayed in a way that made it clear that they made the connection between object and narration. When showing a new scene, we observed that the viewer scanned it for a few moments before settling on an object. We also noticed that any action draws the attention of the viewer, something that is expected. But we also found that two, for us unnoticed, distractors were visible. A dark spot on the wall and a power outlet drew the attention of about 60 % of the participants. When reviewing the video most of the participants were not conscious of having looked at them at all. The short attention time confirms this as being unconscious, which is why one might assume they played a minor role in the cognitive processing, and thus can be considered as silent distractors.

During the review, most participants commented on how they watched the presenter, with surprise that they looked so much at the eyes or mouth. About 40% commented on the use of an upper door on the balance, as those who had used the balance in the laboratory had missed this possibility. The comments were overall positive both towards the video and the experience using eye tracker.

In addition to reviewing the video afterwards, the participants were asked to fill in a questionnaire immediately after watching the video. Almost all participants thought the video, the contents and length of the video was good (3.34 and 3.41 out of 5 respectively). Both the level and the ability to focus on the content were judged as good. The usefulness of the video was judged as slightly lower (2.97). On the question whether they learned anything new, the responses showed a relatively low yield (2.66); still most indicated that they had learned something new. However, one must keep in mind that the intention of the video is to give an introduction and the possibility to review it prior to use of an analytical balance. The participants had already experience of using a balance, which is why the result, on these questions, is not surprising.

In order to assess learning, we also had three questions on the use of an analytical balance. Two were on the theoretical foundation: why does the balance stand on a heavy table, and what does it mean to *tare*. These questions were correctly answered by all. The last question was more on the point when you can take measurement data (the correct answer being when all doors are closed and the balance shows that the measurement is done). This was indirectly addressed in the video both in narration and in action. In this case, 43% gave an incomplete answer, missing one of the options.

DISCUSSION

The multimedia learning principles of Mayer (2006) serve as a guide on how to produce multimedia learning objects. We have applied the principles to instructional video development and performed an eye tracker analysis of viewers' gaze when watching the video. The results indicate that the presenter, in this context, is important, as the participants focus their attention on her, especially her eyes and mouth. This seems to contradict the use of the image principle (Table 1), but this is not the case as we focus more on the personalisation and voice principles. We also avoided including any text in this video, as it was not necessary. The attempt to give more focus on narration (and the presenter) is judged as successful, as the attention was on the presenter for almost 75% of the time the presenter was in view.

The use of video as a learning object is increasing, although with most people not knowing how to design them to be as efficient as possible. In addition to the intentions and scope of the video as a learning object, one has to incorporate rules and conventions of contemporary cinema and television as basic principles, as well as Mayer's multimedia learning principles. This can make the production of instructional videos quite complicated, if the video is to have the desired effect on students' learning. However, it is possible to improve the quality substantially by following a few rules. It is important to remove as many distractors as possible, for example bad sound and video quality. The background should also be as neutral as possible and actions and movements should be limited to those relevant for the purpose of the video. From our results, a "personal" contact with the presenter seems to be important, which is why the presenter should be visible with sufficient resolution, so that the eyes and mouth are clearly visible.

We have shown by employing eye tracker technology that the use of Mayer's' multimedia learning principles help the viewer to focus on what is assumed to be important. The use of eye trackers will also give more insights into design aspects of learning videos and is expected to play an important role in the future, which is why we are continuing our studies on instructional video design using eye tracker technology.

Acknowledgement

We would like to thank Åshild Samseth, who acted as presenter in the video.

REFERENCES

Brooks, G., Hughes, J., Ritchie, L., Roberts, S., & Wright, K. (2005). *Digital Beginnings: Young Children's Use of Popular Culture, Media and New Technologies.* University of Sheffield.

Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & Van de Weijer, J. (2011). *Eye Tracking: A Comprehensive Guide to Methods and Measures*. Oxford University Press.

Just, M. A., & Carpenter, P. A. (1984). Using eye fixations to study reading comprehension. *New Methods in Reading Comprehension Research*, 151–182.

Manovich, L. (2001). *The Language of New Media.* MIT Press.

Mayer R. E. (2003), The promise of multimedia learning: Using the same instructional design methods across different media. *Learning and Instruction, 3*(2), 125–139.

Mayer R. E. (2006). *Multimedia Learning*. Cambridge University Press, ISBN: 978-0-521-51412-5.

NTNU Openvideo (2015) http://video.adm.ntnu.no/pres/540cba26699ec (accessed 2016-03-15).
Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124*(3), 372.
Thompson, K., & Bordwell, D. (2010). *Film History: An Introduction. 3rd Ed.* McGraw-Hill.
Tobii (2015). http://www.tobiipro.com/ (accessed 2016-03-15).