Information Modalities for Procedural Instructions: The Influence of Text, Pictures, and Film Clips on Learning and Executing RSI Exercises

CHARLOTTE VAN HOOIJDONK AND EMIEL KRAHMER

Abstract—Much of the empirical research on the effectiveness of different instructional designs has focused on declarative tasks, where a learner acquires knowledge about a certain topic. It is unclear to what extent findings for learning declarative tasks (which are not consistent on all aspects) carry over to learning procedural tasks, where a learner acquires a certain skill. In this paper, we describe an experiment studying a specific kind of procedural instructions, namely exercises for the prevention of repetitive strain injury (RSI), taking information modality (text versus picture versus film clip) and difficulty degree of the exercises (easy versus difficult) into account. In the experiment, participants had to learn RSI exercises and were asked to execute them. The results showed that an instruction in a picture lead to the shortest learning times followed by an instruction in a film clip. An instruction in text led to the longest learning times. For the amount of practicing the exercises during the learning phase, it was found that the participants in the film clip condition hardly engaged in practicing the exercises during the learning phase. The participants in the picture condition engaged in a moderate amount of practicing of the exercises during the learning phase. The participants in the text condition engaged in the most practicing during the learning phase. The results concerning the execution times showed that an instruction in a picture led to the lowest execution times followed by an instruction in a film clip. The instruction in text led to longest execution times. Finally, for the amount of correctly executed exercises, it was found that learning from a film clip led to the highest learning performance, both for easy and for difficult exercises. Learning from an instruction in text led to a fairly good learning performance, both for easy and difficult exercises. Learning from a picture led to a good learning performance for the easy exercises, but the performance dropped for the difficult exercises.

Index Terms—Document design, information modality, procedural learning.

The emergence of computer-based learning has led to new possibilities for presenting instructions and to a renewed interest in the effects of different information modalities. Instructions can be given in plain text, but also in the form of static visuals (e.g., diagrams, pictures) or in dynamic visuals (e.g., animations, film clips). Naturally, this raises the question which (combinations of) modalities are best for which learning task. This question has been addressed in a large number of recent studies (e.g., [1]–[6], among many others).

Each of the aforementioned presentation modes has its own unique characteristics, and its own advantages and disadvantages from an instructional perspective. Language is the basic form of human communication, and one of its main strengths is that it is expressive and explicit (both for concrete and for abstract subject matter). An additional advantage of its written form (as opposed to the spoken variant) is that it is not transient: written sentences remain visible on paper and can be re-read if desired, while spoken sentences are “gone” once they have been uttered. But, reading a text requires considerable skill and effort, and, moreover, text primarily facilitates linear information processing.

A difference between text and pictures is their symbol system. The symbol system used for text is abstract/linguistic, while the one used for pictures is sensory. This affects the type of information that they can convey. For example, pictures can directly communicate perceptual information (e.g., spatial orientation and location). In other words, pictures have a lower “articulatory distance” [7]. Moreover, pictures are not constrained by the linear structure of text, and are therefore arguably more efficient at representing nonlinear relations among objects. In text these relationships often remain implicit [8].

The focus of many recent studies has been on the instruction effects of dynamic visuals, presumably because such instructions only recently have become a real possibility due to the increased computing power of standard multimedia PCs. A number of reasons have been suggested as to why an advantage of dynamic visuals over other presentation possibilities such as text and static visuals might be expected. For instance, it has
been argued that dynamic visuals are beneficial for learning since they offer a “complete model” of a learning task (e.g., [1], [9]); in other words, they are “informationally complete” [10, p. 249]. When static visuals or text are used, learners themselves will have to construct a mental representation of the temporal aspects in the instruction. It has been argued that dynamic illustrations offer a better representation of these temporal aspects, thereby reducing the level of abstraction and supporting a deeper level of understanding than static visuals would do [9]. Arguably, this should facilitate learning since it would reduce learning times, require less practice, and result in fewer errors.

A substantial number of studies have tried to demonstrate this presumed learning benefit, but with mixed results (e.g., [1], [5], [11], and [12]). Two groups of explanations for these mixed results have been offered: one group of explanations stresses the disadvantages of dynamic visuals; another group points to methodological problems in these studies. The first explanation claims that dynamic visuals have a fixed duration which viewers simply have to watch. This set time period may lead to inherently longer learning times [11]. In a similar vein, Lewalter [1] suggested that dynamic visuals may take more time to process than other presentation modes; the information in dynamic visuals changes continuously, and as a result learners could be overwhelmed with the amount of information. Alternatively, it has been suggested that dynamic visuals are processed somewhat superficially; they require little cognitive effort, as they can be watched rather passively (e.g., [10] and [13]). The second explanation, by contrast, argues that the mixed findings so far are primarily due to methodological problems in the various experimental studies. For instance, to be really beneficial, dynamics should have some added value [14], which is not always the case. When it comes to temporal sequencing or indicating direction of motion, it has been claimed that arrows in static visuals may be just as effective as dynamic visuals (e.g., [15]). Tversky, Morrison, and Bétrancourt [11] point out that in some comparative studies there is a lack of equivalence between dynamic and static visuals in content or procedures (e.g., because the dynamic visuals convey more information or involve interactivity). Some researchers have argued that even when dynamic visuals do not lead to improved learning, they are more attractive and motivating than other instruction forms and should be preferred for that reason (e.g., [11], [16], and [17]). However, this “subjective satisfaction” [18, p. 33] of various instruction modes is often not addressed, and when it is the results are equivocal (e.g., [19]).

A further complicating factor is that the learning domain itself might also have an influence on the relative benefits of information modalities [14]. It seems reasonable to assume that different learning domains benefit from different kinds of instructions (cf., also [20]). In the field of learning psychology, many studies have shown that dynamic visuals are used as learning instructions for descriptive, “scientific” explanations (often describing causes and effects), such as mechanical and electronic systems (e.g., brakes, pumps, generators [21], [22]; electronic circuit systems [23]; mathematical concepts (e.g., algebra [24]); or complex natural phenomena (e.g., electricity [25]; lightning [26]; gravitational lensing [1]; meteorological changes [2]). Arguably, some of these scientific learning domains lend themselves better for dynamic visualization than others. Moreover, what holds for descriptive tasks (such as those above), may not apply to procedural ones, such as bandaging a hand [5], folding origami models [27]; or tying nautical knots [28]. These procedural tasks differ from descriptive ones in a number of respects. Not only is the nature of the task different (procedures are inherently more linear—one step following another), but the learning goal is also different (the focus is not only on understanding, but also on acquiring certain capabilities or skills) [29]. One of the factors that might affect the processing of procedural information is the design format [29]. Procedural information is often presented in a text and/or picture format (e.g., route directions, maintenance instructions, assembly instructions). While pictures may help users to form a mental model of the procedure [30], [31], the combination of text plus picture may not always be helpful. For example, a user’s attention has to switch between the information presented in text and picture, leading to the so-called split attention effect [32]. Therefore, document designers have to understand the strengths and limitations of both text and pictures when designing procedural information [8]. Both the type of picture (e.g., overview versus partial view [33], line drawing versus photo [5], object-centered versus body-centered [34], etc.) as well as the type of textual instruction (e.g., user-centered versus object-centered versus environment-centered [35]) may influence a user’s performance of the procedure. To further complicate the situation, it may well be that besides variation between domains there is also variation within domains. Arguably, some learning tasks in a given domain are easier than others, and this
may have an influence on the relative contribution of various instruction formats for those tasks.

In this paper, we report on a study of the effects of task difficulty and information modality (comparing dynamic visuals with static visuals and text) on learning a specific class of procedural tasks, namely exercises aiming at the prevention of repetitive strain injury (RSI). RSI is a general term for disorders that are caused by repetitious use of hands, arms, and shoulders, often as a result of prolonged computer terminal work [36]. Recently, the term CANS (short for complaints of arm, neck and/or shoulders) was introduced as an alternative term, but here we will stick to the more familiar RSI terminology. It is generally assumed that taking regular breaks during computer work in combination with exercises is beneficial for the prevention of RSI (e.g., [37]–[39]). RSI exercises offer a new and understudied learning domain, with various interesting properties. Generally, these exercises involve little or no abstraction, do not consist of a complicated sequence of actions, and are relatively short. Moreover, the exercises are highly recognizable (almost everybody has two hands). It is interesting to observe that current RSI information websites offer a large variety of prevention exercises, in many different presentation formats (see Fig. 1 for a representative selection). These differences raise the natural question of what is the effectiveness of the various presentation formats. Note also that there is substantial variation in the difficulty level of existing RSI exercises, so that the other factor of interest (variation in task difficulty) can be modelled in a fairly straightforward way in this domain.

The potential influence of both presentation modality and task difficulty on learning performance can be motivated from COGNITIVE LOAD THEORY (e.g., [40]–[43]), a theory which aims to develop optimal instructional designs while considering the limitations of the human mind. Cognitive load theory builds on the broadly accepted assumption that the mind combines a short-term (or working) memory of very limited capacity (where all conscious activity and processing of information occurs [44], [45]) and a long-term memory with a virtually unlimited capacity [42]. According to cognitive load theory, learning amounts to the construction of new (or modification of existing) schemata [46], which are subsequently stored in long-term memory. Since the capacity of working memory is severely limited, the cognitive load of learners should be kept at a minimum during learning. In the current version of the theory [42], [43], three kinds of load are distinguished: INTRINSIC load, caused by the structure and intrinsic nature of the learning task; EXTRANEOUS load, caused by the manner of presentation and its influence.
on working memory; and GERMANE load, caused by learners’ efforts to process and comprehend learning material. The sum of these three kinds of load should not exceed working memory capacity in order to avoid cognitive overload. As argued above, the intrinsic load may vary both between and within task domains. Since the intrinsic load of a particular learning task is fixed, instruction design can only influence the extraneous and the germane load, and, obviously, when the intrinsic load of a particular task is high, there is relatively less room for the extraneous and the germane load.

Germane load is a “positive” form of load, since it encourages learners to engage in cognitive processing that may lead to improved schemata construction [42], [47]. This suggests that optimal instructions are those which minimize extraneous and maximize germane load, while simultaneously avoiding overload. A complicating factor is that the distinction between extraneous and germane cognitive load, although intuitively plausible, cannot reliably be measured [47]. In general, measuring cognitive load in realistic learning situations is not straightforward [48]. While some recent attempts have been made to measure cognitive load directly (e.g., the dual task approach advocated by Brünken, Plass, and Leutner [49]), many studies rely on indirect measures such as behavioural or learning outcome measures (e.g., [47] and [49]). Although these measures only relate to cognitive load indirectly, they do assess learning behaviour directly, which is sufficient for our current purposes. In this study, we concentrate on learning times, amount of practicing during learning, execution times, and execution errors.

Arguably, the different information modalities of interest (text, picture, film clip) have different loading potentials, which may influence performance on one or more of these measures. Of all three modalities, it seems likely that text imposes the highest load: It requires more mental effort to read a text than to watch a picture or a film clip, hence it seems likely that learning from text takes longer than learning from pictures or film clips. Potentially, an additional complication for learning RSI prevention exercises from text is that schemata construction may be more involved than for pictures and film clips. In the text case, learners have to “translate” the textual instructions to manual gestures. Notice that this is a concrete instance of Glenberg’s Indexical Hypothesis, stating that readers associate words and phrases with objects and actions in “the real world,” which should facilitate understanding (e.g., [50]–[52]). An instruction in the form of a film clip and (probably to a lesser extent) a picture depicts the hand and arm movements the learner should make, whereas the learner has to infer these gestures when they are presented in text. Hence, it is hypothesized that learners will practice more often while learning from text than while learning from visual presentations. An interesting question is whether the load imposed by text is only extraneous or also partly germane. It might be that the extra effort required to learn from text leads to good learning results (short execution times, few execution errors), especially when the intrinsic load is low (so that overload can be avoided).

Pictures arguably impose the lowest load of the three presentation modalities: viewing a static picture requires little mental effort. Since a picture captures the essential information of a procedure, it is to be expected that learning times for pictures are relatively short. However, due to their static nature, pictures offer little information about the temporal structure of a procedure, and for more complicated exercises (i.e., exercises with a higher intrinsic load) this may hamper schemata construction and might result in suboptimal execution performance.

To what extent film clips impose cognitive load is uncertain. On the one hand, it can be argued that they may induce load, since the film clips continuously change and learners have to process this information, which reduces the cognitive resources available for germane load. But they may also lessen cognitive load by relieving the learner in the translation process, which is required when reading text. The main strength of film clips might well be their “informational completeness”—learners do not have to infer the exact sequence of movements from text or from a single snapshot because the entire sequence of actions is visualized, which arguably facilitates schemata construction. This “completeness” suggests that learners will not practice much during learning. Whether this will also result in few execution errors is not obvious; it may be that the large amount of external support relieves learners of cognitive load demands that they would be able to fulfil, but which might prevent them from performing beneficial cognitive actions for learning.

The findings of this study will not only be theoretically relevant, but may also inform more practical work on multimodal information presentation. An example of such work is a medical question-answering (QA) system, which is currently developed within the IMOGEN (Interactive
Multimodal Output GENeration) project that is embedded in the IMIX research programme. IMIX (Interactive Multimodal Information eXtraction) is a research programme in the field of Dutch speech and language technology. A QA system is an automatic system that can answer a user’s question posed in natural languages (e.g., What does RSI stand for?) with an answer formulated in natural language (e.g., RSI stands for Repetitive Strain Injury). Today, QA systems are not only expected to give answers to simple questions, like What does RSI stand for? but also to more complex questions, like How should I organize my workspace in order to prevent RSI? or What is a good exercise to prevent RSI in my hands? The answers to these questions might be more informative if they contained a picture or an animation, rather than an extensive textual answer [53]. In the IMOGEN project, different aspects of multimodal information presentation are studied in order to improve the output quality of QA systems.

To find out what the actual strengths and weaknesses of the various information modalities are, we describe an experiment in which participants have to learn and execute 20 RSI prevention exercises in two degrees of difficulty. We measure the following features: the influence of presenting an instruction in text; picture or film clips on learning times; the amount of practicing during learning; execution times; and number of execution errors. Besides these objective measures, participants are also asked for their subjective satisfaction.

METHOD

Participants Participants were 30 young adults, between 18 and 30 years of age. Of the participants, 15 were male and 15 were female.

Design The experiment had a 3 (information modality) × 2 (difficulty degree) factorial design, with information modality (dynamic visual [film clip], static visual [picture], text) used as a between-participants variable, and difficulty degree (easy, difficult) as a within-participants variable, and with learning times, amount of practicing during learning; execution times; and number of correctly executed exercises were used as dependent variables. The participants were randomly assigned to an experimental condition.

Stimuli A total of 20 RSI prevention exercises were chosen from websites on RSI prevention and RSI injury prevention software (e.g., http://web.mit.edu/atic/www/disabilities/rsi/exercises.html and http://www.workpace.com). The chosen exercises were all exercises to prevent RSI in hands and arms. Of the 20 exercises, 10 were assumed to be easy to perform, and 10 were assumed to be difficult. The criterion for a difficult exercise was that it should be either a complex symmetrical movement or an asymmetrical movement. Complex symmetrical movements were classified as movements consisting of at least three sequential atomic movements, in the course of which both arms and hands make the same movements. Asymmetrical movements were classified as movements in which the participant should make a different movement with each arm or hand. Easy exercises were neither asymmetrical nor complex movements. Fig. 2 shows representative examples of an easy and a difficult exercise.

![Easy Exercise](image1)

**Easy Exercise**

Hold your hands in front of you with your palms facing downwards. Lift both your index fingers from the knuckles. Next, gently drop your index fingers.

(English translation of Dutch original, in Dutch this exercise contains 29 words)

![Difficult Exercise](image2)

**Difficult Exercise**

Hold your left arm in front of you and drop the left hand down bending at the wrist. Place your right hand on the knuckles of the left hand. Press your right hand towards. (English translation of Dutch original, in Dutch this exercise contains 30 words)

Fig. 2. Two typical exercises: one easy and one difficult.

Note that this operationalization of easy and difficult tasks is based on the relative complexity of the sequence of motoric movements. To find out to what extent this objective criterion coincided with subjective perception of task difficulty, a pre-test was carried out in which nine participants were asked to classify the 20 exercises (presented in text plus picture format, and in random order). They were instructed to make two piles, the first consisting of the 10 exercises they considered easiest to perform, and the second pile consisting of the 10 exercises they considered more difficult to perform. It turned out that of the 10 exercises classified as easy, 7 were indeed perceived as easy...
by the vast majority (>75%) of the participants, while the remaining 3 were perceived as difficult by a majority of the participants (presumably because these consisted of gestures that are motorically simple but not commonly used and hence with which participants may have been less familiar). Of the 10 exercises classified as difficult, 9 were indeed perceived as such by the vast majority (>75%), while the remaining 1 was perceived as easy by most participants (interestingly, this was an exercise that was motorically complex but familiar to most participants). In sum, for 16 of the 20 exercises, there was a clear and consistent correlation between the objective and the perceived difficulty degree.

The 20 RSI prevention exercises were presented in three different formats. In the text condition, the exercises were explained to the participants in a purely textual format. The total amount of words did not differ between the 10 easy and 10 difficult exercises: both counted 268 words in total. Thus, the average length was almost 27 words per exercise. Since some exercises are a few words shorter and others a few words longer, we will only report on mean **total** averages for the 10 easy and the 10 difficult exercises in each condition. Because natural language is often ambiguous, the text exercises were checked for their comprehensibility in a second pre-test with three participants (different from those in the first pre-test). It turned out that a few exercises led to misunderstandings and so these exercises were reformulated. The new versions were presented to the same three participants, and they agreed that the reformulations solved the misunderstandings.

In the picture condition, each of the 20 exercises was displayed in a single photograph. Digital camera pictures were taken of a female who did the exercises. She wore black clothing and the movements were shot against a black background so that only her hands were visible in the picture. The photo depicted the STROKE of the movement, with the phase of the movement as it unfolds in time containing the “semantic content” of the movement [54]. To indicate the direction of movement, arrows were added to the pictures of 9 difficult and 2 easy exercises. The size of the pictures was 1536 × 1014 pixels.

For the film clip condition, the same female in an identical surrounding was filmed with a digital film camera (25 frames per second). The total amount of frames did not differ between easy and difficult exercises: both counted 1097 frames (average film length was thus 4.39 seconds). Again, since some film clips are somewhat shorter, and others somewhat longer, we will report on mean **total** averages for the 10 easy and the 10 difficult exercises.

The exercises were presented to the participants via a website: one website for each condition. The experiment was run on a multimedia PC, with a 17-inch color monitor. In all three conditions, the presentations of the RSI exercises appeared at the centre of the display. On the video website, the movements were shown in a film clip with a start and a stop button and a slider. The participants had the opportunity to view the film as often as they desired, but not much use was made of this option. Exercises were presented in one of two random orders to control for possible learning effects.

**Procedure** The experiment consisted of three parts: a practice session, the central part in which the 20 RSI exercises were presented, and a questionnaire. The participants completed each part individually. Each participant was invited to an experimental laboratory, where he or she took a seat behind the computer. Participants were told that they would receive 20 exercises which they had to learn and perform to assess to what extent they suffered from RSI. In addition, they were led to believe that their hand and arm movements were filmed because a doctor would later look at the recordings of the participants to determine to what extent they suffered from RSI. The procedure of the experiment is depicted in Fig. 3.

After the participants had read the instructions, they could click on the hyperlink *start*, and the first trial exercise would appear. Depending on their experimental condition, the participants read or viewed the trial exercise until they thought that they could properly execute the exercise. Subsequently, they clicked the hyperlink *next* at the bottom of the page. A new webpage appeared with the text “execute trial exercise” at the centre of the page. During the execution of an exercise, participants could not see the instruction. When they had executed the exercise, they clicked the hyperlink *next exercise* at the bottom of the page. After completing the second trial exercise, participants were asked if they had any questions regarding the experimental procedure. If not, the participants could start the actual experiment, proceeding in the same way as they did in the trial session. During the experiment, there was no further interaction between participants and the experiment leader.
Fig. 3. Procedure of the experiment.

After the participants completed the 20 experimental exercises, they received a questionnaire. In this questionnaire, the subjective satisfaction regarding the content and structure of the website as well as the comprehensibility and the attractiveness of the exercises were measured. The questionnaire consisted of 16 bi-polar seven-point semantic differentials addressing structure and content of the websites, as well as comprehensibility and attractiveness of the exercises (see Appendix A). The presentation order of the adjectives was randomized. For processing, the positive adjectives were mapped to one (very positive) and the negative adjectives were mapped to seven (very negative).

The participants were debriefed at the end of the experiment.

**Data Processing** The following data was collected: learning times; number of practiced exercises during the learning time; execution times; number of correctly executed exercises; and the results of the questionnaire.

The time it took the participants to learn and execute the exercises was computed from the data of the log program ProxyPlus (http://www.proxyplus.com). This program registered the times associated with participants' mouse clicks during the experiment. The time period between clicking the hyperlink next which preceded the instruction of an exercise and the hyperlink next which followed the instruction of an exercise was defined as the learning time (see Fig. 3). The time period between clicking the hyperlink next which followed the instruction of an exercise and the hyperlink next that preceded a new instruction of an exercise was defined as the execution time (see Fig. 3).

A digital video camera was used to record the movements the participants made during the experiment. These video recordings of the participants' hands and arms were used to assess whether the participants practiced the exercises during the learning period and to assess their performances while executing the RSI exercises. Occasionally, a participant performed an exercise in a correct but not intended way. These cases were counted as correctly executed exercises. One judge did the assessment; scoring was easy, and the few difficult cases that did arise were resolved after discussion.

Tests for significance were performed using a repeated measures analysis of variance (ANOVA), with a significance threshold of 0.05. For post-hoc tests the Bonferroni method was used. The internal consistency of the four item sets of the questionnaire was measured using Cronbach's alpha.

**RESULTS**

**Learning Times** Table I summarizes the results. First consider the learning time. We found that the difficulty degree had an effect on the amount of time to learn the exercises ($F(1,27) = 37.35, p < 0.001$). Overall, the participants needed more time to learn the difficult exercises than the easy ones. Also, the information presentation had
an effect on the learning time \( F(2,27) = 4.53, p < 0.025 \). Participants in the picture condition required the shortest learning times, participants in the text condition had the longest learning times, with learning times from film clips falling in between those two. Post-hoc tests indicated that the instruction in text differed significantly from the instruction in a picture \( (p < 0.025) \). There was no significant difference between the instruction in text and instruction in a film clip \( (p = 0.35) \). Also, the instruction in a picture did not differ significantly from the instruction in a film clip \( (p = 0.25) \). No significant interactions between difficulty degree and information modality were found.

**Amount of Practicing During the Learning Period**
Table I also reveals that the participants practiced the difficult exercises more often than the easy ones during the learning period, and this difference was found to be statistically significant \( F(1,27) = 9.00, p < 0.01 \). Also, information presentation had an effect on the amount of practicing during the learning period \( F(2,27) = 6.76, p < 0.005 \). In the film clip condition, participants almost never practiced; in the picture condition, they practiced for about a fifth of the exercises, while in the text condition, participants practiced about half of the exercises. Post-hoc tests showed that the difference between the instruction in text and the instruction in a picture was not significant \( (p = 0.06) \). Text differed significantly from the instruction in a film clip \( (p < 0.005) \). There was no significant difference between the instruction in a picture and a film clip \( (p = 0.43) \), presumably because of the relatively high standard deviation. No significant interaction effects were found.

**Execution Times** The difficulty degree had a main effect on the amount of time needed to perform the exercises \( F(1,27) = 20.84, p < 0.001 \). The participants required more time to execute the difficult exercises than the easy ones. There was also a main effect of information modality on the execution times \( F(2,27) = 26.78, p < 0.001 \). Participants in the text condition had much longer execution times than those in the picture and film clip conditions. Participants in the picture condition needed somewhat less time for execution than the participants in the film clip condition, but the differences are relatively small and the standard deviations are relatively high. Post-hoc tests indicated that there was a significant difference between the instruction in text and the instruction in a picture \( (p < 0.001) \). Also, the instruction in text significantly differed from the instruction in a film \( (p < 0.001) \). There was no significant difference between the instruction in a film clip and in a picture \( (p = 0.35) \). There was no significant interaction between difficulty degree and information modality.

**Amount of Correctly Executed Exercises** A main effect of difficulty degree on the performance was found \( F(1,27) = 11.76, p < 0.005 \). As can be seen in Table I, the participants executed on average 9.1 easy exercises correctly, as opposed to 8.3 difficult ones. There was also a main effect of information modality: the participants who watched the film clip executed the most movements correctly \( F(2,27) = 11.68, p < 0.001 \). A two-way interaction between difficulty degree and information modality was found \( F(2,27) = 3.62, p < 0.05 \). This interaction effect can be explained as follows: For the instruction in text \( F(1,9) = 1.22, p = 0.30 \) and for the information in a film clip \( F(1,9) = 1.00, p = 0.34 \) no significant differences were found in the amount of correctly executed easy and difficult exercises. However, for the instruction in a picture \( F(1,9) = 12.52, p < 0.01 \), a significant difference was found in the amount of correctly executed easy and difficult exercises.

### TABLE I

<table>
<thead>
<tr>
<th>Factor</th>
<th>Difficulty</th>
<th>Text (sd)</th>
<th>Picture (sd)</th>
<th>Film (sd)</th>
<th>Average (sd)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Learning time</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Easy</td>
<td>96.30</td>
<td>(45.04)</td>
<td>55.10</td>
<td>(17.51)</td>
<td>76.10</td>
</tr>
<tr>
<td>Difficult</td>
<td>108.10</td>
<td>(40.64)</td>
<td>69.80</td>
<td>(29.04)</td>
<td>91.20</td>
</tr>
<tr>
<td>Practicing</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Easy</td>
<td>4.70</td>
<td>(4.79)</td>
<td>1.50</td>
<td>(1.84)</td>
<td>0.00</td>
</tr>
<tr>
<td>Difficult</td>
<td>5.30</td>
<td>(4.62)</td>
<td>2.20</td>
<td>(2.30)</td>
<td>0.40</td>
</tr>
<tr>
<td>Execution time</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Easy</td>
<td>164.20</td>
<td>(48.86)</td>
<td>53.80</td>
<td>(15.48)</td>
<td>77.90</td>
</tr>
<tr>
<td>Difficult</td>
<td>196.10</td>
<td>(71.28)</td>
<td>77.00</td>
<td>(27.95)</td>
<td>99.60</td>
</tr>
<tr>
<td>Correctly executed exercises</td>
<td>Easy</td>
<td>8.40</td>
<td>(0.70)</td>
<td>9.20</td>
<td>(0.63)</td>
</tr>
<tr>
<td>Difficult</td>
<td>7.90</td>
<td>(1.10)</td>
<td>7.60</td>
<td>(1.35)</td>
<td>9.40</td>
</tr>
</tbody>
</table>
The participants in the picture condition correctly executed fewer of the difficult RSI exercises than the easy exercises (respectively, 9.2 versus 7.6).

**Subjective Satisfaction** The internal consistency on the four items sets in the questionnaire was measured using Cronbach’s alpha. Following standard practice, we qualify alpha as adequate if its value was higher than 0.70. For the structure of the website the alpha was 0.78, and for the content of the website it was 0.56. In Table II, we report separately on the items concerning the content of the website. The alpha for the comprehensibility of the exercises was 0.82, and 0.83 for their attractiveness.

<table>
<thead>
<tr>
<th>Satisfaction with</th>
<th>Text (sd)</th>
<th>Picture (sd)</th>
<th>Film (sd)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Structure</td>
<td>2.25 (0.99)</td>
<td>1.85 (0.58)</td>
<td>2.23 (1.11)</td>
</tr>
<tr>
<td>Content (general)</td>
<td>4.30 (1.95)</td>
<td>3.40 (1.51)</td>
<td>4.20 (1.40)</td>
</tr>
<tr>
<td>Clarity</td>
<td>2.50 (1.18)</td>
<td>2.70 (1.49)</td>
<td>2.70 (1.06)</td>
</tr>
<tr>
<td>Comprehensibility</td>
<td>2.90 (1.29)</td>
<td>3.30 (1.77)</td>
<td>2.30 (1.43)</td>
</tr>
<tr>
<td>Intelligibility</td>
<td>3.70 (1.70)</td>
<td>3.00 (2.00)</td>
<td>2.40 (1.42)</td>
</tr>
<tr>
<td>Exercises</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Comprehensibility</td>
<td>3.80 (1.05)</td>
<td>3.93 (1.32)</td>
<td>2.80 (1.44)</td>
</tr>
<tr>
<td>Attractiveness</td>
<td>3.57 (1.08)</td>
<td>4.07 (1.19)</td>
<td>3.50 (0.93)</td>
</tr>
</tbody>
</table>

Table II gives an overview of the results of the subjective satisfaction questionnaire. Information modality had no effect on the subjective satisfaction regarding the website and the exercises. No effects were found between the three conditions for the website ($F < 1$) and for the exercises ($F < 1$).

**CONCLUSION AND DISCUSSION**

In this paper, we investigated the learnability of RSI prevention exercises in different presentation formats. What effective ways of learning such exercises is an important research question with clear practical ramifications, in view of the growing awareness of the importance of RSI prevention and the bewildering array of presentation formats for RSI prevention exercises currently being employed on the internet and in RSI prevention software.

An experiment was performed looking at the effect of offering RSI prevention exercises in three different formats (film clips, pictures, or text) focusing on learning time, the amount of practicing, execution time, learning results, and subjective satisfaction. To model variation within a domain, 20 RSI prevention exercises were selected from different sources in such a way that 10 exercises were motorically easy (symmetric and simple) and 10 exercises were more difficult (asymmetric or complex). A pre-test with nine participants revealed that there was a strong connection between objective and perceived difficulty degree. As a manipulation check, we tested whether easy exercises were “easier” to perform than the assumed difficult ones. The results showed that irrespective of presentation format, the easy exercises were associated with shorter learning times, less practicing, shorter execution times, and less execution errors than the difficult exercises. It is worth repeating that the summed length (in terms of frames and number of words) of the 10 easy exercises was exactly the same as the summed length of the 10 difficult ones, so that these effects can only be attributed to the differences in complexity (intrinsic cognitive load).

Of the three presentation formats under investigation, text was expected to impose the highest load. Overall, text indeed led to the longest learning times, which can in part be ascribed to the fact that it takes more mental effort to read a text than to look at a picture or watch a film clip. Furthermore, during the learning phase, people not only read the text, but also engage in a substantial amount of practicing, which takes time as well. People in the text condition by far did the most practicing, which can in part be ascribed to the fact that it takes more mental effort to read a text than to look at a picture or watch a film clip. Furthermore, during the learning phase, people not only read the text, but also engage in a substantial amount of practicing, which takes time as well.

Of course, people in the text condition by far did the most practicing, which can in part be ascribed to the fact that it takes more mental effort to read a text than to look at a picture or watch a film clip. Furthermore, during the learning phase, people not only read the text, but also engage in a substantial amount of practicing, which takes time as well.

The Indexical Hypothesis [50]–[52]: To foster understanding, participants “translated” the textual instructions into actual movements during learning for many exercises. Participants must thus engage in fairly deep processing of the textual instructions, but arguably this is to some extent a form of germane load. The deep processing and practicing appear to be beneficial for learning. Contrary to what one might expect, the relatively long learning phase does not lead to shorter execution times. Still, it does lead to a surprisingly good learning performance. For the easy exercises, participants
in the text condition made a few more errors than participants in the other two conditions, but for the difficult exercises, performance was even slightly better than for pictures, which suggests that the germane activities pay off.

Pictures were expected to impose the lowest load. The average learning times were indeed lowest in the picture condition, as were the average execution times. The pictures did not offer a complete model of the exercise, because only the stroke of the movement was depicted, with arrows indicating motion where applicable. The expectation was that people generally would be able to derive the complete exercise on the basis of this information. This might explain why a moderate amount of practicing took place in this condition (more than for film clips). For easy exercises, learning from pictures led to a good performance. In fact, participants made as few errors for these exercises as participants in the film clip condition. But the performance dropped for the difficult exercises, where as many errors were made as in the text condition.

Concerning the load of film clips, the following two contrasting hypotheses were mentioned: They might induce load because a learner has to process the continuously changing images, but they might also reduce load, freeing the learner by simply presenting a complete, physical model of the task to be carried out. It was found that film clips led to medium length learning times (between picture and text). In part, this can be attributed to the fact that watching a clip takes a fixed length. But it is interesting to see that difficult exercises required longer learning times than easy ones, even though they were of the same average duration (and it is not the case that learners practiced difficult exercises more often), which was probably due to the higher average intrinsic load of the difficult tasks. There was virtually no practicing in the film clip condition, as expected, since the clips offer an informationally complete model of the task. Contrary to expectation, the execution times were not the shortest, which suggests that participants still had to engage in cognitive activity during the execution phase. Film clips did lead to a consistently high learning performance, both for easy and for difficult exercises. Hence, despite practicing during learning, learners do construct the necessary scheme based on germane cognitive processes.

It is interesting to observe that none of the presentation formats appeared to be superior on all dimensions of interest; each had some disadvantage (less efficient for learning, relatively many errors, etc.). In view of this, it is perhaps not surprising that the subjective satisfaction questionnaire did not reveal any significant differences. The fact that no single modality outperformed the others on all dimensions implies that there is no single, straightforward design recommendation on how to present information (e.g., in the context of a medical QA system). The goal of the information presentation influences the type of presentation. For example, if the amount of practicing is considered to be the most important factor, the RSI exercises are best described in text. However, if quick learning is most important, the RSI exercises are best illustrated with a picture. Finally, when the goal of the information presentation is a good overall execution, the RSI exercises are best visualized with a film clip.

RSI prevention exercises offer a new and ecologically valid learning situation with a number of interesting properties. These exercises are quite brief and arguably relatively easy to learn. A downside of this choice is that, with respect to learning performance (number of errors), there may be a ceiling effect, in that easy exercises are learned very well for all three modalities. It would be interesting to redo the experiment with more complex RSI related tasks, and see whether this change would lead to more differentiation between the different modalities where errors are concerned.

While we did our best to make sure that the exercises in the three conditions offered comparable information, following the recommendations of Tversky et al. [11], it turned out that this was not always straightforward. While a static picture combined with an arrow indicating direction or motion can be very informative, it does not make the entire intended movement explicit as a dynamic picture does. In the former case, but not in the latter case, the learner has to infer the full movement, which may lead to errors, especially for the difficult exercises as we have seen. Still, it is interesting to see that the efficiency of pictures (learning and execution times) is higher than that of the other two modalities, and leads to nearly optimal results for easy exercises. This finding indicates that a particularly efficient method for illustrating more complex exercises might be via a series of pictures depicting key stages of the procedure. One would expect that this could lead to both high efficiency and good learning results, for easy as well as for difficult exercises. It would also be interesting to study the effect of series of pictures versus single
pictures, or film clips in multimodal instructions (e.g., the use of visuals plus speech). We intend to explore these possibilities in future research.

In a somewhat similar vein, we found that certain RSI exercises seem to be inherently easier to represent than others, and especially that this ease-of-representation may vary across different presentation formats. Some movements can be concisely described in language because the entire movement has been “coded” in a fixed expression (e.g., make fists), whereas other movements can be rather cumbersome to describe. Also, expressing how a particular movement “feels” (e.g., spread your fingers until a mild stretch between the fingers is felt) is obviously easier in language than in static or dynamic visuals. For such exercises, a textual presentation might have an added value over other presentation formats. It would be interesting to systematically vary the presence or absence of linguistic short cuts (describing complex movements in a few words) and investigate how this variance influences the effectiveness of textual instructions.

APPENDIX

SUBJECTIVE SATISFACTION QUESTIONNAIRE
(ENGLISH TRANSLATION OF DUTCH ORIGINALS)

(1) The structure of the website is: (a) Orderly/Disorderly (b) Clear/Unclear (c) Easy to access/Not easy to access (d) Easy to see through/Not easy to see through
(2) The content of the website is: (a) Informative/Uninformative (b) Clear/Unclear (c) Comprehensible/Incomprehensible (d) Understandable/Unintelligible
(3) The comprehensibility of the exercises is: (a) Easy/Difficult (b) Simple/Hard (c) Clear/Unclear (d) Unambiguous/Ambiguous
(4) The attractiveness of the exercises is: (a) Varied/Unvaried (b) Interesting/Uninteresting (c) Appealing/ Unappealing (d) Fascinating/ Tiresome

ACKNOWLEDGMENT


The current research is part of the Interactive Multimodal Information eXtraction (IMIX) project, which is funded by the Netherlands Organisation for Scientific Research (NWO). Thanks are due to P. Barkhuysen and L. van der Laar for their help in setting up the experiment, and to F. Maes and N. Ummelen for their useful comments on an earlier version of this paper. Moreover, we would like to thank the two anonymous reviewers for their suggestions.

REFERENCES


Charlotte van Hooijdonk is a Ph.D. student at Tilburg University, The Netherlands. Her Ph.D. research focuses on how users process and evaluate multimodal information presentations using different evaluation methods. She has been a member of the IEEE Professional Communication Society since 2005.

Emiel Krahmer is a Full Professor at Tilburg University, The Netherlands. His research interests include multimodal and multisensory information processing, developing and evaluating language technology applications, and the analysis of nonverbal communication of real and virtual humans.
Information Modalities for Procedural Instructions: The Influence of Text, Pictures, and Film Clips on Learning and Executing RSI Exercises

CHARLOTTE VAN HOOIJDONK AND EMIEL KRAHMER

Abstract—Much of the empirical research on the effectiveness of different instructional designs has focused on declarative tasks, where a learner acquires knowledge about a certain topic. It is unclear to what extent findings for learning declarative tasks (which are not consistent on all aspects) carry over to learning procedural tasks, where a learner acquires a certain skill. In this paper, we describe an experiment studying a specific kind of procedural instructions, namely exercises for the prevention of repetitive strain injury (RSI), taking information modality (text versus picture versus film clip) and difficulty degree of the exercises (easy versus difficult) into account. In the experiment, participants had to learn RSI exercises and were asked to execute them. The results showed that an instruction in a picture lead to the shortest learning times followed by an instruction in a film clip. An instruction in text led to the longest learning times. For the amount of practicing the exercises during the learning phase, it was found that the participants in the film clip condition hardly engaged in practicing the exercises during the learning phase. The participants in the picture condition engaged in a moderate amount of practicing of the exercises during the learning phase. The participants in the text condition engaged in the most practicing during the learning phase. The results concerning the execution times showed that an instruction in a picture led to the lowest execution times followed by an instruction in a film clip. The instruction in text led to longest execution times. Finally, for the amount of correctly executed exercises, it was found that learning from a film clip led to the highest learning performance, both for easy and for difficult exercises. Learning from an instruction in text led to a fairly good learning performance, both for easy and difficult exercises. Learning from a picture led to a good learning performance for the easy exercises, but the performance dropped for the difficult exercises.

Index Terms—Document design, information modality, procedural learning.

The emergence of computer-based learning has led to new possibilities for presenting instructions and to a renewed interest in the effects of different information modalities. Instructions can be given in plain text, but also in the form of static visuals (e.g., diagrams, pictures) or in dynamic visuals (e.g., animations, film clips). Naturally, this raises the question which (combinations of) modalities are best for which learning task. This question has been addressed in a large number of recent studies (e.g., [1]–[6], among many others).

Each of the aforementioned presentation modes has its own unique characteristics, and its own advantages and disadvantages from an instructional perspective. Language is the basic form of human communication, and one of its main strengths is that it is expressive and explicit (both for concrete and for abstract subject matter). An additional advantage of its written form (as opposed to the spoken variant) is that it is not transient: written sentences remain visible on paper and can be re-read if desired, while spoken sentences are “gone” once they have been uttered. But, reading a text requires considerable skill and effort, and, moreover, text primarily facilitates linear information processing.

A difference between text and pictures is their symbol system. The symbol system used for text is abstract/linguistic, while the one used for pictures is sensory. This affects the type of information that they can convey. For example, pictures can directly communicate perceptual information (e.g., spatial orientation and location). In other words, pictures have a lower “articulatory distance” [7]. Moreover, pictures are not constrained by the linear structure of text, and are therefore arguably more efficient at representing nonlinear relations among objects. In text these relationships often remain implicit [8].

The focus of many recent studies has been on the instruction effects of dynamic visuals, presumably because such instructions only recently have become a real possibility due to the increased computing power of standard multimedia PCs. A number of reasons have been suggested as to why an advantage of dynamic visuals over other presentation possibilities such as text and static visuals might be expected. For instance, it has
been argued that dynamic visuals are beneficial for learning since they offer a “complete model” of a learning task (e.g., [1], [9]); in other words, they are “informationally complete” [10, p. 249]. When static visuals or text are used, learners themselves will have to construct a mental representation of the temporal aspects in the instruction. It has been argued that dynamic illustrations offer a better representation of these temporal aspects, thereby reducing the level of abstraction and supporting a deeper level of understanding than static visuals would do [9]. Arguably, this should facilitate learning since it would reduce learning times, require less practice, and result in fewer errors.

A substantial number of studies have tried to demonstrate this presumed learning benefit, but with mixed results (e.g., [1], [5], [11], and [12]). Two groups of explanations for these mixed results have been offered: one group of explanations stresses the disadvantages of dynamic visuals; another group points to methodological problems in these studies. The first explanation claims that dynamic visuals have a fixed duration which viewers simply have to watch. This set time period may lead to inherently longer learning times [11]. In a similar vein, Lewalter [1] suggested that dynamic visuals may take more time to process than other presentation modes; the information in dynamic visuals changes continuously, and as a result learners could be overwhelmed with the amount of information. Alternatively, it has been suggested that dynamic visuals are processed somewhat superficially; they require little cognitive effort, as they can be watched rather passively (e.g., [10] and [13]). The second explanation, by contrast, argues that the mixed findings so far are primarily due to methodological problems in the various experimental studies. For instance, to be really beneficial, dynamics should have some added value [14], which is not always the case. When it comes to temporal sequencing or indicating direction of motion, it has been claimed that arrows in static visuals may be just as effective as dynamic visuals (e.g., [15]). Tversky, Morrison, and Betrancourt [11] point out that in some comparative studies there is a lack of equivalence between dynamic and static visuals in content or procedures (e.g., because the dynamic visuals convey more information or involve interactivity). Some researchers have argued that even when dynamic visuals do not lead to improved learning, they are more attractive and motivating than other instruction forms and should be preferred for that reason (e.g., [11], [16], and [17]). However, this “subjective satisfaction” [18, p. 33] of various instruction modes is often not addressed, and when it is the results are equivocal (e.g., [19]). A further complicating factor is that the learning domain itself might also have an influence on the relative benefits of information modalities [14]. It seems reasonable to assume that different learning domains benefit from different kinds of instructions (cf., also [20]). In the field of learning psychology, many studies have shown that dynamic visuals are used as learning instructions for descriptive, “scientific” explanations (often describing causes and effects), such as mechanical and electronic systems (e.g., brakes, pumps, generators [21], [22]; electronic circuit systems [23]; mathematical concepts (e.g., algebra [24]); or complex natural phenomena (e.g., electricity [25]; lightning [26]; gravitational lensing [1]; meteorological changes [2]). Arguably, some of these scientific learning domains lend themselves better for dynamic visualization than others. Moreover, what holds for descriptive tasks (such as those above), may not apply to procedural ones, such as bandaging a hand [5], folding origami models [27], or tying nautical knots [28]. These procedural tasks differ from descriptive ones in a number of respects. Not only is the nature of the task different (procedures are inherently more linear—one step following another), but the learning goal is also different (the focus is not only on understanding, but also on acquiring certain capabilities or skills) [29]. One of the factors that might affect the processing of procedural information is the design format [29]. Procedural information is often presented in a text and/or picture format (e.g., route directions, maintenance instructions, assembly instructions). While pictures may help users to form a mental model of the procedure [30], [31], the combination of text plus picture may not always be helpful. For example, a user’s attention has to switch between the information presented in text and picture, leading to the so-called split attention effect [32]. Therefore, document designers have to understand the strengths and limitations of both text and pictures when designing procedural information [8]. Both the type of picture (e.g., overview versus partial view [33], line drawing versus photo [5], object-centered versus body-centered [34], etc.) as well as the type of textual instruction (e.g., user-centered versus object-centered versus environment-centered [35]) may influence a user’s performance of the procedure. To further complicate the situation, it may well be that besides variation between domains there is also variation within domains. Arguably, some learning tasks in a given domain are easier than others, and this
may have an influence on the relative contribution of various instruction formats for those tasks.

In this paper, we report on a study of the effects of task difficulty and information modality (comparing dynamic visuals with static visuals and text) on learning a specific class of procedural tasks, namely exercises aiming at the prevention of repetitive strain injury (RSI). RSI is a general term for disorders that are caused by repetitious use of hands, arms, and shoulders, often as a result of prolonged computer terminal work [36]. Recently, the term CANS (short for complaints of arm, neck and/or shoulders) was introduced as an alternative term, but here we will stick to the more familiar RSI terminology. It is generally assumed that taking regular breaks during computer work in combination with exercises is beneficial for the prevention of RSI (e.g., [37]–[39]). RSI exercises offer a new and understudied learning domain, with various interesting properties. Generally, these exercises involve little or no abstraction, do not consist of a complicated sequence of actions, and are relatively short. Moreover, the exercises are highly recognizable (almost everybody has two hands). It is interesting to observe that current RSI information websites offer a large variety of prevention exercises, in many different presentation formats (see Fig. 1 for a representative selection). These differences raise the natural question of what is the effectiveness of the various presentation formats. Note also that there is substantial variation in the difficulty level of existing RSI exercises, so that the other factor of interest (variation in task difficulty) can be modelled in a fairly straightforward way in this domain.

The potential influence of both presentation modality and task difficulty on learning performance can be motivated from COGNITIVE LOAD THEORY (e.g., [40]–[43]), a theory which aims to develop optimal instructional designs while considering the limitations of the human mind. Cognitive load theory builds on the broadly accepted assumption that the mind combines a short-term (or working) memory of very limited capacity (where all conscious activity and processing of information occurs [44], [45]) and a long-term memory with a virtually unlimited capacity [42]. According to cognitive load theory, learning amounts to the construction of new (or modification of existing) schemata [46], which are subsequently stored in long-term memory. Since the capacity of working memory is severely limited, the cognitive load of learners should be kept at a minimum during learning. In the current version of the theory [42], [43], three kinds of load are distinguished: INTRINSIC load, caused by the structure and intrinsic nature of the learning task; EXTRANEOUS load, caused by the manner of presentation and its influence.
on working memory; and germane load, caused by learners’ efforts to process and comprehend learning material. The sum of these three kinds of load should not exceed working memory capacity in order to avoid cognitive overload. As argued above, the intrinsic load may vary both between and within task domains. Since the intrinsic load of a particular learning task is fixed, instruction design can only influence the extraneous and the germane load, and, obviously, when the intrinsic load of a particular task is high, there is relatively less room for the extraneous and the germane load.

Germane load is a “positive” form of load, since it encourages learners to engage in cognitive processing that may lead to improved schemata construction [42], [47]. This suggests that optimal instructions are those which minimize extraneous and maximize germane load, while simultaneously avoiding overload. A complicating factor is that the distinction between extraneous and germane cognitive load, although intuitively plausible, cannot reliably be measured [47]. In general, measuring cognitive load in realistic learning situations is not straightforward [48]. While some recent attempts have been made to measure cognitive load directly (e.g., the dual task approach advocated by Brunken, Plass, and Leutner [49]), many studies rely on indirect measures such as behavioural or learning outcome measures (e.g., [47] and [49]). Although these measures only relate to cognitive load indirectly, they do assess learning behaviour directly, which is sufficient for our current purposes. In this study, we concentrate on learning times, amount of practicing during learning, execution times, and execution errors.

Arguably, the different information modalities of interest (text, picture, film clip) have different loading potentials, which may influence performance on one or more of these measures. Of all three modalities, it seems likely that text imposes the highest load: It requires more mental effort to read a text than to watch a picture or a film clip, hence it seems likely that learning from text takes longer than learning from pictures or film clips. Potentially, an additional complication for learning RSI prevention exercises from text is that schemata construction may be more involved than for pictures and film clips. In the text case, learners have to “translate” the textual instructions to manual gestures. Notice that this is a concrete instance of Glenberg’s Indexical Hypothesis, stating that readers associate words and phrases with objects and actions in “the real world,” which should facilitate understanding (e.g., [50]–[52]). An instruction in the form of a film clip and (probably to a lesser extent) a picture depicts the hand and arm movements the learner should make, whereas the learner has to infer these gestures when they are presented in text. Hence, it is hypothesized that learners will practice more often while learning from text than while learning from visual presentations. An interesting question is whether the load imposed by text is only extraneous or also partly germane. It might be that the extra effort required to learn from text leads to good learning results (short execution times, few execution errors), especially when the intrinsic load is low (so that overload can be avoided).

Pictures arguably impose the lowest load of the three presentation modalities; viewing a static picture requires little mental effort. Since a picture captures the essential information of a procedure, it is to be expected that learning times for pictures are relatively short. However, due to their static nature, pictures offer little information about the temporal structure of a procedure, and for more complicated exercises (i.e., exercises with a higher intrinsic load) this may hamper schemata construction and might result in suboptimal execution performance.

To what extent film clips impose cognitive load is uncertain. On the one hand, it can be argued that they may induce load, since the film clips continuously change and learners have to process this information, which reduces the cognitive resources available for germane load. But they may also lessen cognitive load by relieving the learner in the translation process, which is required when reading text. The main strength of film clips might well be their “informational completeness”—learners do not have to infer the exact sequence of movements from text or from a single snapshot because the entire sequence of actions is visualized, which arguably facilitates schemata construction. This “completeness” suggests that learners will not practice much during learning. Whether this will also result in few execution errors is not obvious; it may be that the large amount of external support relieves learners of cognitive load demands that they would be able to fulfil, but which might prevent them from performing beneficial cognitive actions for learning.

The findings of this study will not only be theoretically relevant, but may also inform more practical work on multimodal information presentation. An example of such work is a medical question-answering (QA) system, which is currently developed within the IMOGEN (Interactive
Multimodal Output GENeration) project that is embedded in the IMIX research programme. IMIX (Interactive Multimodal Information eXtraction) is a research programme in the field of Dutch speech and language technology. A QA system is an automatic system that can answer a user’s question posed in natural languages (e.g., What does RSI stand for?) with an answer formulated in natural language (e.g., RSI stands for Repetitive Strain Injury). Today, QA systems are not only expected to give answers to simple questions, like What does RSI stand for? but also to more complex questions, like How should I organize my workspace in order to prevent RSI? or What is a good exercise to prevent RSI in my hands? The answers to these questions might be more informative if they contained a picture or an animation, rather than an extensive textual answer [53]. In the IMOGEN project, different aspects of multimodal information presentation are studied in order to improve the output quality of QA systems.

To find out what the actual strengths and weaknesses of the various information modalities are, we describe an experiment in which participants have to learn and execute 20 RSI prevention exercises in two degrees of difficulty. We measure the following features: the influence of presenting an instruction in text; picture or film clips on learning times; the amount of practicing during learning; execution times; and number of execution errors. Besides these objective measures, participants are also asked for their subjective satisfaction.

**METHOD**

**Participants** Participants were 30 young adults, between 18 and 30 years of age. Of the participants, 15 were male and 15 were female.

**Design** The experiment had a 3 (information modality) × 2 (difficulty degree) factorial design, with information modality (dynamic visual [film clip], static visual [picture], text) used as a between-participants variable, and difficulty degree (easy, difficult) as a within-participants variable, and with learning times, amount of practicing during learning; execution times; and number of execution errors. Besides these objective measures, participants are also asked for their subjective satisfaction.

**Stimuli** A total of 20 RSI prevention exercises were chosen from websites on RSI prevention and RSI injury prevention software (e.g., http://web.mit.edu/atic/www/disabilities/rsi/exercises.html and http://www.workpace.com). The chosen exercises were all exercises to prevent RSI in hands and arms. Of the 20 exercises, 10 were assumed to be easy to perform, and 10 were assumed to be difficult. The criterion for a difficult exercise was that it should be either a complex symmetrical movement or an asymmetrical movement. Complex symmetrical movements were classified as movements consisting of at least three sequential atomic movements, in the course of which both arms and hands make the same movements. Asymmetrical movements were classified as movements in which the participant should make a different movement with each arm or hand. Easy exercises were neither asymmetrical nor complex movements. Fig. 2 shows representative examples of an easy and a difficult exercise.

**Fig. 2.** Two typical exercises: one easy and one difficult.

Note that this operationalization of easy and difficult tasks is based on the relative complexity of the sequence of motoric movements. To find out to what extent this objective criterion coincided with subjective perception of task difficulty, a pre-test was carried out in which nine participants were asked to classify the 20 exercises (presented in text plus picture format, and in random order). They were instructed to make two piles, the first consisting of the 10 exercises they considered easiest to perform, and the second pile consisting of the 10 exercises they considered more difficult to perform. It turned out that of the 10 exercises classified as easy, 7 were indeed perceived as easy.
by the vast majority (>75%) of the participants, while the remaining 3 were perceived as difficult by a majority of the participants (presumably because these consisted of gestures that are motorically simple but not commonly used and hence with which participants may have been less familiar). Of the 10 exercises classified as difficult, 9 were indeed perceived as such by the vast majority (>75%), while the remaining 1 was perceived as easy by most participants (interestingly, this was an exercise that was motorically complex but familiar to most participants). In sum, for 16 of the 20 exercises, there was a clear and consistent correlation between the objective and the perceived difficulty degree.

The 20 RSI prevention exercises were presented in three different formats. In the text condition, the exercises were explained to the participants in a purely textual format. The total amount of words did not differ between the 10 easy and 10 difficult exercises: both counted 268 words in total. Thus, the average length was almost 27 words per exercise. Since some exercises are a few words shorter and others a few words longer, we will only report on mean total averages for the 10 easy and the 10 difficult exercises in each condition. Because natural language is often ambiguous, the text exercises were checked for their comprehensibility in a second pre-test with three participants (different from those in the first pre-test). It turned out that a few exercises led to misunderstandings and so these exercises were reformulated. The new versions were presented to the same three participants, and they agreed that the reformulations solved the misunderstandings.

In the picture condition, each of the 20 exercises was displayed in a single photograph. Digital camera pictures were taken of a female who did the exercises. She wore black clothing and the movements were shot against a black background so that only her hands were visible in the picture. The photo depicted the STROKE of the movement, with the phase of the movement as it unfolds in time containing the "semantic content" of the movement [54]. To indicate the direction of movement, arrows were added to the pictures of 9 difficult and 2 easy exercises. The size of the pictures was 1536 × 1014 pixels.

For the film clip condition, the same female in an identical surrounding was filmed with a digital film camera (25 frames per second). The total amount of frames did not differ between easy and difficult exercises: both counted 1097 frames (average film length was thus 4.39 seconds). Again, since some film clips are somewhat shorter, and others somewhat longer, we will report on mean total averages for the 10 easy and the 10 difficult exercises.

The exercises were presented to the participants via a website: one website for each condition. The experiment was run on a multimedia PC, with a 17-inch color monitor. In all three conditions, the presentations of the RSI exercises appeared at the centre of the display. On the video website, the movements were shown in a film clip with a start and a stop button and a slider. The participants had the opportunity to view the film as often as they desired, but not much use was made of this option. Exercises were presented in one of two random orders to control for possible learning effects.

**Procedure** The experiment consisted of three parts: a practice session, the central part in which the 20 RSI exercises were presented, and a questionnaire. The participants completed each part individually. Each participant was invited to an experimental laboratory, where he or she took a seat behind the computer. Participants were told that they would receive 20 exercises which they had to learn and perform to assess to what extent they suffered from RSI. In addition, they were led to believe that their hand and arm movements were filmed because a doctor would later look at the recordings of the participants to determine to what extent they suffered from RSI. The procedure of the experiment is depicted in Fig. 3.

After the participants had read the instructions, they could click on the hyperlink *start*, and the first trial exercise would appear. Depending on their experimental condition, the participants read or viewed the trial exercise until they thought that they could properly execute the exercise. Subsequently, they clicked the hyperlink *next* at the bottom of the page. A new webpage appeared with the text "execute trial exercise" at the centre of the page. During the execution of an exercise, participants could not see the instruction. When they had executed the exercise, they clicked the hyperlink *next exercise* at the bottom of the page. After completing the second trial exercise, participants were asked if they had any questions regarding the experimental procedure. If not, the participants could start the actual experiment, proceeding in the same way as they did in the trial session. During the experiment, there was no further interaction between participants and the experiment leader.
After the participants completed the 20 experimental exercises, they received a questionnaire. In this questionnaire, the subjective satisfaction regarding the content and structure of the website as well as the comprehensibility and the attractiveness of the exercises were measured. The questionnaire consisted of 16 bi-polar seven-point semantic differentials addressing structure and content of the websites, as well as comprehensibility and attractiveness of the exercises (see Appendix A). The presentation order of the adjectives was randomized. For processing, the positive adjectives were mapped to one (very positive) and the negative adjectives were mapped to seven (very negative).

The participants were debriefed at the end of the experiment.

**Data Processing** The following data was collected: learning times; number of practiced exercises during the learning time; execution times; number of correctly executed exercises; and the results of the questionnaire.

The time it took the participants to learn and execute the exercises was computed from the data of the log program ProxyPlus (http://www.proxyplus.com). This program registered the times associated with participants’ mouse clicks during the experiment. The time period between clicking the hyperlink next which preceded the instruction of an exercise and the hyperlink next which followed the instruction of an exercise was defined as the LEARNING TIME (see Fig. 3). The time period between clicking the hyperlink next which followed the instruction of an exercise and the hyperlink next that preceded a new instruction of an exercise was defined as the EXECUTION TIME (see Fig. 3).

A digital video camera was used to record the movements the participants made during the experiment. These video recordings of the participants’ hands and arms were used to assess whether the participants practiced the exercises during the learning period and to assess their performances while executing the RSI exercises. Occasionally, a participant performed an exercise in a correct but not intended way. These cases were counted as correctly executed exercises. One judge did the assessment; scoring was easy, and the few difficult cases that did arise were resolved after discussion.

Tests for significance were performed using a repeated measures analysis of variance (ANOVA), with a significance threshold of 0.05. For post-hoc tests the Bonferroni method was used. The internal consistency of the four item sets of the questionnaire was measured using Cronbach’s alpha.

**RESULTS**

**Learning Times** Table I summarizes the results. First consider the learning time. We found that the difficulty degree had an effect on the amount of time to learn the exercises ($F(1,27) = 37.35$, $p < 0.001$). Overall, the participants needed more time to learn the difficult exercises than the easy ones. Also, the information presentation had
an effect on the learning time \(F(2, 27) = 4.53, p < 0.025\). Participants in the picture condition required the shortest learning times, participants in the text condition had the longest learning times, with learning times from film clips falling in between those two. Post-hoc tests indicated that the instruction in text differed significantly from the instruction in a picture \(p < 0.025\). There was no significant difference between the instruction in text and instruction in a film clip \(p = 0.35\). Also, the instruction in a picture did not differ significantly from the instruction in a film clip \(p = 0.25\). No significant interactions between difficulty degree and information modality were found.

Amount of Practicing During the Learning Period

Table I also reveals that the participants practiced the difficult exercises more often than the easy ones during the learning period, and this difference was found to be statistically significant \(F(1, 27) = 9.00, p < 0.01\). Also, information presentation had an effect on the amount of practicing during the learning period \(F(2, 27) = 6.76, p < 0.005\). In the film clip condition, participants almost never practiced; in the picture condition, they practiced for about a fifth of the exercises, while in the text condition, participants practiced about half of the exercises. Post-hoc tests showed that the difference between the instruction in text and the instruction in a picture was not significant \(p = 0.06\). Text differed significantly from the instruction in a film clip \(p < 0.005\). There was no significant difference between the instruction in a picture and a film clip \(p = 0.43\), presumably because of the relatively high standard deviation. No significant interaction effects were found.

Execution Times

The difficulty degree had a main effect on the amount of time needed to perform the exercises \(F(1, 27) = 20.84, p < 0.001\). The participants required more time to execute the difficult exercises than the easy ones. There was also a main effect of information modality on the execution times \(F(2, 27) = 26.78, p < 0.001\). Participants in the text condition had much longer execution times than those in the picture and film clip conditions. Participants in the picture condition needed somewhat less time for execution than the participants in the film clip condition, but the differences are relatively small and the standard deviations are relatively high. Post-hoc tests indicated that there was a significant difference between the instruction in text and the instruction in a picture \(p < 0.001\). Also, the instruction in text significantly differed from the instruction in a film \(p < 0.001\). There was no significant difference between the instruction in a film clip and in a picture \(p = 0.35\). There was no significant interaction between difficulty degree and information modality.

Amount of Correctly Executed Exercises

A main effect of difficulty degree on the performance was found \(F(1, 27) = 11.76, p < 0.005\). As can be seen in Table I, the participants executed on average 9.1 easy exercises correctly, as opposed to 8.3 difficult ones. There was also a main effect of information modality: the participants who watched the film clip executed the most movements correctly \(F(2, 27) = 11.68, p < 0.001\). A two-way interaction between difficulty degree and information modality was found \(F(2, 27) = 3.62, p < 0.05\). This interaction effect can be explained as follows: For the instruction in text \(F(1, 9) = 1.22, p = 0.30\) and for the instruction in a film clip \(F(1, 9) = 1.00, p = 0.34\) no significant differences were found in the amount of correctly executed easy and difficult exercises. However, for the instruction in a picture \(F(1, 9) = 12.52, p < 0.01\), a significant difference was found in the amount of correctly executed easy and difficult exercises.

### Table I

<table>
<thead>
<tr>
<th>Factor</th>
<th>Difficulty</th>
<th>Text (sd)</th>
<th>Picture (sd)</th>
<th>Film (sd)</th>
<th>Average (sd)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Learning time</strong></td>
<td>Easy</td>
<td>96.30 (45.04)</td>
<td>55.10 (17.51)</td>
<td>76.10 (17.07)</td>
<td>75.83 (33.28)</td>
</tr>
<tr>
<td></td>
<td>Difficult</td>
<td>108.10 (40.64)</td>
<td>69.80 (29.04)</td>
<td>91.20 (15.42)</td>
<td>89.70 (33.20)</td>
</tr>
<tr>
<td><strong>Practicing</strong></td>
<td>Easy</td>
<td>4.70 (4.79)</td>
<td>1.50 (1.84)</td>
<td>0.00 (0.00)</td>
<td>2.07 (3.48)</td>
</tr>
<tr>
<td></td>
<td>Difficult</td>
<td>5.30 (4.62)</td>
<td>2.20 (2.30)</td>
<td>0.40 (0.70)</td>
<td>2.63 (3.56)</td>
</tr>
<tr>
<td><strong>Execution time</strong></td>
<td>Easy</td>
<td>164.20 (48.86)</td>
<td>53.80 (15.48)</td>
<td>77.90 (13.42)</td>
<td>98.63 (56.52)</td>
</tr>
<tr>
<td></td>
<td>Difficult</td>
<td>196.10 (71.28)</td>
<td>77.00 (27.95)</td>
<td>99.60 (31.41)</td>
<td>124.23 (69.89)</td>
</tr>
<tr>
<td><strong>Correctly executed exercises</strong></td>
<td>Easy</td>
<td>8.40 (0.70)</td>
<td>9.20 (0.63)</td>
<td>9.60 (0.70)</td>
<td>9.10 (0.84)</td>
</tr>
<tr>
<td></td>
<td>Difficult</td>
<td>7.90 (1.10)</td>
<td>7.60 (1.35)</td>
<td>9.40 (0.70)</td>
<td>8.27 (1.44)</td>
</tr>
</tbody>
</table>
The participants in the picture condition correctly executed fewer of the difficult RSI exercises than the easy exercises (respectively, 9.2 versus 7.6).

**Subjective Satisfaction** The internal consistency on the four item sets in the questionnaire was measured using Cronbach’s alpha. Following standard practice, we qualify alpha as adequate if its value was higher than 0.70. For the structure of the website the alpha was 0.78, and for the content of the website it was 0.56. In Table II, we report separately on the items concerning the content of the website. The alpha for the comprehensibility of the exercises was 0.82, and 0.83 for their attractiveness.

**TABLE II**

<table>
<thead>
<tr>
<th>Satisfaction with</th>
<th>Text (sd)</th>
<th>Picture (sd)</th>
<th>Film (sd)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Website</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Structure</td>
<td>2.25 (0.99)</td>
<td>1.85 (0.58)</td>
<td>2.23 (1.11)</td>
</tr>
<tr>
<td>Content (general)</td>
<td>4.30 (1.95)</td>
<td>3.40 (1.51)</td>
<td>4.20 (1.40)</td>
</tr>
<tr>
<td>Clarity</td>
<td>2.50 (1.18)</td>
<td>2.70 (1.49)</td>
<td>2.70 (1.06)</td>
</tr>
<tr>
<td>Comprehensibility</td>
<td>2.90 (1.29)</td>
<td>3.30 (1.77)</td>
<td>2.30 (1.43)</td>
</tr>
<tr>
<td>Intelligibility</td>
<td>3.70 (1.70)</td>
<td>3.00 (2.00)</td>
<td>2.40 (1.42)</td>
</tr>
<tr>
<td><strong>Exercises</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Comprehensibility</td>
<td>3.80 (1.05)</td>
<td>3.93 (1.32)</td>
<td>2.80 (1.44)</td>
</tr>
<tr>
<td>Attractiveness</td>
<td>3.57 (1.08)</td>
<td>4.07 (1.19)</td>
<td>3.50 (0.93)</td>
</tr>
</tbody>
</table>

Table II gives an overview of the results of the subjective satisfaction questionnaire. Information modality had no effect on the subjective satisfaction regarding the website and the exercises. No effects were found between the three conditions for the website ($F < 1$) and for the exercises ($F < 1$).

**CONCLUSION AND DISCUSSION**

In this paper, we investigated the learnability of RSI prevention exercises in different presentation formats. What effective ways of learning such exercises is an important research question with clear practical ramifications, in view of the growing awareness of the importance of RSI prevention and the bewildering array of presentation formats for RSI prevention exercises currently being employed on the internet and in RSI prevention software.

An experiment was performed looking at the effect of offering RSI prevention exercises in three different formats (film clips, pictures, or text) focusing on learning time, the amount of practicing, execution time, learning results, and subjective satisfaction. To model variation within a domain, 20 RSI prevention exercises were selected from different sources in such a way that 10 exercises were motorically easy (symmetric and simple) and 10 exercises were more difficult (asymmetric or complex). A pre-test with nine participants revealed that there was a strong connection between objective and perceived difficulty degree.

As a manipulation check, we tested whether easy exercises were “easier” to perform than the assumed difficult ones. The results showed that irrespective of presentation format, the easy exercises were associated with shorter learning times, less practicing, shorter execution times, and less execution errors than the difficult exercises. It is worth repeating that the summed length (in terms of frames and number of words) of the 10 easy exercises was exactly the same as the summed length of the 10 difficult ones, so that these effects can only be attributed to the differences in complexity (intrinsic cognitive load).

Of the three presentation formats under investigation, text was expected to impose the highest load. Overall, text indeed led to the longest learning times, which can in part be ascribed to the fact that it takes more mental effort to read a text than to look at a picture or watch a film clip. Furthermore, during the learning phase, people not only read the text, but also engage in a substantial amount of practicing, which takes time as well. People in the text condition by far did the most practicing, which is consistent with the Indexical Hypothesis [50]–[52]: To foster understanding, participants “translated” the textual instructions into actual movements during learning for many exercises. Participants must thus engage in fairly deep processing of the textual instructions, but arguably this is to some extent a form of germane load: The deep processing and practicing appear to be beneficial for learning. Contrary to what one might expect, the relatively long learning phase does not lead to shorter execution times. Still, it does lead to a surprisingly good learning performance. For the easy exercises, participants...
in the text condition made a few more errors than participants in the other two conditions, but for the difficult exercises, performance was even slightly better than for pictures, which suggests that the germane activities pay off.

Pictures were expected to impose the lowest load. The average learning times were indeed lowest in the picture condition, as were the average execution times. The pictures did not offer a complete model of the exercise, because only the stroke of the movement was depicted, with arrows indicating motion where applicable. The expectation was that people generally would be able to derive the complete exercise on the basis of this information. This might explain why a moderate amount of practicing took place in this condition (more than for film clips). For easy exercises, learning from pictures led to a good performance. In fact, participants made as few errors for these exercises as participants in the film clip condition. But the performance dropped for the difficult exercises, where as many errors were made as in the text condition.

Concerning the load of film clips, the following two contrasting hypotheses were mentioned: They might induce load because a learner has to process the continuously changing images, but they might also reduce load, freeing the learner by simply presenting a complete, physical model of the task to be carried out. It was found that film clips led to medium length learning times (between picture and text). In part, this can be attributed to the fact that watching a clip takes a fixed length. But it is interesting to see that difficult exercises required longer learning times than easy ones, even though they were of the same average duration (and it is not the case that learners practiced difficult exercises more often), which was probably due to the higher average intrinsic load of the difficult tasks. There was virtually no practicing in the film clip condition, as expected, since the clips offer an informationally complete model of the task. Contrary to expectation, the execution times were not the shortest, which suggests that participants still had to engage in cognitive activity during the execution phase. Film clips did lead to a consistently high learning performance, both for easy and for difficult exercises. Hence, despite practicing during learning, learners do construct the necessary scheme based on germane cognitive processes.

It is interesting to observe that none of the presentation formats appeared to be superior on all dimensions of interest; each had some disadvantage (less efficient for learning, relatively many errors, etc.). In view of this, it is perhaps not surprising that the subjective satisfaction questionnaire did not reveal any significant differences. The fact that no single modality outperformed the others on all dimensions implies that there is no single, straightforward design recommendation on how to present information (e.g., in the context of a medical QA system). The goal of the information presentation influences the type of presentation. For example, if the amount of practicing is considered to be the most important factor, the RSI exercises are best described in text. However, if quick learning is most important, the RSI exercises are best illustrated with a picture. Finally, when the goal of the information presentation is a good overall execution, the RSI exercises are best visualized with a film clip.

RSI prevention exercises offer a new and ecologically valid learning situation with a number of interesting properties. These exercises are quite brief and arguably relatively easy to learn. A downside of this choice is that, with respect to learning performance (number of errors), there may be a ceiling effect, in that easy exercises are learned very well for all three modalities. It would be interesting to redo the experiment with more complex RSI related tasks, and see whether this change would lead to more differentiation between the different modalities where errors are concerned.

While we did our best to make sure that the exercises in the three conditions offered comparable information, following the recommendations of Tversky et al. [11], it turned out that this was not always straightforward. While a static picture combined with an arrow indicating direction or motion can be very informative, it does not make the entire intended movement explicit as a dynamic picture does. In the former case, but not in the latter case, the learner has to infer the full movement, which may lead to errors, especially for the difficult exercises as we have seen. Still, it is interesting to see that the efficiency of pictures (learning and execution times) is higher than that of the other two modalities, and leads to nearly optimal results for easy exercises. This finding indicates that a particularly efficient method for illustrating more complex exercises might be via a series of pictures depicting key stages of the procedure. One would expect that this could lead to both high efficiency and good learning results, for easy as well as for difficult exercises. It would also be interesting to study the effect of series of pictures versus single
pictures, or film clips in multimodal instructions (e.g., the use of visuals plus speech). We intend to explore these possibilities in future research.

In a somewhat similar vein, we found that certain RSI exercises seem to be inherently easier to represent than others, and especially that this ease-of-representation may vary across different presentation formats. Some movements can be concisely described in language because the entire movement has been “coded” in a fixed expression (e.g., make fists), whereas other movements can be rather cumbersome to describe. Also, expressing how a particular movement “feels” (e.g., spread your fingers until a mild stretch between the fingers is felt) is obviously easier in language than in static or dynamic visuals. For such exercises, a textual presentation might have an added value over other presentation formats. It would be interesting to systematically vary the presence or absence of linguistic short cuts (describing complex movements in a few words) and investigate how this variance influences the effectiveness of textual instructions.

APPENDIX

SUBJECTIVE SATISFACTION QUESTIONNAIRE (ENGLISH TRANSLATION OF DUTCH ORIGINALS)

(1) The structure of the website is: (a) Orderly/Disorderly (b) Clear/Unclear (c) Easy to access/Not easy to access (d) Easy to see through/Not easy to see through

(2) The content of the website is: (a) Informative/Uninformative (b) Clear/Unclear (c) Comprehensible/Incomprehensible (d) Understandable/Unintelligible

(3) The comprehensibility of the exercises is: (a) Easy/Difficult (b) Simple/Hard (c) Clear/Unclear (d) Unambiguous/Ambiguous

(4) The attractiveness of the exercises is: (a) Varied/Unvaried (b) Interesting/Uninteresting (c) Appealing/ Unappealing (d) Fascinating/ Tiresome

ACKNOWLEDGMENT


The current research is part of the Interactive Multimodal Information eXtraction (IMIX) project, which is funded by the Netherlands Organisation for Scientific Research (NWO). Thanks are due to P. Barkhuysen and L. van der Laar for their help in setting up the experiment, and to F. Maes and N. Ummelen for their useful comments on an earlier version of this paper. Moreover, we would like to thank the two anonymous reviewers for their suggestions.

REFERENCES


Charlotte van Hooijdonk is a Ph.D. student at Tilburg University, The Netherlands. Her Ph.D. research focuses on how users process and evaluate multimodal information presentations using different evaluation methods. She has been a member of the IEEE Professional Communication Society since 2005.

Emiel Krahmer is a Full Professor at Tilburg University, The Netherlands. His research interests include multimodal and multisensory information processing, developing and evaluating language technology applications, and the analysis of nonverbal communication of real and virtual humans.