

Animated demonstrations vs written instructions for learning procedural tasks: a preliminary investigation

SUSAN PALMITER AND JAY ELKERTON†

Center for Ergonomics, Department of Industrial and Operations Engineering, The University of Michigan, Ann Arbor, MI 48109, USA

PATRICIA BAGGETT

School of Education, The University of Michigan, Ann Arbor, MI 48109, USA

(Received 5 April 1989 and accepted in revised form 20 November 1989)

Animated demonstrations have been created due to the development of direct manipulation interfaces and the need for faster learning, so that users can learn interface procedures by watching. To compare animated demonstrations with written instructions we observed users learning and performing HyperCard™‡ authoring tasks on the Macintosh™ during three performance sessions. In the training session, users were asked either to watch a demonstration or read the procedures needed for the task and then to perform the task. In the later two sessions users were asked to perform tasks identical or similar to the tasks used in the training session. Results showed that demonstrations provided faster and more accurate learning during the training session. However, during the later sessions those who saw demonstrated procedures took longer to perform the tasks than did users of written instructions. Users appeared to be mimicking the training demonstrations without processing the information which would be needed later. In fact, when users had to infer procedures for tasks which were similar to those seen in the training session, the text group was much better at deducing the necessary procedures than the demonstration group. These findings indicate that animated demonstrations, as they were implemented for this study, were not robust enough to aid in later transfer.

1. Introduction

When learning to use a computer application, many users are confronted with instructions in a written formant. Yet these instructions are too often discarded because the user wants to get started immediately, or because the instructions are difficult to follow and to assimilate (Mack, Lewis & Carroll, 1983). Often the user will try to find an expert or colleague who can demonstrate the appropriate interface procedures. Since users may generalize these demonstrated procedures to other tasks, the demonstrations may result in "one-trial" learning (Lewis, Casner, Schoenberg & Blake, 1987). The purpose of this research is to explore these issues and determine whether demonstrated, animated instructions result in faster learning, better retention, and transfer to similar tasks when compared with written instructions.

Recent interest in using demonstrations has been generated by graphical animation software for programming by example (Myers, 1987; Duisberg, 1988).

† Current address: Philips Laboratories, 345 Scarborough Road, Briarcliff Manor, New York, NY 10510, USA.

‡ HyperCard is a trademark of Apple Computer, Inc.

"Watch me do it" demonstrations have appeared in experimental interfaces such as NLS-SCHOLAR (Grignetti, Hausmann & Gould, 1975) and CADHELP (Cullingford, Krueger, Selfridge & Bienkowski, 1982; Neiman, 1982). The NLS-SCHOLAR system used artificial intelligence techniques to teach text-editing skills by showing the user how to perform editing tasks. In CADHELP, an animation program was able to simulate the sequential performance of low-level interface actions by running scripts of these procedures. For example, users could see a demonstration of how the cursor is moved to drag a device. The focus of these efforts has been on the knowledge representations and algorithms required for animation of interface procedures rather than examining the value of demonstrations as a learning aid for users.

Today, demonstrations are being added to software packages, indicating that interface designers think they are effective ways of teaching interface methods. The Apple LisaTM and MacintoshTM, for example, included "getting started" tours for users. In these tours, basic procedural skills (e.g. pointing and dragging) and more advanced skills (e.g. choosing commands and cutting and pasting) are taught through animated demonstrations, textual instruction, and guided user practice (Apple Computer, 1988). Macintosh application developers have also begun to follow this pattern of using guided tours as a teaching aid (VideoWorks IITM†, MacroMind, 1987).

Despite these theoretical and applied efforts, little empirical data exist on the efficacy of animated demonstrations for training or helping users. A limited evaluation (six experienced users) of LisaGuideTM‡ by Carroll and Mazur (1986) suggested that demonstrations may introduce their own usability problems. In another study, Lewis, Hair and Schoenberg (1989a) compared videotaped demonstrations with a tutorial-format manual. Unfortunately, the two instructional formats were not equivalenced (a procedure which allows for the comparison of instructions solely on the basis of media). For example, the video tape was removed after viewing so that the users had to work from memory whereas the manual was available to the other group while they performed the tasks. Nonetheless, results showed that the manual group out-performed the video group in the number of tasks they were able to complete. Therefore, from an engineering perspective, the large amount of work required to develop graphical animations demands that we understand its potential benefits and limitations for usability.

In the current experiment, animated demonstrations were compared with written step-by-step instructions for learning procedural interface tasks. We hypothesized that the two media may be encoded differently as a result of their individual instructional characteristics. Users receiving animated demonstrations lack the explicit, verbal element of the instruction. Instead, they have two visual codes, one from watching the instructions and one from watching themselves perform those actions. With animated demonstrations users also gain the motoric component of the instruction when actually performing a task. Alternatively, the user who reads instructions and then attempts to perform the task, gets verbal information while reading, and visual and motoric components when performing. This scheme is loosely based on the dual-coding theory set forth by Paivio (1971). We have added

† Videoworks II is a trademark of MacroMind, Inc.

‡ Apple Lisa, Macintosh and Lisa Guide are trademarks of Apple Computer, Inc.

the motoric code because evidence suggests that, for procedural tasks, actual manipulation is an important element in learning (see Baggett, 1987).

In contrast to written instructions, animated demonstrations may convey procedural knowledge more directly about the interface. An animated demonstration shows the user how the interface appears as the procedure is executed and also links the input actions with the interface results. Users should be able to experience in concrete visual terms how each procedural step contributes to the overall task goal (Lewis, Casner, Schoenberg & Blake, 1987). Animated demonstrations may allow users to rehearse and plan visually while watching instructions, thereby reducing the additional load of forming a motoric code. Animated demonstrations, therefore, may improve initial learning when compared with written instructions since the amount of cognitive processing will be reduced during the learning stage.

Reduced processing during initial learning with animated demonstrations is also predicted because demonstrations are integrated into the interface to be learned and serve as examples. As articulated by Anderson, Boyle, Farrell and Reisner (1984), learning in the problem context should decrease the difficulty of encoding procedural knowledge. A very similar prediction can be made if animated demonstrations are considered as examples of interface procedures. Although research has focused predominantly on textual examples, these demonstrations seem more like examples than textual procedures because the user is able to see the impact on the interface while the task is being performed. Thus, previous research using textual examples may be applicable to our demonstration examples.

Well constructed examples have been found to improve initial learning in several problem domains (Lewis & Anderson, 1985; Sweller & Cooper, 1985). LeFevre and Dixon (1986) have also found that people rely consistently on examples. They found this preference for examples in a wide variety of conditions and state that people processed written instructions only superficially when examples were available. The reason for users' reliance on examples and performance improvements seems to be the close match between the features of the problem and the example (Lewis & Anderson, 1985) which may hold true for animated demonstrations. In total, users may be more inclined to use the animated demonstrations and also may find them easier to use.

Still, there is the danger that users may passively watch the animated demonstrations and then blindly mimic these procedures with very little processing and encoding (LeFevre & Dixon, 1986). Suppose a user learns a task through an animated demonstration and then is faced with a similar task. Will the animated demonstrations studied earlier aid the user in performing the new task? Some evidence for transfer of training with demonstrated procedures has been provided by Lewis *et al.* (1987). The following is an example of a generalization situation in which users are shown commands which result in a given action (Lewis, *et al.*, 1989b):

Consider the following three commands from a fictitious system:

- (E1) foo baz: deletes the authorization table
- (E2) blee baz: deletes the terminal assignment table
- (E3) foo bar: prints the authorization table

What command would you issue to print the terminal assignment table? Probably you will say "blee bar".

However, the generalization tasks and the demonstrated results used are relatively simple since they are only two word commands. The ability to generalize between more complex procedures, as in computer tasks, may be more difficult.

To understand better the demonstration media, we studied animated demonstrations and written directions for moderately complex interface procedures using HyperCard, a hypertext system for the Apple Macintosh. This application was chosen since the object-oriented procedures of HyperCard provide a concrete, direct manipulation interface to fully explore the possible benefits of animated demonstrations. To assess animated demonstrations in comparison to written instructions in different learning phases, we observed users learning and practising with the instructional media, performing similar and different interface tasks immediately after instruction, and performing these same tasks after several days delay.

2. Method

2.1. PARTICIPANTS

Twenty-eight students and staff (14 of each gender) from the University of Michigan served as participants. Each was required to have four months experience in using both a graphics and a word-processing application on the Macintosh, but no experience in using HyperCard. A background survey of participants showed that they had a fair amount of experience using computers. The age of the participants ranged from 19 to 43 years (average: 25.8 years) and all were either attending the University of Michigan or had a college education.

2.2. EXPERIMENTAL DESIGN

The study was a four variable ($2 \times 2 \times 7 \times 3$), mixed-factor design. The between-subjects variables were the instructional media presented (animated demonstrations or written instructions) and the amount of required practice during instruction (one or three mandatory trials). The practice variable was included to ascertain whether additional performance affected learning and retention of the procedures presented by the two instructional media. Since this practice variable led to no interesting or significant results it will not be discussed further. The within-subjects variables were the seven procedural tasks and the three performance assessment sessions in which those tasks were presented. The performance assessment sessions were: (1) a training session to determine how participants learned the tasks with the animated or written instructions; (2) an immediate test session to test the procedural skills they had just learned; and (3) a delayed time session, approximately three days later, to assess retention of these skills.

2.3. HYPERCARD AUTHORING TASKS

The experiment was conducted using an Apple Macintosh II with an 11 in. monochrome display running the HyperCard application (Apple Computer, 1988). Low-level HyperCard tasks were used for the study. These tasks were selected because they were elemental tasks which would be needed to be able to create (author) a HyperCard stack. Tasks were also chosen to allow us to assess retention

TABLE 1
Same, similar and different tasks in the training, immediate test and delayed test sessions

Training tasks		Immediate test tasks		Delayed test tasks
Copy Text	» » » » » » » »	Copy Text	» » » » » » » »	Copy Text
Link Button	» » » » » » » »	Link Button	» » » » » » » »	Link Button
Modify Field	» » » » » » » »	Modify Field	» » » » » » » »	Modify Field
Create Button	> > > > > > >	Create Field	» » » » » » » »	Create Field
Copy Button	> > > > > > >	Copy Field	» » » » » » » »	Copy Field
Create Card		Delete Card*	» » » » » » » »	Delete Card
Create Stack		Delete Button*	» » » » » » » »	Delete Button

» » » » Task is the *same* between sessions.
 > > > > Task is *similar* between sessions except for type of object.
 * Task is *different* from previous session.

of previously learned skills and transfer of training. Tasks for the immediate and delayed test sessions were either: (1) the *same* as tasks from the training session, (2) *similar* to tasks from the training session (e.g. copy field instead of copy button), or (3) *different* than those used in the training session, but the same in the two test sessions. Table 1 presents the three types of tasks. All tasks were counterbalanced between participants and within sessions using a Latin square.

2.4. INSTRUCTIONAL MATERIALS

Procedural instructions were created for the basic authoring tasks of HyperCard and were presented in two ways. The first was similar to the written procedural instructions found in many online help systems and in textual documentation. An example of the written procedure for copying a HyperCard button in shown in Figure 1. All steps in these written instructions were shown concurrently and filled the Macintosh screen. A time allowance of four seconds per procedural step, determined in pre-testing, equalled the total presentation time. The other type of instruction was a real-time, animated demonstration of the interface procedures generated with the TempoTM† macro facility (Affinity Microsystems Limited, 1986), but without any accompanying written or spoken text. The animated demonstrations

1. Select Button Tool from Tools menu.
2. Click on button to copy.
3. Select "Copy Button" from Edit menu.
4. Select "Paste Button" from Edit menu.
5. Click on middle of button and drag to move the button to correct location.
6. Select Browse Tool from Tools menu to determine if button was modified correctly.

FIGURE 1. Written procedures for copying a button.

† Tempo is a trademark of Affinity microsystems Ltd.

were shown on the display as if another person were performing the steps needed to accomplish an authoring task.

For each task, equivalencing procedures, similar to those conducted by Baggett (1979), were undertaken in an attempt to assure that differences between instructions were in terms of media rather than content. Four additional participants were given labeled written instructions and asked to look for dissimilarities or missing items in the animated instructions. Based on the equivalencing procedures, six changes were made to the 45 total instructions in the seven tasks used during the training session. Most of these changes were in the wording of the written instructions to provide greater consistency between the two media.

As shown in Table 2, the presentation times for the two media types were not equivalent. Four seconds per procedural step equalled the presentation time for the written instructions while the animated demonstrations had longer presentation times per step (5.2–8.9 sec). The longer presentations for the animated demonstrations were the result of additional time required for mouse movements and system response time. Although this was a potential time advantage for the users of the animated demonstrations, we felt that the serial nature of the demonstrations necessitated the additional time. Users of the animations were not allowed to review parts of the demonstrations or return to previous portions of the instruction, whereas users of the written instructions were afforded reading advantages, such as being able to scan and re-read text. Thus, we felt the time allowances made for a fair comparison between the two treatments.

2.5. PROCEDURE

The experiment was conducted on two days for each participant, the first lasting approximately 90 min and the second lasting about 20 min. During the first day each participant received preliminary training on HyperCard, learned authoring tasks using one of the two media in the training session, and performed authoring tasks without any instructions in the immediate test session. Before leaving, participants filled out a questionnaire, and were asked not to use HyperCard until they had completed the entire experiment. On the second day, 3 to 7 days

TABLE 2
Number of procedural steps and times for task instructions for both media

Learning tasks	Procedural steps	Total presentation time (sec)	
		Text	Animation
Create Card	2	8	14-40
Create Stack	4	16	35-50
Copy Text	5	20	30-40
Copy Button	6	24	33-27
Link Button	7	28	40-72
Modify Field	8	32	42-09
Create Button	9	36	46-56
Means	5.86	23.43	34.71

later, participants received some initial warm-up trials and performed the same authoring tasks as those in the immediate test session.

2.5.1. Preliminary HyperCard training

During the first day each participant was familiarized with the basic functions of HyperCard by reading a hardcopy tutorial. Included were online skills in how to browse through cards, the basics of how cards are arranged in stacks, and the major components of the system. Short criterion tests were administered after the preliminary HyperCard training to insure that the participant had understood the concepts as well as how to navigate through the HyperCard system. Those areas in which the participant was still deficient were re-taught and tested again. The training period, including criterion tests, lasted approximately 30 min.

2.5.2. Training trials

For the training tasks shown in Table 1, a goal name or task was presented along with a precise description for the goal. Each participant was told to read the goal and its description thoroughly and then to invoke the instruction for that task. To begin, participants initiated a Tempo "macro" which began the animated or written instructions. This invocation sequence was practised in a series of simple non-Hypercard goals before data collection began. Once invoked, the instructions appropriate for the stated task were immediately provided.

After the textual or animated instructions were presented for a task, participants clicked the mouse on a "start" button, performed the task, and clicked on a "done" button when they were finished. The system then tested the results for correctness and provided limited feedback about their performance.† If incorrect, participants were required to ask for the instructions and perform the task again until done correctly.

2.5.3. Test sessions

On the immediate and delayed test sessions, participants performed the tasks shown in Table 1 which were the same, similar and different from the tasks in the training session. In these sessions they were required to perform the tasks without the instructions.

3. Results

The following sections give the performance results for: (1) tasks that were identical over all three sessions; (2) tasks that were similar over the three sessions; and (3) tasks that were identical over the last two sessions. Results for tasks which were different between the training and testing sessions will not be reported here because there were only two different tasks and no interesting results. In each section, the dependent measures will be discussed in the following order: (1) *total time* which

† Within the HyperCard system, we were able to examine if the user had executed the essential portions of the task correctly by checking what commands had been used and how the end object appeared. For example, if the task was to create a new button and put it in a specific location, a trial would not be considered correct if the user copied an existing button or placed it incorrectly. When users improperly performed the task, they would be notified of this when signalling that they had finished.

excludes instruction time and includes the time for both correct and incorrect trials; (2) *percentage of correct trials* which is the number of correct trials divided by the number of total trials for each task; and (3) *time per attempt* which is the total time divided by number of trials for each task.

3.1. IDENTICAL TASKS

An analysis of variance (ANOVA) on the mean total times to complete the three tasks which were the same in all three performance sessions revealed main effects of task ($F[2, 48] = 3.93, p = <0.03$) and session ($F[2, 48] = 4.92, p = <0.01$). Figure 2 shows the means for the two media over the three performance sessions. The task effect was expected since the tasks were of different length and difficulty. This effect will not be discussed further since there was no interaction of task and medium ($p = >0.5$) to indicate that one instructional medium was more or less effective with some of the tasks. However, there was an interaction between medium and session ($F[2, 48] = 5.77, p = <0.01$) indicating that the instructional medium was differently effective in the three sessions.

As can be seen in Figure 2, users of the written text took more than 50% more time than users of the animated demonstrations during the training session. However, after the training session there was a large decrease in total time for users of the written instructions. This led to total times for the users of the written instructions which were substantially below those of the users of the animated demonstrations in the immediate and delayed testing sessions. To interpret these data, a *post hoc* simple main effects analysis was conducted. Over the three sessions, users that received the written instructions improved significantly ($F[2, 48] = 9.58$,

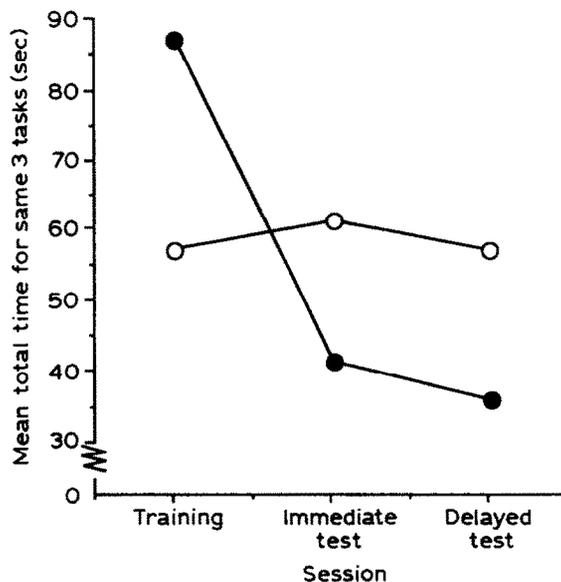


FIGURE 2. Mean total time to perform the same three tasks at each performance session. Key: ○, animation; ●, text.

$p = <0.001$) whereas there was no corresponding improvement for users of the demonstrations ($p = >0.5$). Differences between the media groups at each session were also non-significant ($p = >0.05$). Therefore, the data indicate that users of written instructions improved significantly on the same tasks, while users of the animated demonstrations remained at the same level.

An analysis of the percentage of correct trials followed a pattern similar to the total time data. Main effects of task ($F[2, 48] = 5.59, p = <0.007$) and session ($F[2, 48] = 5.41, p = <0.008$) were found. Once again, differences between tasks were expected and not of specific interest to this investigation since task and medium did not interact statistically ($p = >0.35$). However, there was an interaction of session and medium ($F[2, 48] = 5.15, p = <0.015$) indicating that percentage of correct trials depended on the performance session. The percentage of correct trials data for the interaction of session and medium are shown in Figure 3. As illustrated, the percentage of correct trials in performing the tasks was generally high (90% +) with the exception of users in the text group during the first session (79.3%). A *post hoc* simple main effects comparison found that text group was significantly less accurate than the animation group during the training session. Moreover, over the three sessions the percentage of correct trials for the text group improved significantly ($F[2, 48] = 7.77, p = <0.002$) while the animation group did not improve significantly ($p = >0.5$). These results confirm that the percentage of correct trials did not diminish as speed increased in later sessions.

To check on the rate at which tasks were performed, an ANOVA on the time per attempt was conducted. This analysis found no significant differences between the text and animation groups or in the interaction between medium and sessions ($p = >0.5$). Therefore, all groups during all sessions were performing at approxi-

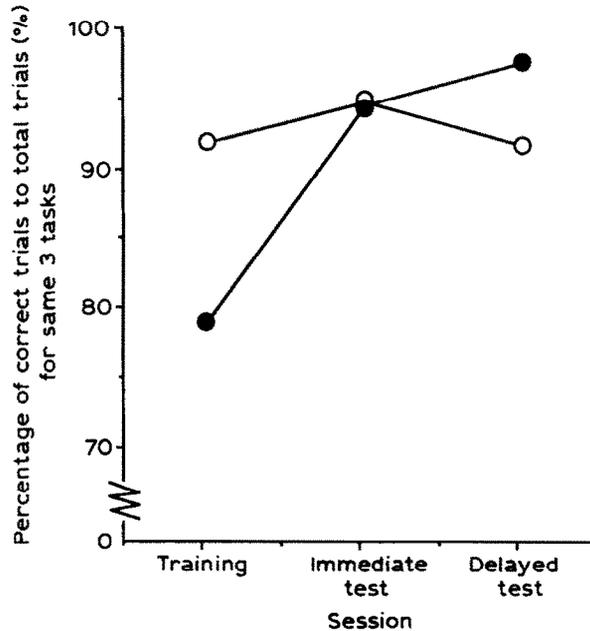


FIGURE 3. Mean percentage of correct trials to total trials while practising the same three tasks at each performance session. Key: ○, animation; ●, text.

mately the same pace. These results indicate that text users committed more task errors during training, thereby increasing the average number of trials needed to complete training, and ultimately increasing the average total time.

3.2. SIMILAR TASKS

To study transfer of training in this study, those tasks which were similar between the training and testing sessions were compared: Copy Button→Copy Field and Create Button→Create Field (Table 1). The tasks from training to testing sessions differed only in the object on which they operated. For example, in the training session the task "Copy Button" was changed to "Copy Field" in the later sessions. The operations for these tasks are very similar except that the change of objects (Button to Field) required a few changes in the procedures (e.g. "Select Field Tool" instead of "Select Button Tool").

Mean total performance times for the groups over the three sessions are seen in Figure 4 and were analysed using paired *t*-tests. *T*-tests were used because the tasks between the training and testing sessions were similar, but not identical (Table 1), and thus could not be compared using ANOVA. Between the training session and the immediate test session the users in both groups had a significant change in performance (animation: $t[13] = -2.25$, $p = <0.045$; text: $t[13] = 2.90$, $p = <0.015$). The text group participants were able to complete the similar tasks more quickly than the original tasks, whereas the animation group participants took more time when faced with a similar but different task. The results suggest that when presented with these similar tasks, text users experienced positive transfer and animation users experienced negative transfer. After performing these similar tasks in the immediate

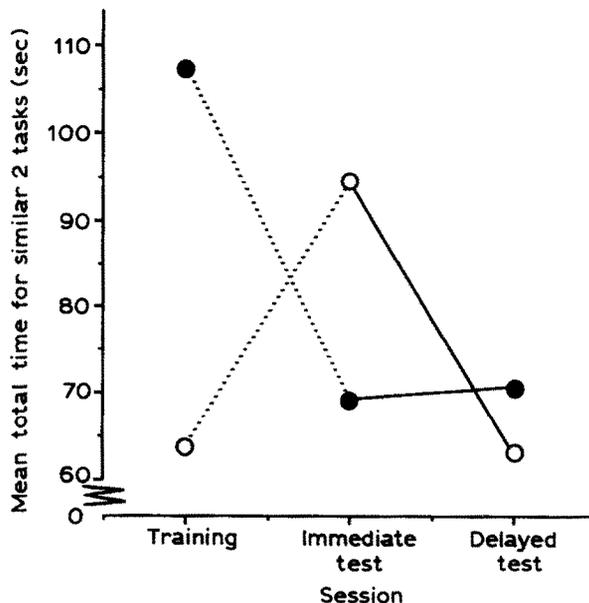


FIGURE 4. Mean total time to perform two similar tasks at each performance session (dotted lines signify that different, but similar, tasks are compared between the training and later test sessions). Key: ○, animation; ●, text.

test session, there was a significant reduction in time to perform these same tasks in the delayed test session for the animation group ($t[13] = 3.23, p = <0.0066$) and no corresponding change in time for the text group ($p = >0.5$). It appears that once the tasks were learned, the animation users were able to perform during the delayed test session at a level similar to the training session (63.9 s at training and 63.1 s at delayed test session).

No significant differences were found between the two media on the basis of accuracy between sessions for similar tasks. However, for time per attempt, shown in Figure 5, there was an increase for both media groups between the training and immediate test session (animation: $t[13] = -4.74, p = <0.0004$; text: $t[13] = -2.39, p = <0.035$). In addition, the time per attempt for the animation group decreased between the immediate test session and the delayed test session ($t[13] = 5.90, p = <0.0001$). These data suggest that the occurrence of a similar, but new task causes users to slow down their rate of performance perhaps due to an increase in time to problem solve in the immediate session.

3.3. IDENTICAL TASKS OVER THE TWO TEST SESSIONS

The amount of information which was retained between immediate and delayed test sessions was analysed by looking at the seven tasks which were the same between the two sessions. To a large extent the results of these analyses were dominated by task effects. For example, highly significant task effects were found for all dependent measures (total time: $F[6, 144] = 9.14, p = <0.0001$; the percentage of correct trials: $F[6, 144] = 5.86, p = <0.0001$; time per attempt: $F[6, 144] = 19.79, p = <0.0001$).

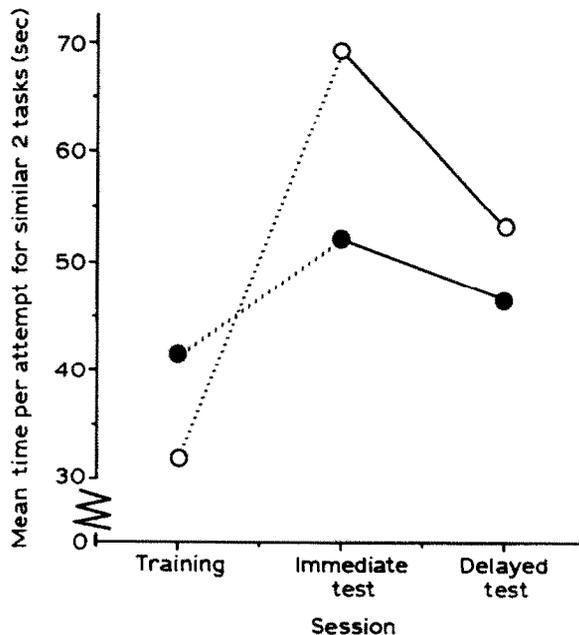


FIGURE 5. Mean time per attempt to perform two similar tasks at each performance session (dotted lines signify that different, but similar, tasks are compared between the training and later test sessions). Key: ○, animation; ●, text.

In addition, smaller interaction effects of tasks with other independent variables were found. For the total time measure a third-order effect of task, session and mandatory trial variables was found ($F[6, 144] = 2.25, p = <0.04$) and for the percentage of correct trials measure an interaction of task and session was found ($F[6, 144] = 3.19, p = <0.01$). Since none of these effects and interactions involved the media type, the main focus of this study, further presentation of the data will not be reported. The lack of significant differences due to media on these identical tasks in the test sessions suggests that the type of medium does not affect how well the procedural skills are retained and remembered over a period of time.

4. Discussion

We expected that animated demonstrations would reduce the amount of translation needed when getting started, serve as an interface example, and provide instructions that are more explicit because they are integrated with the interface. To some extent these predictions were true. The animation group was almost 50% faster than the text group in the training session. Yet, once the instructions were removed the animation group did not improve as the text group did. As we discuss the possible reasons for these paradoxical results, we will also explore how these animated demonstrations could be improved for advanced learning and retention along with the faster training capability already seen.

The two groups exhibited different behavior which might help explain the unexpected results. Some members of the animation group anecdotally reported that during the training session they had "... just done what the computer had done. . . ." Evidence for this mimicking behavior is also suggested by the data. For the same tasks across sessions, performance for the animation group was remarkably stable. In contrast, participants in the text group, although slower and more inaccurate during the training session, improved their performance in the immediate and delayed sessions so that they were approximately 20s faster and 10% more accurate than those in the animation group. Since performance did not improve for the animation group, they may have been stuck at the mimicking stage with their representation based solely on a rote procedure.

One reason for this mimicking behavior may be the different processing required by text and animation which may account for the different performance results. As discussed earlier, the text group was exposed to a richer encoding medium because it contained verbal, visual and motoric codes, whereas the animation group had only the visual and motoric components. In fact, the text group could have visualized the instructions as they read them, providing an even richer visual code during instruction and performance. Other evidence for the different processing of the animated demonstrations was that some participants vocalized what they saw. Users seemed to want a verbal component with the demonstrations suggesting that all three codes (verbal, visual and motoric) may be desirable. Indeed, having all three codes during learning and performance could lead to a deeper and richer level of processing. This may explain the extra time required for text users during the training session. These users probably needed additional time to translate the verbal instructions into visual and motoric codes. Thus, although reading instructions is commonly disliked and often skipped (Carroll, Mack, Lewis, Grischkowsky &

Robertson, 1985), the improved performance by the text group suggests that this processing may be beneficial.

Perhaps animated demonstrations and written instructions should be used in combination. In a study by Booher (1975), the relative comprehensibility of various picture-word formats was compared for procedural instructions. The highly pictorial formats which included text were consistently faster and more accurate than the other formats. Using only highly textual formats increased performance time whereas a picture-only format increased task errors. The results of the Booher study, which concentrated on static pictures and not animated demonstrations, bear a similarity with the results of this study. Visual instructions appear to provide the direct mapping from the instruction to the action to be performed. Booher suggests that users pick up advance information from pictures which helps them to organize. Thus, it may be that demonstrations with accompanying text would allow for faster learning and improved retention. This result is supported in the comparison of video-taped demonstrations and text by Lewis *et al.* (1989b). They suggested that the addition of text to demonstrations should call attention to key aspects of the operations, and missing connections between steps should be indicated clearly.

The text group was also able to infer more quickly the needed procedures when similar tasks were presented. It was expected that animated demonstrations would serve as an interface example of a task, making it easier to infer procedures. Yet, the animation group was much slower than the text group at determining the corresponding procedures for a similar task (approximately 20 s per task). The slower performance of similar tasks may be due to the inability to remember the task from the training session or the inability to grasp the consistency in methods between the similar tasks. In addition, if participants were only mimicking the training procedures, resulting in a superficial encoding of the task, they were probably unable to alter the procedures from the training task when a similar task was presented in the immediate test session.

Therefore, these results call into question the hypothesis that animated demonstrations might serve as interface examples. It has been noted that well developed examples can aid learning (Lewis & Anderson, 1985; Sweller & Cooper, 1985). However, the degraded performance of animation users between the training and immediate test session for similar tasks suggests that these users did not view the training tasks as generalizable examples. Instructions given to the participants before the training session included a warning that they would see tasks that were "... the same or similar to the tasks you will learn about. ..." Yet, they seem to have viewed each task's procedure as a solitary method to accomplish a single goal without looking for consistencies between the training tasks and similar immediate test tasks.

Finally, the type of animated demonstrations used in this study may be the reason why the animation users did not perform as well as expected. The amount of time given to the animation group to watch the training sequences was determined somewhat arbitrarily, but was uniformly longer than for the text group. This time variation, combined with the user's inability to stop the demonstration or begin a demonstration in the middle, may have led to poorer performance. Although the text group was also given a time constraint, they were able to read any part of the written instructions and at any speed. In fact, it would seem that since the

demonstration group received more time during training, they would perform better. This was not the case. In addition, the structure of some tasks may not be suitable for animated demonstrations. If a task is a combination of lower level tasks, the task structure may not be apparent if the subtasks are simply strung together in a serial demonstration.

5. Summary

Animated demonstrations will not be an appropriate online help aid for all interfaces. Their use is limited to interfaces which are highly graphical and contain few hidden responses to user input. This technique is being used to teach procedural methods in direct manipulation interfaces such as the Macintosh and CAD/CAM systems. Furthermore, although demonstrated procedural instructions might appear to have a clear advantage over other, less high-tech media such as written instructions, it could not be deduced from this study. Thus, the way in which the animated demonstrations are constructed should be carefully undertaken to complement the task and human capabilities.

Animated demonstrations, as were implemented for this study, appeared initially to have an advantage over text because of the faster training trials. Paradoxically, this advantage in training was translated into poorer performance once the demonstrated instructions were removed. Those who used the demonstrations were not as adept at transferring their newly acquired knowledge to new situations. Unfortunately, the demonstrations did not require the processing which would aid in future learning. This major result should be heeded when creating demonstrations.

While creating the demonstrations for this study, watching users learn with them, and from results reported, a list of guidelines have been compiled. The following points should be considered when creating animated demonstrations for procedural tasks:

- Create animations which will not simply be mimicked.
- Include the verbal component with the demonstrations so that the verbal, visual, and motoric components are all present.
- Do not assume that simply because a visual component is present, that the instructions will be encoded into long term memory.
- Do not assume that users of animated demonstrations without text will be able to infer similar procedures.
- Do not make procedures that are too complex or too long.
- Insure that all results of an action are visible during a demonstration.
- Be very careful with how you pace a demonstration—pre-testing is probably necessary.

Ongoing research will look more closely at these guidelines. Another study is in progress which has an increased number of tasks while introducing new types of tasks. In addition, more thorough and new types of equivalencing of the two types of instructions will be used. In addition, the future research will help to determine ways of combining the speed in learning with later retention and problem solving by combining text with demonstrations.

This research was supported under ONR Contract number N00014-87-K-0740 with John J. O'Hare serving as the technical monitor. The authors wish to thank Richard Gong for his help in data collection and for his insights. The conclusions within this paper are those of the authors and not necessarily those of ONR.

References

- AFFINITY MICROSYSTEMS, LTD. (1986). *Tempo Manual*. Boulder, CO.
- ANDERSON, J. R., BOYLE, C. F., FARRELL, R. & REISNER, B. J. (1984). *Cognitive Principles in the Design of Computer Tutors*, Technical report #ONR-84-1. Pittsburgh: Carnegie-Mellon University, Department of Psychology.
- APPLE COMPUTER, INC. (1988). *HyperCard® User's Guide*. Cupertino, CA.
- APPLE COMPUTER, INC. (1988). *Macintosh® II Owner's Guide*. Cupertino, CA.
- BAGGETT, P. (1979). Structurally equivalent stories in movie and text and the effect of the medium on recall. *Journal of Verbal Learning and Verbal Behavior*, **18**, 333–356.
- BAGGETT, P. (1987). Learning a procedure from multimedia instructions: The effects of film and practice. *Applied Cognitive Psychology*, **1**, 183–197.
- BOOHER, H. R. (1975). Relative comprehensibility of pictorial information and printed words in proceduralized instructions. *Human Factors*, **17**, 266–277.
- CARROLL, J. M., MACK, R. L., LEWIS, C. H., GRISCHKOWSKY, N. L. & ROBERTSON, S. R. (1985). Exploring exploring a word processor. *Human-Computer Interaction*, **1**, 284–307.
- CARROLL, J. M. & MAZUR, S. A. (1986). Lisa learning. *Computer*, **19**, 35–49.
- CULLINGFORD, R. E., KREUGER, M. W., SELFRIDGE, M. & BIENKOWSKI, M. A. (1982). Automated explanations as a component of a computer-aided design system. *IEEE Transactions on Systems, Man and Cybernetics*, **12**, 168–181.
- DUISBERG, R. A. (1988). Animation using temporal constraints: An overview of the Animus system. *Human-Computer Interaction*, **3**, 275–307.
- GRIGNETTI, M. C., HAUSMANN, C. & GOULD, L. (1975). An “intelligent” on-line assistant and tutor—NLS-SCHOLAR. In *Proceedings of the 1975 National Computer Conference*, **44**, pp. 775–781, Anaheim, CA: AFIPS Press.
- LEFEVRE, J. A. & DIXON, P. (1986). Do written instructions need examples? *Cognition and Instruction*, **3**, 1–30.
- LEWIS, C., HAIR, D. C. & SCHOENBERG, V. (1989a). Generalization, Consistency and Control. In *Proceedings of CHI 1989*, pp. 1–5. New York: ACM.
- LEWIS, C., HAIR, D. C. & SCHOENBERG, V. (1989b). Learning from a demonstration, Technical report CU-CS-436-89. Boulder: University of Colorado, Department of Computer Science.
- LEWIS, C., CASNER, S., SCHOENBERG, V. & BLAKE, M. (1987). Analysis-based learning in human-computer interaction. In *Proceedings of INTERACT '87*, pp. 275–280. New York: IFIP.
- LEWIS, M. W. & ANDERSON, J. R. (1985). Discrimination of operator schemata in problem solving: learning from examples. *Cognitive Psychology*, **17**, 26–65.
- MACK, R. L., LEWIS, C. & CARROLL, J. M. (1983). Learning to use a word-processor: problems and prospects. *ACM Transactions on Office Information Systems*, **1**, 254–271.
- MACROMIND, INC. (1987). *VideoWorks II. Animated Presentations for the Office, School or Home*. Chicago, III.
- MYERS, B. A. (1987). Creating dynamic interaction techniques by demonstration. In *Proceedings of CHI + GI 1987*, pp. 271–278. New York: ACM.
- NEIMAN, D. (1982). Graphical animation from knowledge. In *Proceedings of National Conference on AI*, pp. 373–376. Menlo Park, CA: AAAI.
- PAIVIO, A. (1971). *Imagery and Verbal Processes*. New York: Holt, Rinehart & Winston.
- SWELLER, J. & COOPER, G. A. (1985). The use of worked examples as a substitute for problem solving in learning algebra. *Cognition and Instruction*, **2**, 59–89.