



This item was submitted to Loughborough's Institutional Repository by the author and is made available under the following Creative Commons Licence conditions.



CC creative commons
COMMONS DEED

Attribution-NonCommercial-NoDerivs 2.5

You are free:

- to copy, distribute, display, and perform the work

Under the following conditions:

BY: **Attribution.** You must attribute the work in the manner specified by the author or licensor.

Noncommercial. You may not use this work for commercial purposes.

No Derivative Works. You may not alter, transform, or build upon this work.

- For any reuse or distribution, you must make clear to others the license terms of this work.
- Any of these conditions can be waived if you get permission from the copyright holder.

Your fair use and other rights are in no way affected by the above.

This is a human-readable summary of the [Legal Code \(the full license\)](#).

[Disclaimer](#) 

For the full text of this licence, please go to:
<http://creativecommons.org/licenses/by-nc-nd/2.5/>

Major extract from Computers and Graphics Vol 17, No. 3, pp 205 – 218, 1993. The text has been modified and expanded in places.

Experiments using Multimedia Interfaces in Process Control: Some Initial Results

James L. Alty, Marius Bergan and Penny Craufurd
Loughborough University of Technology
Department of Computer Studies
Loughborough, Leicestershire LE11 3TU, England
Email: J.L.Alt@uk.ac.lut, M.Bergan@uk.ac.lut

Ciaran Dolphin
Work Research Centre Ltd.
22 Northumberland Road
Dublin 4, Ireland

Abstract: This paper reports a series of experiments using different combinations of multimedia interfaces. The task used in the experiments is the Crossman Waterbath. The media combinations are compared and contrasted to draw out pointers towards the effects of different media. It is concluded that warnings affect comprehension and should be minimised during exploratory learning. Sound results were disappointing, and had a detrimental affect on performance. However, it was thought that the problem lay with the lack of discrimination between different sound levels. Those subjects who found the task hard were greatly helped by speech warnings. In conclusion, the experiments show that different presentational styles do indeed matter, and it is possible that multimedia interfaces are particularly useful for aspects of the interface which are difficult to understand.

1. Multimedia Experiments in the Behaviour Laboratory.

1.1. The Experiment

The PROMISE project is an ESPRIT project funded by the EU, investigating the usefulness of multi-media interfaces in Process Control. It uses, as exemplars, the interfaces to a Nuclear Power station simulator (Scottish Power) and a large Chemical Plant in Terneusen (Netherlands). Because experiments and observations in industrial exemplars were expected to take some time to set up and execute, the project decided initially carry out some experiments in the laboratory to inform later work in the plants. The subjects taking part in the experiment would be Computer Science and Engineering Postgraduate students. The task chosen, therefore, should be relevant to process control (i.e. a relatively complex dynamic task with multiple state and control variables), it should also have scope for multimedia presentation, should be readily learned by non-experts, and should be straight forward to implement (i.e. in a two-three week period).

The task selected for this experiment was Crossman's Water bath (Crossman and Cooke, 1974). This task is closely related to the process control domain (Sanderson et al. , 1989) and we felt there was scope for multimedia presentations. It has been used on numerous occasions (Moray et al., 1986, Moray and Rotenberg, 1989) since it was first used by Crossman and Cooke to show operator progression from a second order controller to exhibiting open-loop control.

The Water bath Task, illustrated in figure 1, is a simulated thermal hydraulic system which consists of a single tank (in the variation used for this experiment). An inflow and an outflow pipe are connected to the tank. There is a valve on each pipe which may be used to

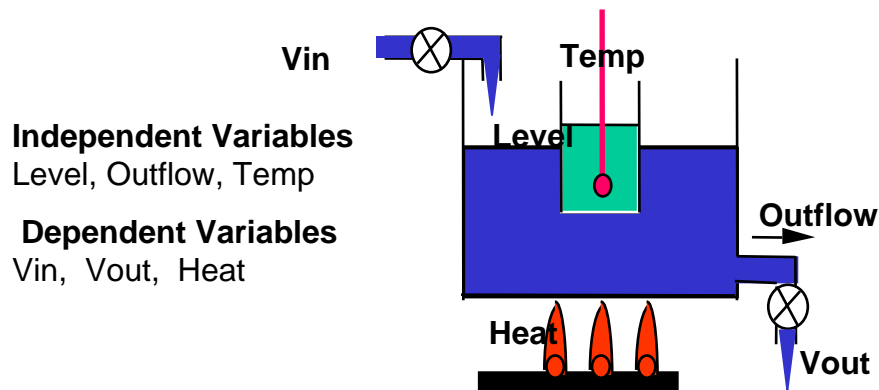


Figure 1 Crossman's Waterbath

regulate the flow. The heater is situated immediately underneath the tank. Inside the tank is an insulated container containing a fixed amount of water and a thermometer. There are also imaginary sensors to measure the level of water in the tank and the rate of flow out of the tank.

Thus there are three control variables: V_{in} , V_{out} and $Heat$. In our implementation of the system the controls can be set to an integer value between 0 and 100, where 0 means completely shut (or off) and 100 means completely open (or on). The state of the process is indicated by the three variables $Level$ (in millimetres), $Outflow$ (millilitres per second) and $Temp$ ($^{\circ}C$).

Subjects had to solve a task that involved achieving new steady state process states from a steady state starting point. The target state was defined as a range within which each of the system variables had to lie after stabilisation of the system. Subjects were placed in front of a PROMISE terminal and their actions were recorded on two television cameras. Sound and speech activity were also recorded. In addition, some gesture information was recorded (as it happened) on the videotape by a hidden observer.

2.2. The Procedure

More than 50 subjects participated in the experiment. They were mainly undergraduate students in Computer Science or Engineering and were also mainly male. The few females who took part were balanced across the experimental conditions. Each session lasted about 2 1/2 hours including a debriefing interview. Subjects were given a brief introduction to the experiments which explained the task and described the control and state variables. No principles on how to operate the system were given. Subjects were only told how to operate the interface to control the simulation.

Each session consisted of two halves. During the first half, the subject completed 21 problems of increasing difficulty. These problems only involved simple increases or decreases in one or more variable. The number of variables which had to be increased or decreased is referred to as the *dimensionality* of the task by Sanderson et al. Another attribute of the task is its *compatibility*. This corresponds to the minimum number of actions required to solve a problem. Sanderson et al. called problems that could be solved with a single action compatible, and problems which required two actions incompatible. For example, the task of increasing the $Level$ and increasing the $Outflow$ is compatible because the dual goal can be achieved with the single action of opening V_{in} from its initial setting. The task of increasing the $Level$ whilst decreasing the $Outflow$, however, is incompatible because it requires two actions, such as a reduction in V_{out} (to increase the $Level$) followed by a reduction of V_{in} (to

decrease the *Outflow*). It was shown by Sanderson et al. that compatibility had a higher impact on performance than the dimensionality of a problem.

The subject controlled the progression through the session by clicking with the mouse on a "next task" button in the top right corner of the screen after the simulator had recognised that an acceptable steady state had been reached on the previous problem. Subjects were asked to verbalise their beliefs about the system and their reasoning behind actions. They were also asked to minimise the number of actions used to solve each problem.

After completing the first 21 tasks, the subject was given sets of semantic differential scales on which to state their ratings on various aspects of the interface and on the mental effort required to perform the task. They were also asked to answer three specific questions about the relationship between the control and state variables. The questionnaires were repeated upon completion of the second half of the session.

The problems in the second half were different from the first in that the target ranges for the state variables were smaller and placed further away from the starting point for the variables. This meant that the amount of change in the controls was often as important as choosing the correct control variable. Most subjects were able to complete all 13 of these problems in the time available. Finally the subject was interviewed about various aspects of the presentation and any other comments were recorded.

In addition to the subjective ratings of the subjects, the complete session was recorded on videotape. As the experiments took place in a professional behaviour laboratory we were able to mix video from three sources on the recorded tape. The three shots available were: of the subject's face, the screen and an overhead shot showing the mouse and keyboard activity. The sound was recorded through a microphone attached to the subject.

The final data source for the analysis was a log of all user actions and system generated events. Complete snapshots of the system state were recorded so as to be able to replay sessions if necessary. The logged data was used for the quantitative analysis of performance.

2.3. Experimental Conditions

As the experiment was of an exploratory nature it was decided to use a wide range of different media interfaces with relatively few subjects in each group rather than having two or three larger groups.

Condition	Text	History Table	Sound	Spoken Warnings	Text Warnings	Graphics	History Graph
T	Yes						
G						Yes	
TMM	Yes	Yes	Yes	Yes	Yes		
GMM			Yes	Yes	Yes	Yes	Yes
TS	Yes		Yes				
GS			Yes			Yes	
GSP				Yes	Yes	Yes	
GW					Yes	Yes	

Table 1 Experimental Conditions

(T = text only, G = graphics only, TMM = text multimedia, GMM = graphics multimedia, TS = text and sound, GS = graphics and sound, GSP = graphics and spoken warnings, GW = graphics and textual warnings)

We wanted to observe as many media and combinations of media in action as we could. This had implications for the likelihood of achieving statistically significant results but it was felt that these were of secondary importance compared with the exploratory observations we would be making. There were eight different conditions with five or six subjects in each. Each subject used the same interface throughout the whole session. The conditions were T, G, TMM, GMM, TS, GS, GSP and GW as shown in Table 1 overleaf.

T. Text only.

Current values for the state variables were shown numerically with the bounding values for their target ranges. In this interface the subjects altered the controls by entering a letter indicating which control should be altered ('I' for in-valve, 'O' for out-valve and 'H' for heater) followed by a numeric value between 0 and 100 into a large command field. The current values for the control variables were displayed on the screen.

G. Graphics.

The current values for the state variables were graphically coded in a realistic impression of the waterbath simulation. The level was shown as blue water rising and falling within the tank, the temperature was shown by rising and falling mercury (which happened to be red) in a thermometer and the flow rate was shown as a scaled up bar which indicated the flowrate. The target ranges were also graphically coded with yellow marker lines indicating where the extremes of the target ranges were for each variable. The state variable values were numerically annotated on the diagram. Sliders were used for the three controls in this condition so no keystrokes were necessary.

TMM. Text Multi Media.

Numerical values of the state variables were shown in scrolling lists. These lists therefore showed the recent history of a state variable's values. Target ranges were shown numerically. In this condition the subject also received spoken warnings when abnormal situations arose. There were two types of warnings. For emergency warnings, such as when the tank was about to overflow with water and the system automatically shut off the in-valve, a female voice was used. For trend warnings, which occurred whenever a state variable was persistently moving away from its target range over a long period, a male voice was used to draw the subjects attention to that fact. The flowrate variable was indicated with a continuous sound of flowing water which varied in volume according to the value of the flow rate.

GMM. Graphics Multi Media.

A scaled-down version of the graphic impression of the system used in G above was shown together with a line graph which showed the trends and current value of each of the state variables as different coloured lines. The spoken warnings and the water flow sound were also present.

TS. Text and Sound.

This was the same as T with the addition of the water sound for the flow rate. It was included to enable us to examine the effect of sound in isolation.

GS. Graphics and Sound.

The same as G with the inclusion of the water sound.

GSP. Graphics and Spoken Warnings.

The same as G with the inclusion of the spoken warnings in the event of erroneous process states. The warnings were in fact provided both textually in the form of messages appearing at the top of the screen and verbally.

GW. Graphics and Textual Warnings.

This is the same as G except that the warnings were only given in the textual manner.

2.4. Independent Variables

With the verbal protocol analysis, the subjective ratings and comments, and logged data there are clearly a great many ways in that data can be analysed and results presented. The results are presented under main headings which correspond to dependent variables under investigation. By comparing results from individual or combinations of conditions, the following main variables can be isolated (of which all but one relate to media) .

a) The presence / absence of warnings.

By comparing conditions G vs GW and GSP, the effect of having warnings can be examined. Admittedly, the way the warnings are presented is likely to influence the overall effect of the warnings. A comparison of the two ways to present the warnings is also carried out .

b) No Sound vs Sound.

Comparing conditions T and G vs TS and GS isolates the use of sound as the only difference between the two groups.

c) No Speech vs Speech.

GSP vs GW above. An alternative grouping here could be T, G, TS, GS and GW (i.e. conditions without spoken warnings) vs TMM, GMM and GSP (conditions with spoken warnings). However, this would introduce a number of other variants which could affect the analysis, such as the presence of trend information in some conditions, the presence of sound in some conditions and the use of text (numbers) as a main source of information in conditions some and graphical representations in others. Thus comparing conditions GSP and GW is most appropriate as the only difference between the two is that in condition GW the warnings are redundantly presented with speech.

d) Text vs. Graphics.

T and TMM vs. G and GMM. There is pair-wise correspondence between the conditions in each of these groups. Within each pair, the information content and redundant media usage are the same, but the primary media for presenting the information is textual in the first condition and graphical in the second condition. To obtain an indication of whether an interface with a textual inclination or one with a graphical inclination is better for operating, these pairs can be analysed individually or the two groups mentioned above can be analysed.

2.5. Other Measures

2.5.1. Performance measures

For the performance analysis of the data logs, the completion time for a task, the number of actions needed to complete the task and the number of warning situations entered, were chosen as the dependent variables.

A) Time taken to complete the tasks

Completion time is perhaps the most straightforward measure. Obtaining a good (low) time score meant the session was completed quicker and it gave the subjects the feeling of being in control of the simulator. Thus, there was a clear incentive for the subjects to optimise their performance in this respect. Another reason for using the completion time is that the concept of time saving is extremely relevant to the process control domain.

b) Number of actions

The number of actions used is perhaps not as robust a measure as completion time. Although the subjects were asked to minimise the number of actions taken, it became very clear that in situations where there might be a trade-off between speeding up the process and minimising the actions, subjects opted to speed up the process at the cost of additional actions. However, we include this measure in our analysis and attempt to offer explanations where the number of actions seems to deviate from the other two measures.

c) Number of warnings given

The number of warning situations should provide a good indicator of how well the subjects operated the simulator. Although not all the conditions included any actual warnings, the number of abnormal situations that would cause a warning to be triggered could still be counted. A large number of such situations would indicate a poor performance. Subjects were observed doing the tasks and were debriefed afterwards. They were also required to answer value judgement questionnaires about the interfaces.

2.5.2. User opinion of the interface

a) Subjective rating questionnaires

In the middle, and at the end of each session, users were asked to fill in a questionnaire to rate all the different aspects of the interface (graphics, text, sound, speech etc.) on a number of bi-polar adjective scales. These were then analysed and compared between the different interfaces to find out the users perception of the strengths and weaknesses of the different interface styles. The subjective rating scales used were those reported by Boucherat et al, 1990, as part of their study on the specification of evaluation tools and methodology. A list of bipolar adjective pairs, for example 'pleasing - irritating', on a scale ranging from 1 to 7 was presented. The positioning of the positive and negative adjectives on the top end or the bottom end of the scale was balanced between the pairs on the questionnaires to prevent response bias. However when presented here the positive adjective is always placed at the top end of the scale to ease comparison.

b) Debriefing interviews

After the experimental session the subjects were interviewed and asked their views on the different styles of presentation, any tasks or procedures that they had found particularly difficult, any possible improvements they could suggest and so on. Some of the interviews were videotaped and some of them were recorded on paper.

c) Observation of testing

The recordings of the subjects taken during the experimentation were watched to glean any further information. For example:

additional comments about the interfaces
 problems the users had with the tasks
 misunderstandings/misconceptions about the system, controls or displays
 strategies used to achieve goals and avoid problems
 learning patterns

2.5.3. Measures concerning user understanding of the system

a) Verbal protocols analysis

The subjects were asked to speak out loud as they were doing the tasks, verbalising why they used the control actions that they used. They were also asked to state any connections that they noticed between the control and state variables and any other thoughts or opinions that they had about the system. The experimental sessions were videotaped and the statements concerning connections between the variables were then analysed to find the percentage that were correct. The dependent variables used for the verbal protocol analysis correspond to those used by Sanderson et al. For 50% of the first 21 problems, the number of correct and incorrect statements about a particular system or control variable was coded. The number of correct statements is expressed as a proportion of the total number of statements.

b) Knowledge state questionnaire

In the middle, and at the end of each session, the subjects were asked to complete a table which gave them control actions using each of the controls and asked them what the effect would be on each of the dependent variables. These tables were analysed to give an idea of the subjects' understanding of the effects of the different controls.

3. Results and Discussion

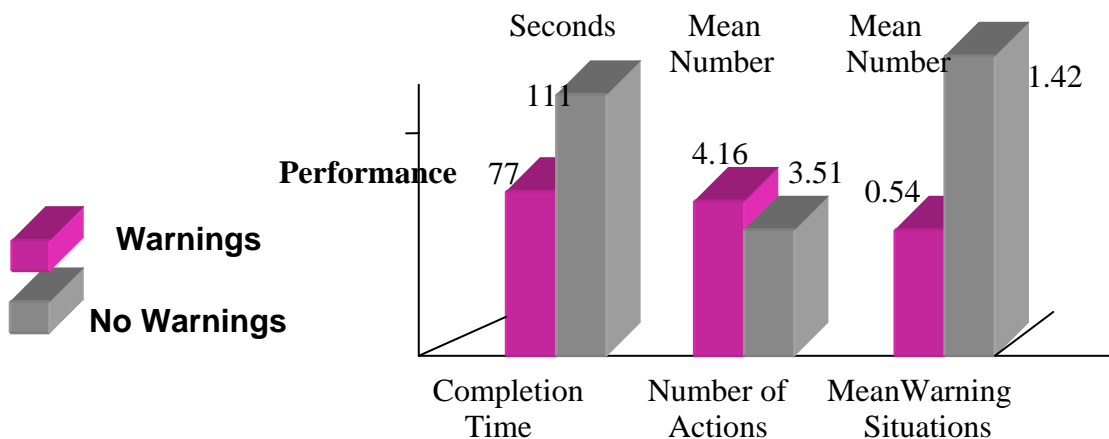


Figure 2 The Effect of Warnings on Performance

3.1. Warnings

The first effect analysed is that of receiving warnings about persistently bad trends in the state of the process. Figure 2 shows that the warnings improved performance with respect to completion time and number of situations where warnings should be given, but slightly increased the number of actions. However, the difference in the number of actions does not reach significance whereas the other two differences are significant ($p < 0.02$ for completion time and $p < 0.04$ for warning situations). The different pattern of results between the number

of actions and the other two measures raises a number of interesting issues. In most cases, receiving a warning about a bad trend simply enabled the subject to take corrective action earlier than would have been done without the warning. Following that reasoning, the number of actions would not benefit from the warnings as corrective action would have to be taken at

some point in any case in order to solve the problem. However, the tendency for the number of actions to be greater in the warnings situation may indicate that warnings can elicit or trigger an initial pressurised response which may not be the best or even correct action and must be followed up by a further corrective action. Without any warnings, subjects strayed far more frequently into situations where a warning would normally have been given. Thus warnings tend to keep a subject within a narrower envelope. When subject comprehension is examined, warnings did have a significant detrimental effect on the number of correct statements made about *Vout* and the *Outflow* (figure 3).

These are the variables which are the most difficult to comprehend for subjects in conditions both with and without warnings. Warnings may have a detrimental effect on subjects' understanding of the most difficult aspects of the system because they may be inherently disruptive to the thought process of the subjects. In many situations a subject may be in the middle of observing how the system behaves, trying to build a picture of cause and effect in the system, when a warning suddenly disrupts this process and forces a corrective action. The action taken may not always be the correct one, as seems to be indicated by the performance results, and it leaves the subject none the wiser about the underlying principles of what is going on.

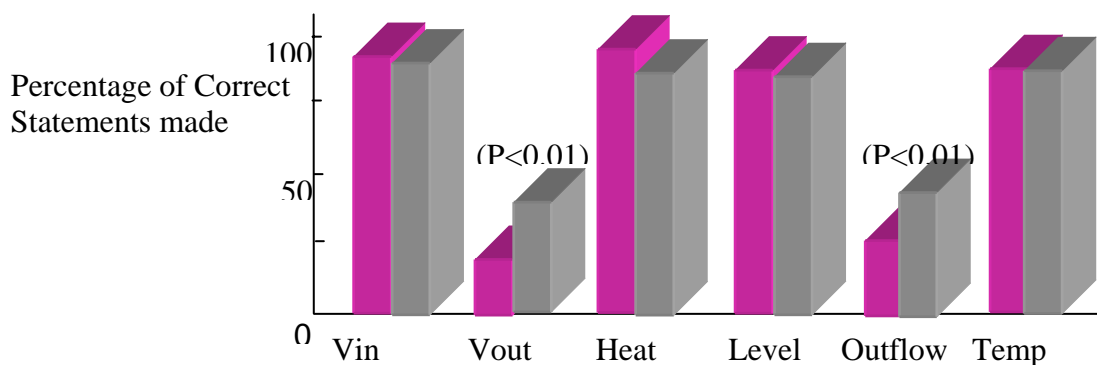


Figure 3 Effect of Warnings on Comprehension

The subjective rating scales indicated that warnings were generally received very well by the subjects. This was also evident from the interviews. However, some subjects commented that warnings sometimes persisted after they were aware of the problem (warnings were repeated at regular intervals if the bad trend that caused the initial warning had not been turned round). This was often due to the slow response of the system, particularly with respect to the temperature. However, we did observe situations where a subject thought he had corrected a bad trend but had either taken the wrong corrective action or had not altered the controls sufficiently to turn round the trend. In these cases it may have been more appropriate to use a different warning which could bring the subjects attention to the mistake.

3.2. Sound (Non-verbal)

As mentioned previously, sound was intended to provide subjects with redundant information about the flowrate. We chose to use a realistic sample of water flowing from a tap which was

played continuously and varied in volume according to the flowrate. In doing so we followed the philosophy that realistic sound would be more easily understood than would a synthesized unrealistic sound. This view has been strongly put by Gaver (1989).

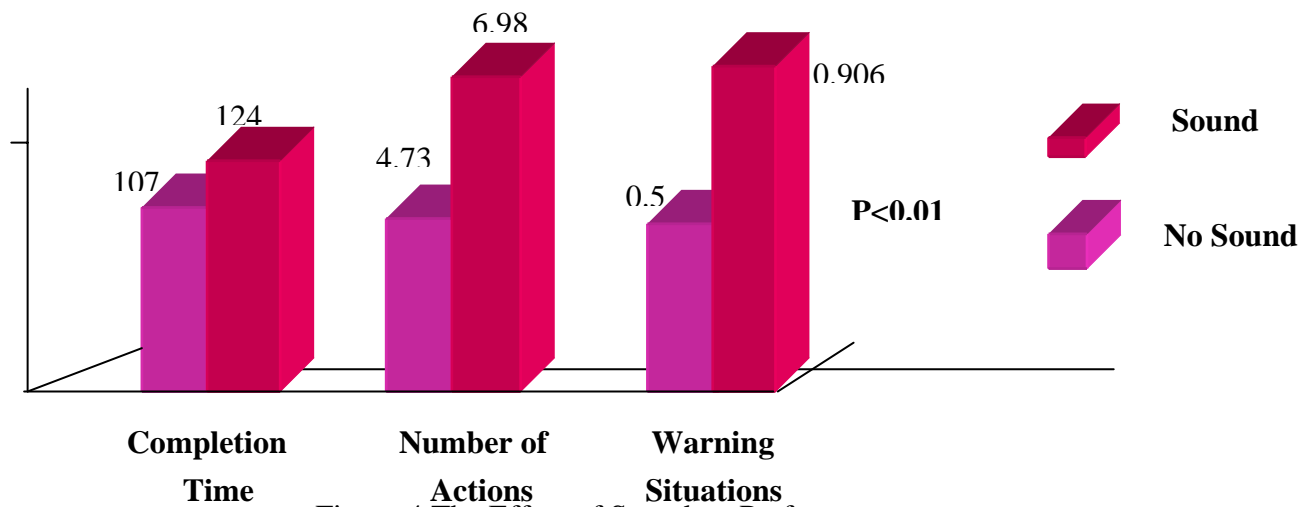


Figure 4 The Effect of Sound on Performance

The trends in figure 4 indicate that sound had a detrimental effect on performance overall. However, only the number of warning situations approaches significance ($p < 0.1$). In order to see whether the sound was equally bad for all types of task we broke down the tasks into three categories of difficulty. In category 1 we put the compatible tasks from the first half of

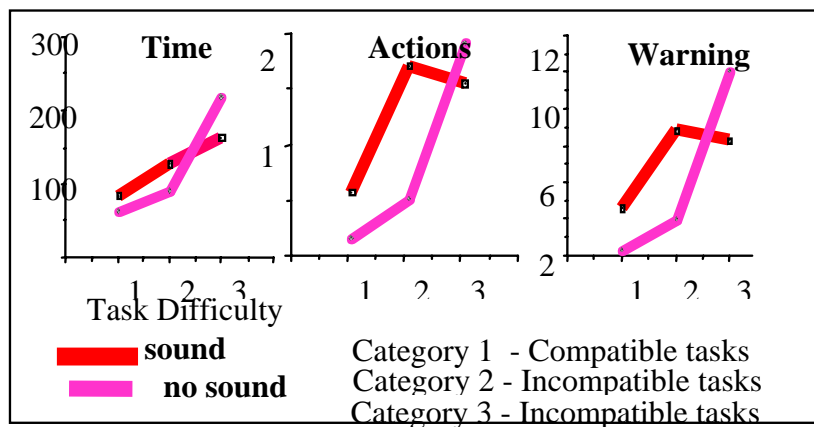


Figure 5 The Effect of Sound over Task Difficulty

the session, that is, the compatible tasks which had unrestricted target ranges for the state variables. In category 2 we put the incompatible tasks with unrestricted target ranges. In category 3 we put the tasks from the second half of the session with restricted target ranges for the state variables. These categories also correspond roughly to the order in which subjects had to solve the problems, so it may also reveal something about the learning effects. A repeated measures analysis of variance with the difficulty of the tasks as within subjects factor revealed the following pattern.

The clear pattern which emerges from the graphs in figure 5 is that operator performance with interfaces using sound was better for the difficult tasks than for the easy ones. The trend shows a marked difference between the category 2 and category 3 tasks.

However, multi-variate analysis does not show a significant interaction between the difficulty factor and the use of sound. The only dependent variable approaching significant interaction

is the number of warning situations ($p < 0.15$).

Although the trend is not significant it is tempting to speculate about what caused this trend. The fact that category 3 tasks had restricted target ranges which were often placed a considerable distance away from the starting points for the state variables generally meant that larger changes in the Outflow were required in these tasks. The larger changes in the volume of the sound may have enabled subjects to make some use of the audio information or at least made it less irritating and distracting. It may also be that since the category 3 tasks were towards the end of the session the subjects had got accustomed to it and perhaps managed to use it to their advantage.

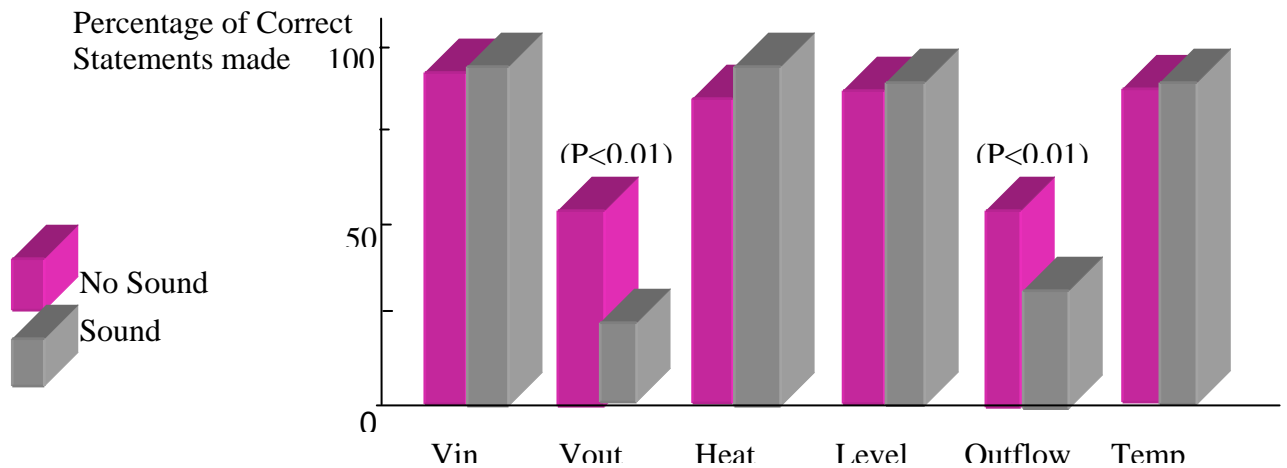


Figure 6 The Effect of Sound on Comprehension

The verbal protocol analysis of number of correct statements made under this condition (figure 6) indicates that sound decreases comprehension about the very aspects of the system in relation to which it was meant to provide redundant information. This result was wholly unexpected. At worst we expected there would be no difference. The only reasonable explanation seems to be that the sound did in fact disturb and distract people from reasoning about the system. Note that the verbal protocol analysis was only based on selected tasks from categories 1 and 2 as described above where the sound appeared to have a worse impact on performance than in category 3 tasks. Also it is important to remember that the flowrate and the out-valve were the aspects of the system which were most difficult to understand overall.

Sound was generally received badly by the subjects. Sound was thought to be less important, more difficult to use, incomplete, unclear and irritating than the speech. It is tempting to condemn sound completely based on our findings. However, it is only the way in which we used sound that seems to have had a poor impact on performance, understanding and subjective ratings. The main problem with the sound we used was that it was not possible to detect small changes in the sound volume. We were able to play the sound at 100 different volumes, but it appears that this scale was not sufficiently graded for the small changes in the flowrate which were often important. Additionally, subjects appeared unable to pick up the smallest volume changes. By opting for a sampled sound we had hoped to exploit associations with water flowing on behalf of the subjects. We succeeded in this in that no subjects were in any doubt that the sound they heard was the sound of water flowing out of the bath. However, changing waterflows produces changes in the pitch of the sound as well as the volume. We did not have sophisticated enough equipment to produce both of these. Thus we did not exploit fully associations with *changing* waterflows on behalf of the subjects which was perhaps the most important association for the subjects to make. This would

account for the sound not being a benefit to the subjects.

To account for the decrease in performance the most obvious reason is that the sound was continuous. this was clearly an irritating factor which put many subjects off the task at hand. This does not mean that sound cannot be used to convey continuous information. In the Arkola simulation described by Gaver et al. (1991) they used discrete sounds to convey continuous information about the state of a bottling plant with considerable success.

On a positive note we should not forget the trend found in figure 5. The fact that the sound became more useful towards the end of the sessions may indicate that despite the inappropriate way in which we employed the sound, people still showed signs of using it with practice. This suggests that people have a remarkable ability to utilise the audio channel.

3.3. Speech and Visual Warnings

We have seen that warnings about bad trends in the system had a positive effect on performance, particularly for the difficult tasks. Interestingly, we also found that the warnings

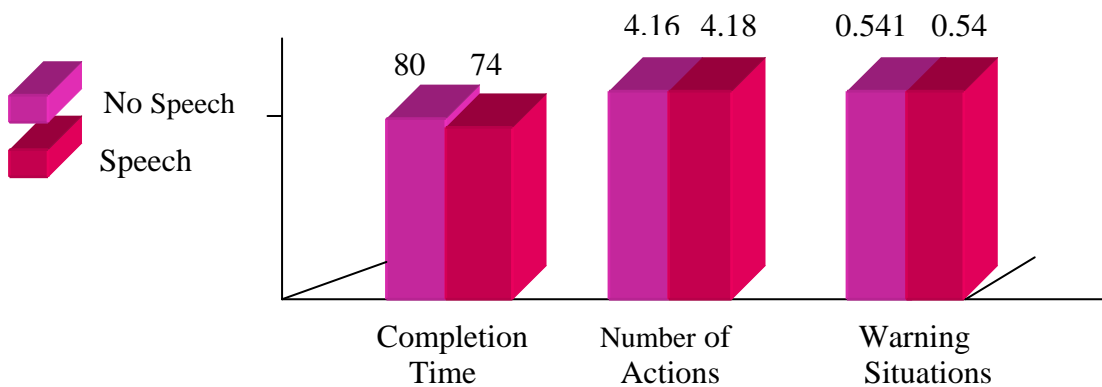


Figure 7 The Effect of Speech on Performance

decreased subjects' understanding of the system. The next question to be posed is whether the way in which the warnings were presented to the users had an impact on these effects. Here we compare the condition where the warnings were given only textually (GW) and the condition in which the warnings were spoken as well as given textually (GSP).

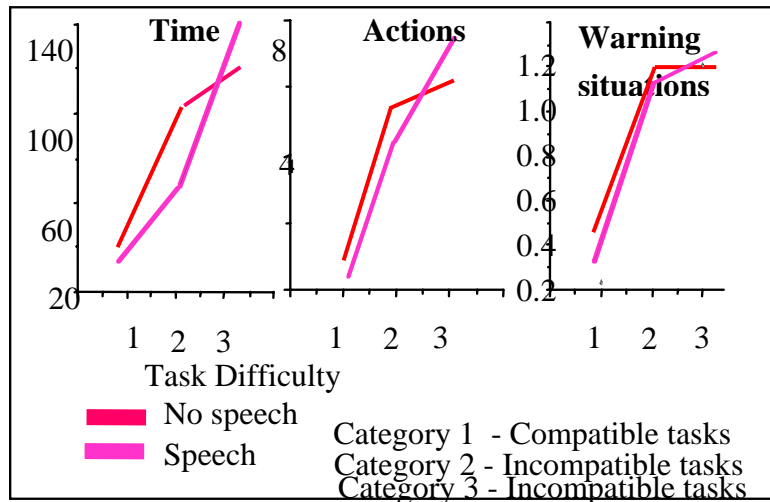


Figure 8 Performance of Speech over Task Difficulty

It can clearly be seen (figure 7) that overall, there is no difference in performance due to the way in which the warnings were given. This was not anticipated because the speech had generally been well received by the subjects and they claimed it helped them solve the problems. It appears that the real help came from the fact that they had warnings at all.

Again we inspect where the differences (if any) between the speech and the no-speech conditions lie with respect to tasks of varying difficulty. The pattern in these graphs in figure 8 shows that speech was slightly more beneficial to subjects in the earlier/easier tasks. However, the interaction between difficulty and use of speech is only significant for completion time ($p < 0.05$). Nonetheless, it is possible to offer a plausible explanation for this pattern. In the beginning, the subjects react slightly better to spoken rather than printed

warnings. This is probably due to the fact that a spoken warning is seldom (if ever) completely missed. As the session goes on the subjects in conditions with spoken warnings may come to rely on the warnings to help them solve the problem. They may relax in the knowledge that 'it will tell me something is wrong anyway'. As the warnings only come as a result of established bad trends, this is not an optimal way to operate the system.

We received a great deal of positive feedback about the speech. Another interesting comparison is between the subjective ratings of spoken and textual warnings.

There was no significant differences between the way subjects perceived spoken and textual warnings although there is a quite large difference with respect to the rating of importance. A closer look at the data showed that there was a bi-modal distribution for the ratings of importance for text warnings. One group of subjects had rated the textual warnings as very important and another group had rated them as rather unimportant. Further investigation revealed that these two groups corresponded very well to people who had rated the overall task as difficult and those who had rated the task as easy. This prompted a two-

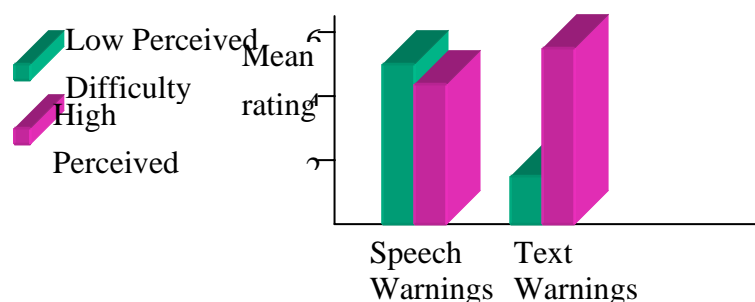


Figure 9 Subjective Ratings of Warning Medium by Perceived Task Difficulty

way analysis of variance to see if this was a significant pattern.

Figure 9 shows that this result is highly significant. There is a clear difference in the rating of the text warnings. Those who found the task easy found the text warnings unimportant and those who found it hard valued the text warnings as being very important. This difference does not exist for speech warnings which was rated as important by all subjects.

The likely reason for this result is again that the speech is very intrusive and is never missed. Another important attribute of speech is that it can be processed in parallel with the visual perception of the screen which means no extra effort is required to take in the spoken information. Thus all the subjects, irrespective of how hard they found the tasks, picked up and employed the spoken warnings. Textual warnings on the other hand can be easily missed. It also requires that the subject switches attention from operating the task to read the warning message. Subjects who found the problems easy to solve may have been more focussed at solving the problems and not paying so much attention to events in the periphery of their vision. It should be noted that the textual warnings were by no means obscure.

Thus the earlier suggestion that spoken warnings invites the subjects to rely on them to a greater extent than the textual warnings, may have to be qualified to be true only for people who find the tasks easy to perform.

3.5 Text and Graphics Differences

The final variable examined was Mental Processing Code. The natural assumption here would be that graphical representations would perform better if the task was spatial rather than verbal. The effect of mental code (Graphical or textual) on performance can be seen in Figure 10 where the three variables of Time to complete, Actions, and Warning situations are displayed. The graphically oriented interfaces (GMM and G) yield better performance than

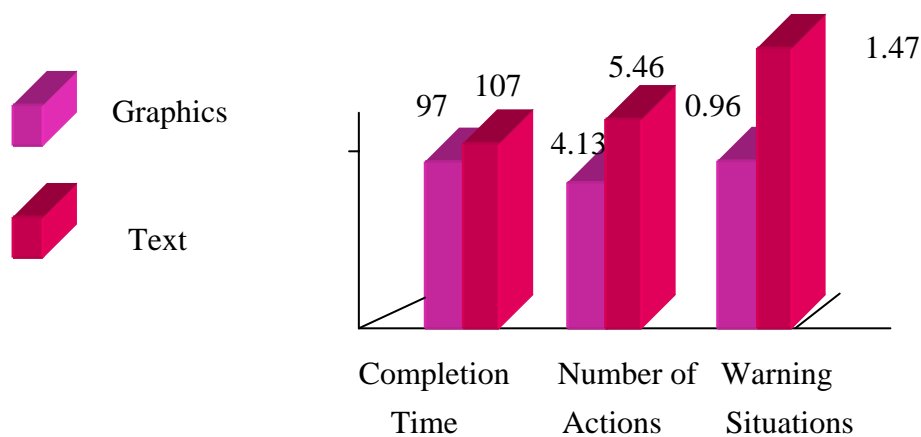


Figure 10 Effect of Mental Code on Performance

the textual ones (T and TMM). A multivariate test of significance shows the overall effect approaching significance ($p < 0.09$). However, the most dramatic difference is found in the number of actions ($p < 0.03$). It appears that the textual interface encourages fine tuning more than the graphical ones. This may be due to a lack of overview of what is going on when in the textual condition. In these conditions the subjects may have been forced into taking more actions in order to observe the effects this had on the system.

An even more interesting result is obtained when we separate out the task difficulty. As in

Figure 11.

A clear pattern emerges from the graphs in the Figure. It is that the graphical interfaces are advantageous for more difficult tasks than for easy tasks. This trend is significant for the number of actions ($p < 0.003$) and approaches significance for the warning situations ($p < 0.08$).

This superiority in graphical presentations over textual ones is also clearly shown in the comprehension tests. As usual the four variables *Vin*, *Level*, *Heat* and *Temp* pose no real difficulty for the users, but *Outflow* and *Vout* cause problems.

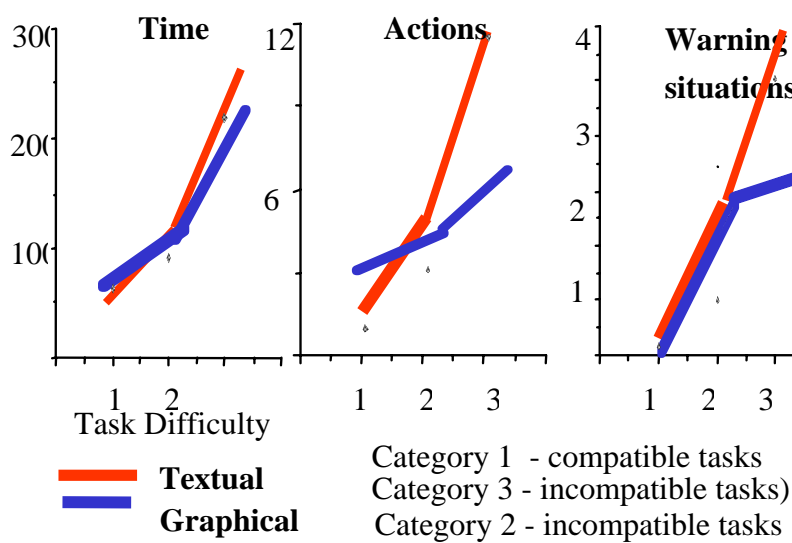


Figure 11. The effect of mental code on performance with Task difficulty

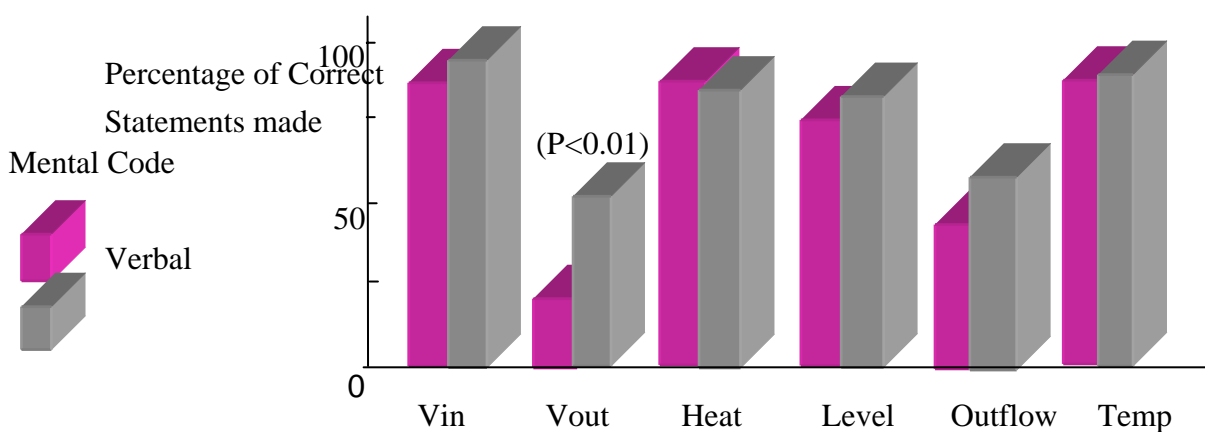


Figure 12 The Effect of Sound on Comprehension

3.6. Does the Medium Matter?

A question that may be legitimately asked about multimedia interfaces is whether they actually make any difference. Does the presentation affect the way in which information is picked up and consequently reasoned about ?

Figure 13 indicates that the presentation does indeed matter. It can be seen that in these experiments there were certain aspects of the system that was very well understood by all subjects irrespective of what interface was used. In particular, comments about the behaviour of *Temp* and the operation of *Vin* were consistently above 95% correct for all interfaces. Some insignificant variations due to the interface was found for the *Heat* and the *Level*, whereas significant variations were found for the *Outflow* and *Vout* which averaged round 50% and 30% correct statements made respectively.

It is clear that the presentation matters considerably for understanding more difficult aspects of a task. This opens up an interesting question of whether there is exists a point where a problem becomes so difficult that the type of presentation again becomes irrelevant. It seems like a reasonable proposition. If the information presented is not sufficient or pitched at the

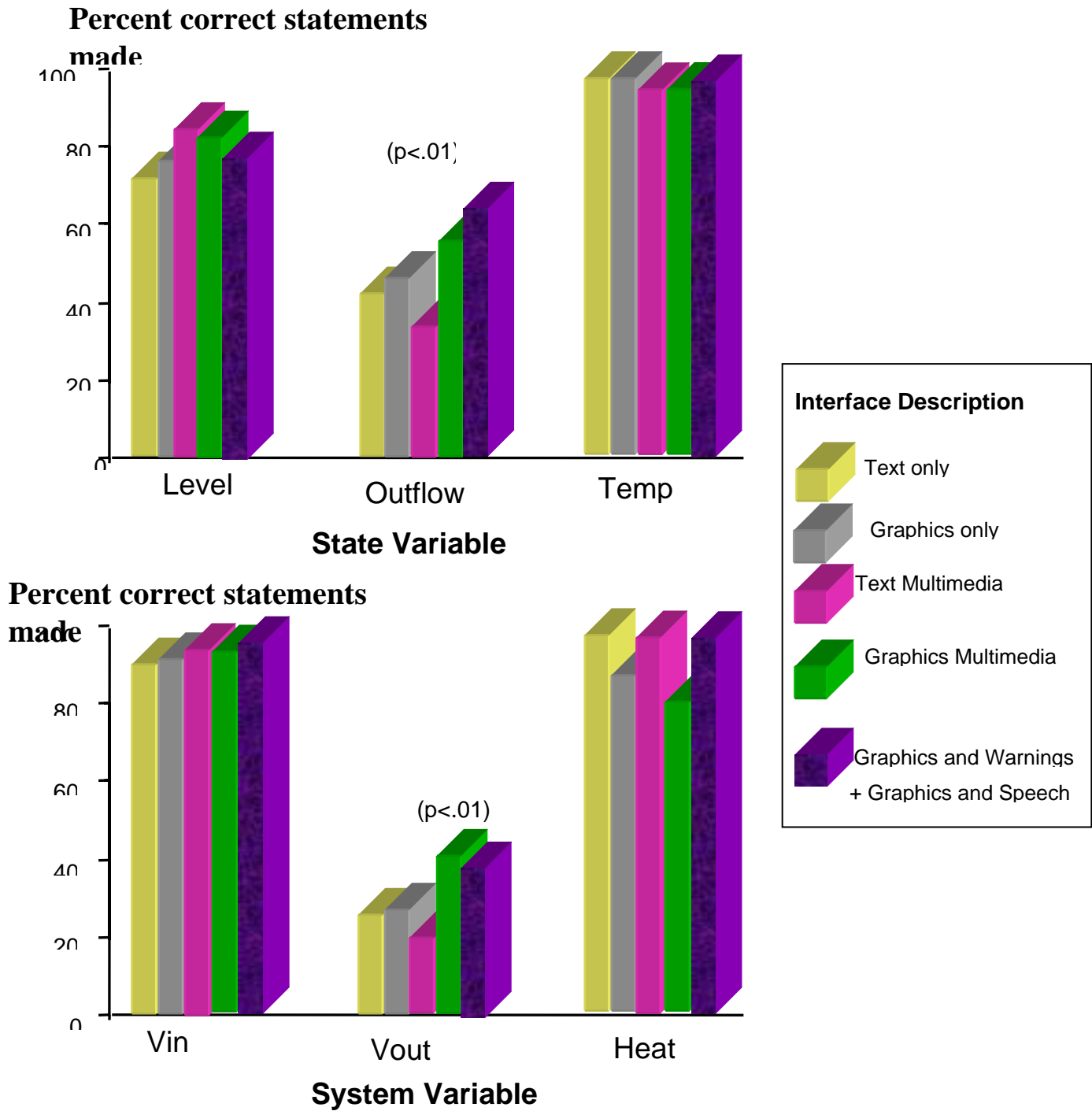


Figure 13 Effect of Media on Comprehension

wrong level of abstraction for solving a difficult task it is unlikely that whether it is presented graphically or textually or otherwise will not make a great deal of difference. This is to say that presenting the information an appropriate manner is indeed important but presenting the right information to solve the problem is still more important.

The users perception of the interfaces varied with the nature of presentation of information. The text only interface was the least popular being scored overall as the least helpful and the most irritating. Conversely the Graphics Multimedia interface received the most positive response on both scales.

4. Some Possible Conclusions from the Results

The first point to be made (which will be rather obvious to psychologists) is the difficulties involved in this type of experimentation and the dangers in assuming that Engineers and Scientists can produce optimal interface designs by chance. The adaptability of human beings, the affect of a task upon human behaviour, and the variability between human beings always makes experimentation difficult. Rarely can one come out with a simple definitive statement.

However we can make some interesting observations from the data. Warnings seem to affect comprehension because they can often intrude when a subject is trying to model the system. They are effective in assisting a subject to attain a particular system state but are not so helpful if the goal of the user is improved understanding. Warnings should be particularly useful to trained operators but should be minimised during exploratory learning. This explains why subjects found the warnings useful and yet showed a lower understanding when warning levels were higher.

Sound is generally thought to be useful in delivering warnings and for context switches. In these experiments sound was found to have a detrimental effect on performance. Even more worrying, it seemed to have a detrimental effect upon the level of understanding of the very variables (e.g Outflow) it was designed to help. However our use of a complex naturalistic sound may not have had the discriminability required. Also human beings can distinguish quite small changes in sound. However they are not good at using sound as an absolute measure. In this particular experiment it was the absolute value that also mattered as well as the change. It was important to know if the flow rate was changing in the right direction when a target limit was being approached, but the sound output changes were not sufficient even to provide for this. The continuous nature of the sound also had a deleterious effect.

Most subjects liked the speech warnings. Speech did seem to improve performance during learning but it is very intrusive and tends to dominate other information. Those who found the task easy also found the textual warnings irrelevant, Those who found the task difficult found the textual warnings important. However this was not followed through with respect to speech. Even those who found the task easy regarded the speech warnings as important. It seems that speech warnings, in themselves, command special respect possibly because of parallel processing of audio information.

Finally, the media used for presentation does indeed matter. Those aspects of the system that were particularly difficult to understand benefitted greatly from an appropriate presentation confirming our hypothesis that higher bandwidth for a difficult concept should be of benefit.

5. Acknowledgements

We would like to thank all our partners in the PROMISE consortium - The Chemical Engineering Department at The University of Leuven (Belgium), Scottish Power in Glasgow, DOW Benelux in Terneuzen (Netherlands), EXIS (Italy) and Tecsiel (Italy). We also gratefully acknowledge support in project No. 2397 from the ESPRIT programme of the European Commission.

6. References

- J. L. Alty and C.D. McCartney, Design of a Multi-Media Presentation System for a Process Control Environment, in *Multimedia: Systems, Interactions and Applications*, L. Kjelldahl (Ed.), Springer-Verlag, Berlin, 293 - 306 (1991).
- E. R. Crossman, and F. W. Cooke, Manual Control of Slow Response Systems, in *The Human Operator in Process Control*, E. Edwards and F. Lees (Eds.), London: Taylor and Francis, 51 - 64 (1974).
- P. M. Sanderson, A. G. Verhage and R. B. Fuld, R. B., State-space and Verbal Protocol Methods for Studying the Human Operator in Process Control, *Ergonomics* **32(11)**, 1343 - 1372 (1989).
- N. Moray, P. Lootsteen, and J. Pajak, The Acquisition of Process Control Skills, *IEEE Transactions Systems, Man and Cybernetics*, **SMC-16**, 497-504 (1986).
- N. Moray, and I. Rotenberg, Fault management in Process Control: Eye Movement and Action, *Ergonomics* **32(11)**, 1319-1342 (1989).
- P. R. Boucherat, R. H. Benton, H. Shahnavaz, E. Wang, Emulation of a multimedia workstation for the investigation of integration principles in design for usability, *Human Factors in Telecommunications, Proc of the 13th Synopsium*, Torino, Italy, 253-260 (1990)
- W. W. Gaver, The SonicFinder: An Interface that uses Auditory Icons, *Human-Computer Interaction* **4**, Lawrence Erlbaum Associates Inc., 67-94, (1989).
- W. W. Gaver, R. B. Smith and T. O'Shea, Effective sounds in Complex Systems: The Arkola Simulation, *Proceedings of the Conference on Human Factors in Computing Systems*, CHI '91, 85-90 (1991)