Effects of extra information on TV viewers' visual attention, message processing ability, and cognitive workload

By Jan Kallenbach, Silja Narhi, Pirkko Oittinen
Interactive TV, [Vol. 5, No. 2]

DOI: 10.1145/1279540.1279548

This article explores how viewers respond to and remember extra information shown on television. Its purpose is to investigate the effects of typical print content on TV viewers' visual attention, message processing ability, and cognitive workload when print content appears simultaneously with video. Test subjects looked at various video clips simultaneously with print, while variable text and graphics changed the structural complexity of the message. Data of psychophysiological measurements such as eye-tracking and skin-conductivity were obtained during the test; memory and cognitive workload measurements were self-reported afterwards. The results show that visual attention was equally balanced between video and print content for lengthy texts, whereas short texts with an increasing number of images caused subjects to look longer and more often at the video. Both video and text recognition were better on lengthy texts. Subjective cognitive workload data shows that lengthy texts resulted in higher ratings for psychological stress than short ones did. We conclude that variable text influences visual attention, memory, and cognitive workload more than pictorial information does; this conclusion supports the limited capacity model.

1. Introduction

Entertainment media content attracts an increasing number of viewers. Although computer and video games are the most prominent examples of this tendency, television, and in particular interactive television, is not an exception [Vorderer 2001]. In fact, it serves as a best case example, since programs on serious subjects like politics and economics are losing viewers while entertainment TV now dominates. Moreover, people are deciding more freely what they want to enjoy, when, and through which media. In this context, the complexity of the content, i.e. its (simultaneous) presentation in various modalities for example, seems to influence the peoples feeling of being entertained little only. Actually, entertainment seems "more and more to be a crucial condition for successful information processing" [Vorderer 2001].

The DigiTV and Print project explores the provision and distribution of print content over the digital broadcast network to media consumers. Carried out by the Helsinki University of Technology and VTT Information Technology with the participation of several media companies, the project employs self-developed prototypes as a research platform. Our project was motivated, on the one hand, by the availability of high-definition coding formats and high-resolution displays that allow consumers to view and interact with high-quality content and, on the other hand, by the rising costs to newspapers and magazines for the delivery of printed issues, especially to remote locations (although in Finland these costs are covered by digital terrestrial television).

If we wish to present typical magazine content that comes in various page sizes simultaneously with TV video, various questions arise: How (long) will people look at both the video and print stimuli on the TV screen, assuming they pay attention to both? What effect will this have on their ability to understand and remember both the video and print content, assuming their capacity to process information is limited? Moreover, what will be the effects on their cognitive capacity?

Clearly, exposure to information in different modalities simultaneously demands more from the viewer. This is especially true when dynamic, time-based stimuli such as moving pictures and aural messages are mixed with static content like text and still images Cognitive processes and structures such as attention and memory are taxed by the simultaneous processing of different information modalities.

Hence, the goals of this article are to explore how the simultaneous presentation of television and print content influences viewers' visual attention, message processing ability, and cognitive workload.

• 1.1 Visual Attention

Research on this topic investigates peoples' cognitive and neural mechanisms and activities, as well as the effects of looking at other people or objects. Research focuses on, for example, eye movement of infants, the effects of impaired vision, visual attention span, and so on.

Media and communication research investigate how much attention people pay to mediated messages. Traditional methods such as eyes-on-screen observations apply to investigations of sequential presentations of information [Krugman et al. 1995]. However, these methods do not work when attempting to research those elements in a mediated message that will attract peoples' attention. Eye tracking is a more modern way to measure visual attention: it is highly accurate; it records data in real-time during exposure to stimuli; and is minimally, or not at all, intrusive. Being also a measure of visual perception, eye-tracking data provides information about the repetitive sequences and saccades of the human eye, known as scanpaths [Josephson and Holmes 2002]. The data allows identifying regions of interest by showing how often and how long people look at them.

We want to find out how print content affects visual attention when shown simultaneously with the television broadcast stream. This investigation becomes especially interesting when considering the (spatial) structural complexity of such television messages. Print content is mostly text and a fixed number of images. Digital television, and interactive television in particular, allow the dynamic and personalized delivery of additional information. Hence the amount of text and pictorial information can vary and influence the viewer's visual attention span. We expect that textual information will influence visual attention the most, since in order to understand the content, it requires reading as a conscious activity.

- **1.2 The Limited Capacity Model**

The limited capacity model [Bolls et al. 1996; Lang 1995; 2000; Lang et al. 1996; Lang et al. 2000] considers humans as information processors that have a limited capacity for processing mediated messages. It distinguishes three simultaneously and continuously working subprocesses: encoding, storage, and retrieval. When people look at television messages their mental resources are individually and flexibly allocated to each subprocess. On the one hand, allocation occurs automatically when viewers process the structural information in the message; on the other hand, viewers can control the allocation according to their own needs, goals, and interests. As long as there are enough resources to allocate to all three subprocesses, the act of watching television is successful. But processing decreases when sufficient resources are not available, which affects encoding, storage, and retrieval differently.

According to the model, variations in the structure of mediated messages cause distinct effects on the processes of remembering and sense-making. Hence encoding attracts special interest when addressing the simultaneous processing of dynamic and static content in media. "The encoding subprocess is a two-step process through which specific bits of information contained in the original message are selected from the myriad information bits available in the sensory store and transformed into activated mental representations in working or short-term memory" [Lang 2000, p. 48]. The first step is to select information to be transformed into mental representations; selection takes place in both a controlled and in an automatic way. The person's goals, needs, and tasks influence both controlled and automatic selection. Automatic selection occurs unintentionally and unconsciously as a response to a "change or an unexpected occurrence in the environment" [Lang 2000, p. 49].

The majority of empirical studies demonstrate that most of the available mental resources are used for encoding [Bolls et al. 1996]. In their experiment, Bolls et al. changed the speed of a message by modifying the number of cuts, and hence increased the viewers' cognitive load (since they must now process the message's structural components (light, sound, editing techniques) as well as its content elements (narrative structure, genre, and emotion). With regard to encoding, the study shows that an increase in speed (i.e., temporal structural complexity) increases memory to a certain level, where it stays constant or decreases. Lang et al. [1996] hypothesized that a higher structural complexity, indicated by a greater number of cuts, will increase skin conductivity (measured online). They stated that a higher physiological arousal entails an increased capacity for encoding a complex message. Their findings show that skin conductivity does not change significantly; but electrodermal activity has a significant interactive effect on the sensation, complexity, and duration of the message. Another study by Lang et al. [2000] also reports that an increase in the speed of a message increases the allocation of processing resources to it, indicated by increased skin conductivity. Moreover, high speed and arousing messages can cause an overload in information processing capacity, indicated by a decrease in memory.

In light of our study and based on the empirical findings above, we argue that extra information in the form of texts of varying lengths and number of images increases the overall (spatial) structural complexity of a mediated message if it is shown simultaneously with the television stream. We repeat the argument in Section 1.1 that such stimuli affect visual attention, and subsequently the process of encoding. To investigate these effects, this study incorporates eye tracking as a measure of visual attention, post-experimental memory tasks via multiple- choice tests to measure encoding [Lang 2000], as well as measuring electrodermal activity (EDA) that indicates changes in the allocation of processing capacity.

- **1.3 Cognitive Workload**

This article investigates cognitive workload by means of an adaptation of the established self-evaluation method SWAT (Subjective Workload Assessment Technique) called A_{SWAT}, after Luximon and Goonetilleke [2001]. A_{SWAT} is a multidimensional subjective rating scale that measures three kinds of workload: time, cognitive effort, and psychological stress. Time load is the degree to which a task must be completed in a given amount of time. Cognitive load consists of demands on attention, such as attending to multiple sources of information and performing calculations. Psychological stress refers to variables such as fatigue, level of training, and emotional state of the human being. Unlike SWAT, the A_{SWAT} method uses no pretask procedure to assign weights to the different dimensions; the overall mental workload is obtained by using an unweighted average of all three dimensions.

According to Luximon and Goonetilleke [2001], A_{SWAT} is very sensitive to low levels of cognitive workload (which was a problem for SWAT). Although studies of cognitive workload in entertainment are relatively rare compared to other more traditional areas such as driving a car or flying, some of their concepts can be applied to media research as well: for example, research on divided attention for in-vehicle entertainment or navigation systems [Landsdown et al. 2004; Park and Cha 1998].

In a television presentation that simultaneously shows print and video content, viewers must continuously divide their attention in order to understand and process the information. Hence, we argue that this affects their cognitive workload. Our study utilizes the A_{SWAT} methodology, via self-reporting, as a subjective evaluation of cognitive workload.

At this point we allude to an important difference between the limited capacity model and the underlying assumptions of the cognitive workload methodology. While the limited capacity model assumes a fixed common capacity for encoding, storage, and retrieval; the cognitive workload approach can also measure information overload (the scale is evaluated subjectively, allowing for large interpersonal variances).

2. Experiment

• 2.1 Design

The experiment incorporates a 2-factor within-subject 2 (text length) X 3 (number of images in the text) design. It has 6 different texts of varying lengths (3 short and 3 long) containing a varying number of images (0, 1, or 2). There are 6 print articles of varying structural complexity associated with 6 different television video clips. The design calls for every participant to watch 6 stimulus screens during which physiological measurements were taken. Immediately after watching each screen, the subjects answered the questions and assessed their workloads.

• 2.2 Subjects

Twelve undergraduate and postgraduate students at the Helsinki University of Technology and the Laboratory of Media Technology took part in the experiment. The sample was made up of 6 women and 6 men with a mean age of 25.5 years (SD=2.75).

• 2.3 Stimuli

The six stimuli screens consisted of a television video clip, which occupies approximately the upper-left quarter of the screen and an adjacent short article covering approximately the right-half of the screen.

The video clips were scenes extracted from the first episode of the US television sitcom "Joey." At the time of the experiment the series had not (yet) aired in Finland, which minimized the risk that some participants in the study may already have seen it. The scenes were shown in order of appearance to ensure a continuous story line. The clip length varied between 57 seconds and 145 seconds (M=123.33, SD=33.02); the lower-left quarter of the screens was occupied by the logo "Joey" (see **Figure 1**).

The short texts varied between 82 and 87 words (M=83.67, SD=2.89), whereas the longer texts varied between 154 and 171 words (M=160.33, SD=9.29). In texts with one image, the image was placed at the beginning of the text. In texts containing two images, one was placed at the beginning and one at the end. No texts were directly related to the content of the associated video, but did contain information about the series itself or the main actor.

Short and long texts were presented alternately; the first stimulus screen had a short text (S) with one (1) image, followed by a long text (L) with no images (0). The third screen again had a short text, but now with two images (2). This sequence of images was repeated for the remaining three screens. Thus, the complete order of the stimuli screens was S1-L0-S2-L1-S0-L2. Their presentation time was the length of their associated videos; **Fig. 1** shows an example of the 6^{th} stimulus screen.

All stimuli screens were shown as PowerPoint presentation on a Fujitsu General PDS 4242E-S 42" wide plasma display connected to the stimulus computer. The original resolution of 1600 X 1200 pixels was widened to fit the 16:9 aspect ratio (93cm X 52.5cm) of the plasma display.

• 2.4 Environment

The room was modified to resemble a living room: a couch, a carpet, plants, as well as dim lighting and a higher room temperature (22°C) contributed to this scheme. Office dividers were used to separate the "living room" from the operational and testing areas.

• 2.5 Measuring the Physiological Response

A mobile eye-tracking system from SensoMotoric Instruments (SMI) was used to track the subjects' eye movements. The helmet-mounted tracking hardware was connected to a computer running iViewX software. The program recorded the eye data with a sample rate of 50Hz. The variables of interest were the horizontal and vertical points of regard (focus), recorded as calculated pixels on the stimulus screen.

The subjects' electrodermal activity was measured with a biofeedback system from Thought Technologies Ltd. The Ag/Ag/Cl electrodes were placed onto the index and ring fingers of the nondominant hand. The ProComp Infinity encoder amplified and rectified their signals and sent them to the stimulus computer. The BioGraph Infinity software recorded the skin conductivity values in µS with a sampling frequency of 32Hz.

- **2.6 Self-Report Measures**

After each stimulus screen, subjects were presented with a questionnaire to assess how well or badly they remembered the content and their cognitive workload.

Encoding was measured by asking six questions about the content about both video and text; there were no questions about the images. For each question, the subjects selected one of four alternatives: three of them recounted information about objects or events in the video, but only one alternative provided the correct answer. The fourth alternative allowed the subjects to indicate that they did not know or remember the answer.

Finally, the subjects assessed their individual workloads using the A_{swat} methodology. Mental effort, time, and stress loads were indicated on a continuous scale from 0 to 100 %.

All questionnaires were completed electronically and saved using a Fujitsu Siemens Stylistic ST4110 W Tablet PC running on Microsoft InfoPath.

- **2.7 Procedure**

The subjects were briefly informed of the tasks they needed to do in the experiment and, on a general level, on the nature of the experiment; they were not told of the study's objectives or variables. The subjects were asked to relax and enjoy the content during the experiment. After mounting the EDA electrodes and the helmet with the eye-tracking system, the left eye of the subject was calibrated to the plasma display using a five-point calibration procedure (the four corners and a center point). Following this, the base levels of skin conductivity were recorded, during which subjects were told to relax and sit still with eyes open. The plasma display was switched off and silence maintained. Then the base-level data was stored on the stimulus and eye-tracking computers. The experiment was started when all data was free of obvious noise and disturbance.

The stimulus and the eye-tracking computers were operated independently by one experimenter each. Due to a lack of synchronization between the various hardware and software components (see Ravaja [2004]), the electrodermal activity was recorded first, followed by the recording of eye data. Finally, the subject was alerted, the plasma display was switched on, and a stimulus screen was shown.

After each stimulus ended, the recordings of electrodermal activity and eye-tracking were stopped. Then the plasma display was switched off and the questionnaire was given to the subjects. Finally, all data was saved in individual files. This procedure was repeated for all six stimulus screens.

- **2.8 Data Reduction and Analysis**

A data-cleaning method was applied to the eye data. Since blinks of the eye or temporary tracking errors prevented the proper detection of eye movements, this data was excluded. Hence histograms were created for every time series and the upper and lower bounds were decided individually to separate outliers from the valid data.

Means were calculated from all psychophysiological measures over time, that is, for every stimulus screen and each participant. The change in skin conductivity was calculated as a change in percent against the base-level means for each subject. The amount of time the subjects fixed their gaze on the screen was computed using the Matlab 7 extension ILAB 3.6.4 [Gitelman 2002]. The software scaled the stimulus resolution of 1600 x 1200 down to 640 x 480. The amount of time during which the subjects fixed their gaze on the video or on the text were calculated using the velocity/distance model with a half-degree of horizontal and vertical variation and a fixed gaze time of 100 milliseconds. A separate analysis was done to distinguish time spent looking at the video from that looking at the text; the times were standardized on a per minute basis to enable comparisons among the subjects.

Answers to text and video recognition questions were checked for correctness and the means for the correct video and text answers were calculated in percentages.

Finally, means for the psychophysiological data and the self-report of the six stimulus screens were analyzed via the variance method (ANOVA).

3. Results

- **3.1 Visual Attention**

Due to our interest in how frequently and how long the subjects look at the text or video, the following research questions arise:

> ***RQ1:*** *How often do subjects look at either the video or the articles?*

> ***RQ2:*** *How long do subjects look at either the video or the articles?*

The ratio between the number of fixed looks at the video and those at the text is shown in **Fig. 2**. The graph shows that for short texts the number of images were a main factor, whereas for long texts it was not. That is, in a short text with two images the subjects looked almost twice as often at the video than at the text, whereas in the short text with no images the ratio was almost the opposite. In contrast, the subjects looked at long texts more often than at the video, independent of the number of images in the texts. The ANOVA results show the number of images have a significant main effect, $F(2,66)=6.74$, $p<0.01$; the text length has a significant effect, $F(1,66)=9.13$, $p<0.01$, and a significant interaction effect, $F(1,66)=10.81$, $p<0.01$.

The number of looks is tightly coupled to their duration. Thus, the analysis of the duration of looks produced a similar result, shown in **Fig. 3**. Basically, all videos are viewed longer than texts, as indicated by the means ratio being greater than one. In particular, for the short text with two images, the subjects looked at the video more than three times longer than they did at the text, whereas for the short text with no images, the video was looked at almost as long as the text; the number of images in the long texts did not have an effect. Here the subjects looked at the videos almost two times longer than at the texts. Again, ANOVA shows that the number or images had a significant main effect, i.e., $F(2,66)=6.71$, $p<0.01$; the text length had a significant effect, $F(1,66)=11.44$, $p<0.01$; and a significant interaction effect, $F(1,66)=12.11$, $p<0.001$.

- **3.2 The Limited Capacity Model**

According to the hypotheses and findings of Lang et al. [1996], a higher structural complexity of a mediated message causes better encoding, demonstrated by higher recognition rates. With regard to independent variables, text-length, and number of images in the text, we hypothesize that:

> ***H1a:*** *Test length has a main effect, such that recognition scores increase as texts grow longer, demonstrating better encoding.*
>
> ***H1b:*** *The number of images have a main effect, such that recognition scores increase as the number of images do, demonstrating better encoding.*

Fig. 4 shows the sum of scores for video and text recognition. The graphs show that short and long texts make a significant difference in the scores, $F(1,66)=6.22$, $p<0.05$, whereas the number of images do not, $F(2,66)=0.92$, ns; this data supports statement H1a and does not support H1b. The scores for short and long texts without images are similar on average, whereas one image in a short and a long text caused the biggest difference between them, and almost caused a significant interaction effect, $F(1,66)=3.61$, $p<0.1$.

Tables I and **II** list separate scores for video and text recognition. The standard deviations of the means appear in parentheses under the scores; parentheses under F indicate degrees of freedom.

As more structural message complexity is predicted to increase sensory arousal [Lang et al. 1996; 1999; 2000], it is hypothesized that extra information causes greater sensory arousal, as demonstrated by greater electrodermal activity in the subjects.

> ***H2a:*** *The length of a text has a main effect, such that arousal increases with the length of a text, demonstrated by greater electrodermal activity.*
>
> ***H2b:*** *The number of images has a main effect, such that sensory arousal increases as the number of images do, demonstrated by greater electrodermal activity.*

Fig. 5 shows the change in the subjects' electrodermal activity measured against their base levels. There is a significant interaction effect, $F(1,66)=4.20$, $p<0.05$, but no statistically significant effects due to text length or the number of images. Thus hypotheses H2a and H2b are not supported. However, skin conductivity seems to increase for long texts. A short text without images caused greater activity than a long text without images, whereas a short text with one image caused the least activity overall.

- **3.2 Cognitive Workload**

Landsdown et al. [2004], Luximon and Goonetilleke [2001], and Park and Cha [1998] show that subjectively reported levels of cognitive workloads increase with the amount of information that has to be processed. However, since the content of the images was not assessed, and it is usually processed much faster than text, the number of images do not have an effect, whereas text length does.

> ***H3:*** *Text length has a main effect, such that the self-reported cognitive workloads increase*

along with text.

This hypothesis is supported: text length has a main effect on subjectively assessed cognitive workload, as shown in **Figure 6**, $F(1,66)=4.41$, $p<0.05$. This result was mainly due to the psychological stress load component (**Figure 7**)–the ANOVA resulted in a significant main effect of $F(1,66)=8.53$, $p<0.01$. Cognitive effort and time did not have statistically significant effects.

4. Discussion

This work supports in part the various hypotheses derived from previous studies and applied in the current context. Empirical evidence was found for the three main concepts on which this work is based. Data indicates that the limited capacity model may also be applied to multimodal television presentations containing print content. A_{SWAT}, which measures cognitive workload, is sensitive to mediated stimuli, particularly to psychological stress.

• 4.1 Visual Attention

The outcomes for the number of looks and their duration were surprising. Although the number of images in long texts did not have a main effect, they did have a strong effect in short texts. We argue that long texts represent the right amount of textual information to create a balance between the time spent looking at the text and looking at the video. In relation to the limited capacity model in which controlled and automatic selection of information starts the encoding process, we interpret the results to mean that images in short texts elicit more automatic selection of visual stimuli than controlled selection does. This may be due to the visual impression that a short text with images is easier and faster to read and memorize than a short text without images. Thus subjects were likely to be more attentive to the video because they feel they have more time to process its content and/or due to unexpected visual changes in the video. However, a short text without images and long texts in general, may appear to take a longer time to read and understand, thus eliciting more controlled attention and resulting in similar numbers and duration of looks.

• 4.2 The Limited Capacity Model

We suggest that our results support the model. The scores for the recognition tasks support the prediction that encoding will be better where there is higher structural complexity. Long texts containing images resulted in higher recognition scores than short texts with images, showing that the length of a text has an effect on the structural complexity of a message, more so than the number of images. However, a short text without images produced a mean similar to a long text without images. As in the discussion regarding visual attention, this may be due to the impression that reading and memorizing a short text without images might take as long as reading and memorizing a long text. We suggest that in this case subjects consciously selected the information to encode, whereas they selected short texts containing images more automatically, as video content is delivered continuously over time. This resulted in greater attention being paid to the video and lower recognition scores for short texts.

The results for sensory arousal measured via electrodermal activity did not support the second hypotheses. Although there was a significant interaction effect, no major effects were found that support H2. But the results do confirm the empirical findings of Lang et al. [1996], where higher structural complexity did not result in a significant increase in skin conductivity. However, we argue that the presence of images had an influence on sensory arousal, which increased with the number of images. This supports the findings of Lang et al. [2000].

Generally, a short text with one image produced an outlying result. **Table II** shows the recognition scores for the content of the texts. Clearly, the mean score for the short text containing one image was the lowest, a pattern that was repeated for electrodermal activity. These low values reveal an effect whose cause can be found in our experimental design. Although the stimulus screen sequence was intended to be pseudo-randomized, its fixed order was not, and caused an order effect for all subjects. Thus our advice to the subjects to relax had a strong impact on the results mentioned above. The subjects were indeed relaxed and enjoyed the first stimulus screen. However, although we introduced the task and explained that the experiment was mainly about recall/recognition and that the questions were about the video and text, their expectations did not match reality. This had an order effect for the other videos, since that the subjects were prepared for the task and so were able to perform better.

• 4.3 Cognitive Workload

The A_{SWAT} methodology demonstrated the highest sensitivity in its psychological stress-load component. We argue that this result supports the outcomes of Lang et al. [1996] by the interaction between the timing of the stimuli and the amount of information that had to be encoded. Thus, with higher structural complexity (long texts) the participants felt more psychologically stressed reading and memorizing the information. From another standpoint, this result was surprising, as we expected that the time-load component would have a significant effect as well; but we do not have an explanation for this outcome.

5. Conclusions

This study provides a first base for applying the limited capacity model to the presentation of print media as additional information on television. The high accuracy in the design of the experiments may prove the study's applicability. In the context of entertainment research on physiological responses to various information modalities shown simultaneously, this study may provide deeper insights into cognitive processes. Although the subjective evaluation methodology A_{SWAT}, which measures cognitive workload, was only sensitive to the psychological stress load, we recommend further research in this direction.

Acknowledgments

We thank TEKES and the participating companies of the DigiTVandPrint project for funding this study.

References

Bolls, P., Potter, R., and Lang, A. 1996. Television arousal and memory: The effects of production pacing and arousing content on encoding, storage, and retrieval of television messages. Paper presented at the annual Midwest Artificial Intelligence and Cognitive Science Conference, Bloomington, IN.

Gitelman, D. R. 2002. ILAB: A program for postexperimental eye movement analysis. *Behavior Research Methods, Instruments and Computers 34*,4, 605–612.

Josephson, S. and Holmes, M.E. 2002. Visual attention to repeated Internet images: Testing the Scanpath theory on the World Wide Web. In *Proceedings of the Eye Tracking Research and Applications Symposium 2002*. ACM, New York, 43–49.

Krugman, D. M., Cameron, G. T., and Mckearney-White, C. 1995. Visual attention to programming and commercials: The use of in-home observations. *J. Advertising XXIV*, 1, 1–12.

Landsdown, T.C., Brook-Carter, N., and Kersloot, T. 2004. Distraction from multiple in-vehicle secondary tasks: vehicle performance and mental workload implications. *Ergonomics 47*, 1, 91–104.

Lang, A. 2000. The limited capacity model of mediated message processing. *J. Communication 50*, 46–67.

Lang, A., Zhou, S., Schwartz, N., Bolls, P. D., and Potter, R. F. 2000. The effects of edits on arousal, attention, and memory for television messages: When an edit is an edit can an edit be too much? *J. Broadcasting and Electronic Media 44*, 94–109.

Lang, A., Bolls, P., Potter, R.F., and Kawahara, K. 1999. The effects of production pacing and arousing content on the information processing of television messages. *J. Broadcasting and Electronic Media 43*, 4, 451–475.

Lang, A., Bolls, P., and Kawahara, K. 1996. The effects of arousing message content and structural complexity on television viewers' level of arousal and allocation of processing resources. Paper presented to the Midwest Artificial Intelligence and Cognitive Science Conference, Bloomington, IN.

Lang, A. 1995. Defining audio/video redundancy from a limited-capacity information processing perspective. *Communication Research 22*, 1, 86–115.

Luximon, A. and Goonetilleke, R.S. 2001. Simplified subjective workload assessment technique. *Ergonomics 44*, 3, 229–243.

Park, P. and Cha, D. 1998 Comparison of subjective mental workload assessment techniques for the evaluation of in-vehicle navigation system usability. In *Proceedings of Session T56 of the 5th World Congress of ITS 98.* (CD-ROM).

Vorderer, P. 2001. It's all entertainment–sure. But what exactly is entertainment? Communication research, media psychology, and the explanation of entertainment experiences. *Poetics 29*, 247–261.

Received July 2006; revised July 2006; accepted July 2006

Authors

Author's address: Helsinki University of Technology, Helsinki, Finland; emails: **jan.kallenbach@tkk.fi**, **snarhi@tkk.fi**, **poittinen@tkk.fi**

Figures

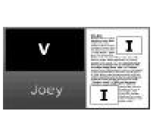
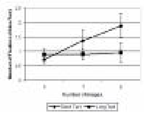
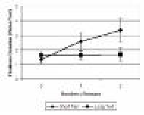


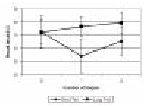
Figure 1. Example of the 6th stimulus screen (L2); V–Video, I–Image.
(http://deliveryimages.acm.org/10.1145/1280000/1279548/figs/f1.html)

Figure 2. Ratio of the number of times looking at videos and articles.
(http://deliveryimages.acm.org/10.1145/1280000/1279548/figs/f2.html)

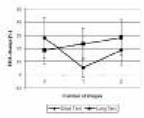




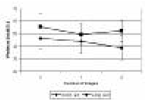
[**Figure 3. Ratio of the amount of time looking at videos and articles. (http://deliveryimages.acm.org/10.1145/1280000/1279548/figs/f3.html)**](http://deliveryimages.acm.org/10.1145/1280000/1279548/figs/f3.html)



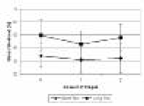
[**Figure 4. Sum of scores for video and text recognition. (http://deliveryimages.acm.org/10.1145/1280000/1279548/figs/f4.html)**](http://deliveryimages.acm.org/10.1145/1280000/1279548/figs/f4.html)



[**Figure 5. Electrodermal activity. (http://deliveryimages.acm.org/10.1145/1280000/1279548/figs/f5.html)**](http://deliveryimages.acm.org/10.1145/1280000/1279548/figs/f5.html)

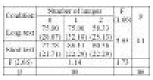


[**Figure 6. Average of subjective cognitive workload ratings. (http://deliveryimages.acm.org/10.1145/1280000/1279548/figs/f6.html)**](http://deliveryimages.acm.org/10.1145/1280000/1279548/figs/f6.html)



[**Figure 7. Subjective psychological stress load ratings. (http://deliveryimages.acm.org/10.1145/1280000/1279548/figs/f7.html)**](http://deliveryimages.acm.org/10.1145/1280000/1279548/figs/f7.html)

Tables



[**Table I. Mean Recognition Rates for Video Content by Text Length and Number of Images (http://deliveryimages.acm.org/10.1145/1280000/1279548/figs/t1.html)**](http://deliveryimages.acm.org/10.1145/1280000/1279548/figs/t1.html)

[**Table II. Mean Recognition Rates for Text Content by Text Length and Number of Images (http://deliveryimages.acm.org/10.1145/1280000/1279548/figs/t2.html)**](http://deliveryimages.acm.org/10.1145/1280000/1279548/figs/t2.html)