

Comparing speech versus text displays for alarm handling

N.A. Stanton and C. Baber †

Department of Psychology
University of Southampton,
Highfield,
Southampton.
SO17 1BJ

†Industrial Ergonomics Group,
School of Manufacturing &
Mechanical Engineering,
University of Birmingham,
Birmingham.
B15 2TT

phone: 01703 592586

fax: 01703 594597

email: nas@soton.ac.uk

The rise of computers in command and control domains has meant that control operations can be performed via desk-based visual display terminals. This trend has also produced the potential to display information to operators in a variety of formats. Of particular interest has been the use of text-based displays for alarm presentation. There are possible limitations to the use of text for alarm presentation, not least of which is the need for a dedicated alarms display screen (or, at least, a display page). Given the capability of computers to synthesize speech, it is possible that speech-based alarms could generate the same information as text-based displays without the need for dedicated screen space. In this paper an experimental comparison of speech-based and text-based displays for presentation of alarms is reported. The findings show that speech leads to longer response times than text displays, but that it has minimal effect on the efficacy of fault handling. The results are discussed within the alarm initiated activities framework and implications for alarm system design are outlined.

KEYWORDS: Alarms, Speech-displays, Text-displays, alarm handling

1. Introduction

Alarms are an essential part of the control interface in a wide range of domains. Their function is to inform operators when certain system parameters exceed tolerance limits, e.g., in industrial process control, or when changes in system state are approaching critical levels, e.g., the approach of enemy aircraft in military surveillance. The information in an alarm is triggered by a discrete event which is to be communicated to the operator, whereas in other displays (such as engine temperature, reactor pressure, radar detection) information is continuously available (or can be made available on demand) and is constantly being updated. In order for alarms to fulfil their role effectively, certain criteria need to be met. The 'trigger' parameters for the alarm need to be defined and tolerance limits set (Usher, 1994). More importantly, the alarm media need to be designed to communicate the information effectively to the operator. Effective communication requires a display which presents an unambiguous message in a form which can be easily and quickly comprehended by the operator. There is still much debate on the best way to achieve effective communication of alarm messages (Stanton, 1994, a). Typically, alarm information is communicated to the operator using a variety of media, for example, auditory alarms, annunciator panels (translucent tiles on which alarm messages are inscribed, which are illuminated when the alarm occurs), mimic displays (a graphical depiction of the process, plant or system highlighting the item affected), text-based displays (textual presentation of the alarm message on a visual display screen).

With the increased use of computers for command and control in systems with limited space, e.g., reconnaissance aircraft and submarines, and with the trend towards decreasing the size of control rooms in industrial process control, there is a move away from traditional hard-wired displays. This means that annunciator panels and mimic displays are being replaced by computer-based displays. In some cases, computer screens are used to display information in a format which was previously

presented on a large display, e.g., screen-based mimic displays. In other cases, new forms of display are being employed, e.g., the substitution of text-based displays for annunciator panels. Given the amount of information presented on visual display units (VDUs), it might be felt that providing additional information, e.g., alarms, would place an unwarranted burden on operators. The process page displays on VDUs in petrochemical industries, for instance, may contain up to 800 pages of information (Stanton, 1991). Providing a page solely for alarms in such a system could easily lead to the information being missed. By way of compensation, contemporary systems may employ three VDUs for each workstation. One of the VDUs provides overview information, often including alarms, while the others are used for control operations. In addition to the 'three VDU workstation', control rooms are characterized by the need for operators to move between workstations, either to consult other operators or to cover for colleagues. Thus, systems may come about which have a very high visual information load and which have a requirement for operator mobility. This would appear to be an excellent candidate for auditory alarms. However, given the variety of alarms which can be anticipated in the control room, it is unlikely that a set of auditory tones would suffice; rather there would be a need to employ alarms based on speech. In this paper, a comparison of text-based and speech-based alarms displays is conducted in order to evaluate the potential for speech in control room operations.

1.1. *Verbal alarm media* □ □

Alarm media can be distinguished on the basis of three properties: channel, access and duration (Stanton *et al.*, 1992). The channel refers to the human sensory modality of the communication, e.g., auditory or visual. In the case of speech and text display, although alarm messages can have exactly the same content, their channel will obviously differ. The second factor, access, refers to the means by which operators can call-up, or access, the alarm message, e.g., the message could be presented without operator intervention or could be made available for the operator to call at a later time.

It is, of course, possible to combine the two. However, an important factor in access is how the operator is alerted to the fact that an alarm message is available. To a certain extent, this returns us to the discussion of channel; it might be anticipated that the auditory channel would be more alerting than the visual channel, especially in a task involving high visual load. The third factor, duration, refers to the time for which the message is presented. Scrolling-text displays are displaced by incoming messages, but remain on the screen (at least, until they reach the bottom of the list). Speech, on the other hand, is transitory. This means that, unlike text-based displays, it needs to be attended to immediately, otherwise it will be lost (unless the operator can call for the message to be repeated or the message is written down, but writing materials may not always be to hand or easily accommodated in the workspace).

Verbal alarms, whether text-based or speech-based, share several characteristics, e.g.,

- they normally consist of several alarms, rather than a single message;
- the time of their occurrence can be important for diagnosis;
- they contain complex information, rather than simple messages;
- they tend to be grouped, e.g., in terms of time of arrival or in terms of the location of the fault;
- they tend to arrive in bursts of messages relating to a specific alarm.

Stanton (1991) points out that few alarms are genuine (as little as one percent of the alarms presented in power generation and manufacturing domains were genuine, the remainder were either 'false alarms' or confirmatory messages). If speech was reserved solely for genuine alarms, then it might prove effective. However, the main problem for speech-displays is auditory clutter, which will occur when several alarms are presented simultaneously (Baber et al., 1992).

Text-based displays present alarm messages on a VDU page within a list of other messages. Typically, the message will contain the time of occurrence of the alarm, the plant area and item affected, the nature and severity of the fault, the current status of the message (whether it has been accepted or not). Often the most recent message is presented at the top of the display and the list scrolls downwards, although this is not always the case and there is no research advising which format is most appropriate. To accept a message, the operator is required to select the relevant message and press an 'accept' key. This action changes the status of the alarm message, from 'unaccepted' to 'accepted'. It is worth noting that the alarm-message list can cover several pages. Thus, the text-based display represents a signal to which the operator responds. The operator then seeks further information from a graphical display. □

If the role of the alarm display was simply to capture the operators' attention and specify the location of a fault, then it might be felt that an auditory display would have superior conspicuity to visual displays, e.g., the auditory display would be omnidirectional and would break-in on whatever activity the operator was performing. Thus, in workstations with high visual load and minimal physical space, auditory alarms could be extremely useful. Cowley *et al.* (1990) provide a set of eleven guidelines for incorporating speech rather than text into an application. Briefly, they compare speech with text displays and suggest that if a message is short and simple, and if it calls for immediate action without the need for remembering the message, then speech should be used in preference to text. □□ Typically, alarms are supposed to elicit immediate responses from operators and are intended to draw the operators' attention to alternative information sources. Thus, speech-based alarms would appear to be preferable to text-based alarms on these criteria.

1.2. Multiple resources and verbal alarm displays

It has been suggested that verbal information is an inappropriate medium for task involving fault diagnosis (Robinson and Eberts, 1987). This suggestion is based on the notion that operators' knowledge of the system in which they are working is primarily spatial in nature. In this respect, the display of the system is most effective if it is graphical, e.g., in a mimic display. However, this does not tell us how an operators attention should be attracted to a fault. While colour coding or flashing could be used to call attention to an item on a specific page, given the number of pages with which the operator has to deal, it is important to first call the page to the screen. Automatically calling up a page with an alarm will have a number of obvious and severe consequences for operator performance. Thus, there is a trend to provide text-based alarm displays, which provide the operator with the information that an alarm condition is present and with information regarding the alarm parameter in question.

An interesting question is whether it might be anticipated that a verbal alarm will interfere with the performance of a visual-spatial task, such as monitoring an industrial process. Payne *et al.* (1994) have demonstrated that spoken messages can interrupt some types of visual-spatial task but have no effect on others. Furthermore, the extent of disruption appears to be related to the intelligibility of the speech used. For instance, spoken messages with an intelligibility of less than 40% impaired performance on tasks involving either mathematical reasoning or spatial decision making (i.e., comparison of the size of shapes to say if they are the same or different). However, there was no impairment when the intelligibility level rose above 50%, nor was the impairment on tasks involving tracking. The explanation of these findings is that when intelligibility falls below 50%, attending to spoken messages represents a demand on verbal attention (i.e., in terms of multiple resource theory, Wickens, 1992), and this demand will impair performance on tasks with a decision making component. This introduces the question of how synthesized speech can be expected to impair performance.

Baber *et al.* (1992) have shown that performance using synthesized speech is inferior to that using human speech, especially if the tasks involve processing of the message rather than simply spotting words in the message. Thus, synthesized speech can be considered analogous to low intelligibility speech, and it might be anticipated that it will have an impact of performance of decision making tasks, such as fault diagnosis. In order to explore this issue with reference to the design and use of alarm displays, it is important to consider the nature of alarm handling.

1.3. *Alarm initiated activities*

Alarm initiated activities (AIA) is a generic description of alarm handling behaviour by operators. It has been developed on the basis of research in a variety of industries (Stanton *et al.*, 1992, Stanton, 1994 b; Stanton and Baber, 1995), and is illustrated by figure one.

FIGURE ONE ABOUT HERE

Figure 1. Alarm initiated activities

AIA defines seven stages through which alarm handling can progress. The initial stage (observe) involves the operator detecting the alarm. Stanton and Baber (1995) argue that, in this stage, the operator can be either active, i.e., searching for alarm information, or passive, i.e., receiving alarm information as it is displayed. At this stage, one might anticipate differences between the speed with which an operator will respond to text-based or speech-based displays.

The second (accept) stage of AIA involves the operator accepting the alarm and hence, changing its status. Following the accept stage, the operator then proceeds to the third (analyse) stage. In this stage, the operator decides on the appropriate course of action.

Often accept and analyse occur simultaneously, i.e., a 'nuisance alarm' occurs, the operator analyses the alarm and accepts it to turn it off. If necessary, the operator will proceed to the fourth (investigate) stage, in which the source of the alarm will be sought and its cause diagnosed. From the above discussion, one might anticipate that there will be a difference in performance at this stage, when using speech or text based displays. Once an alarm has been diagnosed, the fifth stage is to perform some corrective action on the process in order to remove the alarm condition, and then to monitor the change in the process, before resetting the alarm.

This brief discussion of AIA has indicated that there may be differences between alarm media at specific stages in alarm handling. This means that an alarm medium may be useful for one stage, but less useful than another. One potential solution to this problem would be to combine alarm media so that operators can draw on information which is presented in a fashion appropriate for a specific activity (Baber *et al.*, 1992, b). For example, Selcon *et al.* (1995) found that a combination of text and speech enhanced performance in a simulated missile warning approach task. In their study, Selcon *et al.* (1995) found a 'redundancy gain' of presenting more than one source of information, which they explained in terms of the parallel distributed processes theoretical framework.

2. Method

2.1. *Participants*

Thirty undergraduate psychology students at Aston University participated in this study (15 male and 15 female). The participants were aged between 20 to 24 years. They were allocated to one of three experimental conditions (see below) such that each condition contained five men and five women.

2.2. *Design*

The experiment consisted of three stages: training, data gathering and a recall test. All participants performed each stage. A between-subjects design was used, with three groups each performing the same task under different experimental conditions.

2.3. *Equipment*

An initial training video was presented on a Ferguson Videostar videocassette recorder, using a Sharp 14" colour television. The experiment was run on an Acorn Archimedes 310 microcomputer, via a Taxan 770 14" colour monitor using mouse and keyboard for input. Software was specifically written for the experiment, and performance data were collected from the computer. Synthesized speech was generated through a synthesis-by-rule program, called "Speech!", running on the Archimedes.

2.4. *Task*

The experimental task required participants to monitor a simulated industrial 'process' (see figure 2) and had been used previously in Baber (1991).

FIGURE TWO ABOUT HERE

Figure 2. Simulated industrial process.

The process represented a simple distillery, in which liquid was heated to a specified temperature and then passed through a condenser to fill a tank of distilled liquid. Participants had control over the valves (which could be opened or closed) and boiler heating (which could be set to off, low, medium, high or very high). They could also inspect the status of plant elements, e.g., tank levels, valve positions and boiler temperature. Inspection was performed by calling up the display for a specific element. The display disappeared after 3 seconds on the screen. Using this information, participants were required to produce as much distilled liquid as possible. Output was shown in the final tank and the total units were measured.

There were several parts of the process which could fail: the source liquid could run out, the supply pipe could crack, the temperature of the boiler could be outside limits, the coolant liquid could run out. Each failure had an alarm associated with it. When an alarm occurred, participants were required to accept the alarm and then investigate and correct the fault. Timeliness in fault recovery was of paramount importance to overall performance on the task. As fault correction times increase, so output performance would decrease correspondingly.

In addition to the principal task of monitoring the process, participants were also presented with a spatial secondary task. Baber *et al.* (1991) had previously found this task to be affected by process monitoring. Briefly, participants were presented with figure rotation task (pairs of 'stick people' holding circles in either their left or right hands, one of the pair being rotated in either the X-Y or Z plane relative to the other). Figures appeared in a window at the top of the screen. Examples of the figures are given in figure 3. Participants had to indicate whether the pairs were the same or different by pressing the 'same' or 'different' keys, which were the left and right mouse buttons respectively. The secondary task provided a monitoring task,

which had changing parameters and required monitoring. The intention was to add a monitoring task to the process control task in order to reflect the demands of monitoring normally associated with human supervisory control systems.

FIGURE THREE ABOUT HERE

Figure 3. Examples of figures used in the secondary task.

2.5. Procedure

The procedure for the experiment comprised seven steps, which can be summarized as follows:

Stage One - Training

1. Participants were presented with a ten minute video showing a training programme describing the use of the software and the task to be performed.
2. Participants practiced the task using the software in its 'unmasked' state, i.e., with all plant information displayed. Practice was performed to criterion, which was set at 80% of maximum output. The training session ensured that participants could recognize the synthesized speech, which being phoneme-driven, was comparable to commercially available speech synthesis systems. The list of the alarm messages presented are indicated below:

- Boiler temperature low
- Boiler temperature high
- Tank one empty
- Tank two empty
- Tank three empty
- Tank four empty
- Pipe eight damaged

- Boiler empty
- Output tank empty
- Output tank full
- Condensor empty
- Product to waste
- Furnace tripped

In the 'speech' condition these alarm messages were presented aurally, whereas in the 'text' condition the alarm messages were presented on the screen in an alarm message box. In the 'speech and text' condition both presentation media were used.

Stage Two - Experiment

3. Participants were read a set of standard instructions, explaining the nature of the task and informing them that the plant would be run in a 'masked' state. This would require participants to call up information regarding the status of plant information, using 'inspect' commands.

4. Participants monitored the process, and were presented with an unanticipated, unpracticed emergency. Participants were expected to deal with one emergency and return the process to normal operation. During the task the spatial reasoning task as and when the demands of the primary task allowed.

5. Participants were debriefed on the nature of the study for a period of approximately five minutes.

Stage Three- Recall test

6. Participants were presented with an unanticipated recall test, in which they had to recall as many of the alarms presented during the experiment as possible.

7. Participants were thanked for their participation and given an opportunity to discuss the nature of the study should they so wish.

2.6. Measurement

Every input was logged automatically by the software. In addition, the alarms generated and the 'process output' per unit time was also logged, as was final output. Furthermore, participant response times to alarm messages were recorded. The times to diagnose a fault and restore normal process conditions were also recorded. From the data recorded, it was possible to determine whether participants had engaged in any inappropriate activities or had made any keying errors during the experiment. Finally, the accuracy of performance on the secondary task was recorded and participants' performance on the recall test noted.

2.7. Analysis

Performance data were analysed by one way analysis of variance(ANOVA), with alarm media compared for output performance and times to respond to all alarms, and time to respond to and recover from the pipe break alarm as between-subject variables. The data on inappropriate activities and recall performance were analysed using a Mann-Whitney test.

3. Results

The results section will be divided in terms of the analyses performed. Results for the output performance are presented first, followed by the time data. Finally, results for the recall test and comparison of inappropriate activities will be presented.

3.1. *Output performance*

The level of output produced by participants in each of the three conditions was recorded and the mean and standard deviation of these data calculated. For the speech alarm condition, the mean output level was 38.6 (sd = 32.5) units. For the text alarm condition, the mean output level was 74.3 (sd = 26.6) units. For the combined text and speech alarm condition, the mean output level was 64.85 (sd = 31.2) units. However, there is a high level of variation in participants performance, and an ANOVA was conducted in order to ascertain whether the apparent difference between groups was statistically significant. These data were compared using a one-way ANOVA. The results are shown in table 1.

TABLE ONE ABOUT HERE

Table 1. Analysis of variance table for output performance

From table 1, it is clear that a significant main effect ($F = 3.74$; $p < 0.05$) can be claimed for the output level data. Post-hoc analysis of the data, using a Scheffé F-test, indicated that this difference was confined to the speech versus text conditions ($F = 3.486$; $p < 0.05$). Thus, we can say that participants in the speech alarm condition produced significantly lower levels of output than those in the text alarm condition. However, the combined text and speech alarm condition was not significantly different to the other conditions.

3.2. *Alarm handling*

This section reports the results of three measures which examine participant's ability to handle alarm information presented via different media.

3.2.1. *Time to accept alarms*

The time interval between an alarm appearing and a participant manually accepting the alarm were logged by the computer. The mean time to accept an alarm in the speech alarm condition was 11.68 (sd = 12.3) seconds. The mean time in the text alarm condition was 0.29 (sd = 0.2) seconds, and the mean time in the combined text and speech alarm condition was 1.74 (sd = 2.5) seconds. Again, there would appear to be a clear distinction between conditions. This proposal was tested using a one way ANOVA. The use of ANOVA is justified by the standard deviation being proportional to the mean in all cases.

TABLE TWO ABOUT HERE

Table 2. Analysis of variance table for alarm accept times

Table two shows a significant main effect ($F = 7.418$; $p < 0.005$). A post-hoc, Scheffe F-test was used to explore this result further. Significant differences were found between speech versus text alarm conditions ($F = 6.475$; $p < 0.05$) and between speech versus combined speech and text alarm conditions ($F = 4.933$; $p < 0.05$). There was no significant difference between text vs. combined speech and text conditions.

3.2.2. *Time taken to analyse common alarms*

The time taken to diagnose common alarms appeared to quite similar across conditions, e.g., the mean time with speech alarms was 7.9 (sd = 5.7) seconds, with text alarms it was 6.5 (sd = 3.6) seconds and with combined speech and text alarms it was 10.9 (sd = 6.3) seconds. A one way ANOVA confirmed that there was no significant difference between conditions on this measure.

3.2.3. *Time spent correcting common alarms*

The mean times for this measure were also similar, e.g., the mean time with speech alarms was 11.98 (sd = 6.8) seconds, with text alarms it was 7.74 (sd = 4.2) seconds and for combined speech and text alarms it was 12.7 (sd = 6.4) seconds. Thus, one might anticipate any differences to favour the text alarm condition, but the variation in the data is sufficiently large to confound this effect. Indeed, no significant difference was found between groups using a one way ANOVA.

3.3. *Fault handling*

This section reports measures relating the participants ability to deal with a fault presented under all conditions, i.e., a break in pipe 4.

3.3.1. *Time to accept break*

The mean time to accept the break was 8.74 (sd = 5.0) seconds for the speech alarm condition, 0.67 (sd = 0.6) for the text alarm condition and 1.4 (sd = 1.3) for the combined text and speech alarms condition. Comparison of these data was performed using a one way ANOVA and the results presented in table three.

TABLE THREE ABOUT HERE

Table 3. Analysis of variance table for fault accept times

Table three shows a significant main effect ($F = 21.986$; $p < 0.0001$). A post-hoc Scheffé-F test was used to compare pairs of conditions, and significant differences were found between speech versus text conditions ($F = 19.015$, $p < 0.05$) and speech versus combined speech and text conditions ($F = 15.674$, $p < 0.05$), but not between text vs. combined speech and text conditions.

3.3.2. *Fault investigation time*

The mean times to investigate this fault appeared to be slightly different, e.g., 4.7 (sd = 7.8) seconds for speech alarm versus 1.2 (sd = 1.5) seconds for text alarm versus 1.2 (sd = 0.6) seconds for combined speech and text alarms. The time to investigate the fault with speech alarms is over twice that of the other conditions. However, one way ANOVA indicated that the difference is not significant.

3.3.3. *Fault correction time*

The mean times spent correcting from this fault appeared to differ, e.g., 10.86 (sd = 9.1) seconds for speech alarm versus 4.4 (sd = 2.8) seconds for text alarm versus 3.6 (sd = 1.9) seconds for combined speech and text alarms. A one way ANOVA was used to test this difference, which was statistically significant ($F = 4.916$, $p < 0.05$) and

a post-hoc Scheffé-F test did indicated a significant difference between the speech versus the combined speech and text conditions ($F = 4.236, p < 0.05$).

3.4. *Inappropriate actions*

The number of inappropriate actions was compared between conditions using a Kruskal-Wallis one way ANOVA. The mean ranks were 21 (speech condition), 14 (speech and text condition) and 10 (text condition). The resulting H value was significant ($H = 8.661; p < 0.05$) and the data were further analysed using a Mann-Whitney U-test. From this analysis, we found a significant difference between speech versus text conditions ($U = 15.5, p < 0.01$) and between speech versus combined speech and text conditions ($U = 22.5, p < 0.05$). This shows that more inappropriate actions were performed in the speech condition.

3.5. *Secondary task performance*

There did not appear to be a significant difference on performance on the secondary task, in terms of accuracy. Data were analysed using Kruskal-Wallis one way ANOVA and found to be non-significant. This is taken to suggest that the secondary task gave an equal loading in each condition.

3.6. Recall test

Comparison of performance on the recall test did yield interesting differences. A Kruskal-Wallis one way ANOVA produced an H, corrected for ties, of 11.571, which was significant, $p < 0.005$. The mean ranks were 8 (speech condition), 18 (speech and text condition) and 20 (text condition). Post-hoc analysis was performed using Mann-Whitney U-test. Significant differences were found between the speech versus text conditions ($U = 7, p < 0.001$) and between speech versus combined speech and text ($U = 18, p < 0.01$). This shows that participants in the speech condition were less able to recall alarm messages.

3.7. Summary of results

The results of the preceding analyses are summarized in table 4. The table shows that relative significant differences between conditions on the dependent variables recorded in this study.

TABLE FOUR ABOUT HERE

Table 4. Summary of statistical comparisons of data

The level of significance is indicated by the following convention. "ns" indicates that no significant difference was found between conditions, "*" indicates a significant difference at the $p < 0.05$ level, "***" at the $p < 0.01$ level and "****" at the $p < 0.001$ level.

4. Discussion

The study reported in this paper compares two verbal alarm media. The media differ in terms of channel and duration, but access and content were similar across conditions. The study demonstrates a significant effect of channel on performance. Possible explanations for this finding are the quality of the synthesized speech, the intrusiveness of the speech on the control task or participants waiting for the speech to finish before taking action. We believe that the quality of the synthesized speech was not the cause of the performance difference for two reasons. In a study comparing human with synthesised speech, Baber *et al.* (1992) reported that there were no differences between the two conditions for a matching task. The training session enabled participants to learn to discriminate between the messages and identify the variable to which they referred. However, it is clear that performance using text-based and speech-based displays differs largely at the 'Accept' stage of AIA. One might feel that this was simply

a function of message length, with the synthesized speech display taking an appreciable time. In other words, the results could have been confounded by differences in duration. Indeed, Ito *et al.* (1989) have shown that people often wait until a synthesized display is completed before proceeding with their task. Certainly, this effect was observed during this study. However, given the nature of the alarm messages, the duration of the synthesised speech message did not exceed three seconds in length. Thus, while one would anticipate possible time effects, there is somewhere in the order of a ten-fold difference between speech and text displays. This implies that the difference is due to more than simply duration.

The fact that there was a significant difference in recall when speech was used implies that the synthesized speech was of sufficiently low intelligibility to cause participants problems (see also Baber *et al.*, 1992). However, the problem arising from intelligibility appeared to be resolved within the acceptance stage and did not have an impact on subsequent performance. From the work of Payne *et al.* (1994) one might anticipate that the speech display would lead to inferior performance on tasks involving decision making. We found significant effects in terms of both reaction time and activity. Specifically, the speech condition made significantly more inappropriate actions than the other groups.

The differences in performance could be explained, in part, by an appeal to multiple resource theory. For instance, the intelligibility of the synthesized speech represented an attentional demand which had an impact of task performance, both in terms of time and activity. However, the fact that there was little effect, in terms of performance time, on stages other than accept suggests that participants were not using the information in the alarm message to perform subsequent tasks. However, the notion of attentional demand does not allow extension to consider whether a similar effect would occur with speech of higher intelligibility. Certainly the work of Payne *et al.* (1994) would imply

that simply making the speech more intelligible would increase performance. However, we are not sure that this would necessarily be true for this experiment.

Given the fact that participants were able to understand the speech-based display (and that it was of similar quality to other commercially available synthesised speech), then one could argue that the speech had sufficient functional intelligibility for the task. Given also the problems which will arise if the speech-based display becomes too similar to human speech (Baber *et al*, 1992), one might anticipate that simply improving the quality of the speech may not be possible.

There is some evidence to suggest that irrelevant speech may have adverse effects upon performance (Salamé and Baddeley, 1982; Smith, 1989). The effects appear to occur independently of intensity, within the range of 55-95 dB (A). Consider the participant working to maintain the process whilst speech alarms are being presented. The alarm information may be described as 'irrelevant' if it does not relate to the particular task in hand, and performance is disrupted. Task relevancy is an important concept. Whilst alarm information may be important for the overall task of maintaining plant production and efficiency, it may not be immediately relevant to the sub-task being performed at the specific time that the alarm is presented. Stanton & Baber (1995) identify two types of alarm observation activities: active extraction and passive reception. Active extraction is characterized by operators sampling displays, searching out information that relates to their current goals and intentions. Passive reception is characterized by the alarm attracting the attention of operators and forcing them to focus on changes in the system's status. Arguably, the participant has no control over the sampling of speech alarm messages, this can force them into the passive reception mode. The lack of control over sampling and task irrelevancy of information presented may lead to increased attentional and memory demands. These increased demands may lead to a general performance decrement. The differences in recall performance

indicate that participants in the 'speech' alarms condition did not commit the messages to memory to the same degree as the participants in the 'text' alarms condition.

The main difference in performance occurs at the accept stage, resulting from a delay between the alarm and action in the speech condition. This delay can be said to parallel a phenomenon termed 'cognitive lock-up' (Moray and Rotenberg, 1989). Operators might observe abnormal values, but fail to act because they are already busy dealing with another fault and may wish to finish dealing with one problem before tackling a new one. If the speech display placed an attentional demand on the operator, then dealing with the alarm display itself becomes a problem. This would result not only in a time delay pertaining to the length of the speech message, but also an additional time delay resulting from processing and handling the message.

5. Conclusions

It is suggested that speech alone as a medium for alarm displays cannot be recommended for tasks where; there is a memory component, there is likely to be some delay before the fault is attended to, there is likely to be more than one alarm presented at a time, and the operator is required to assimilate information from a variety of sources using spatial reference. If speech is to be incorporated into the alarm system for 'process control' tasks, it is recommended that it be paired with other media such as a scrolling text display. However, speech-based alarms might be appropriate for tasks where: an immediate response is required, the 'operator' is away from the interface, the situation is typically one-alarm to one-event, and fault management is serial in nature. Further investigation is needed before this latter proposal can be firmly recommended.

Acknowledgements

The authors are grateful to Sally Taylor-Adams who collected the data for this study and Ray Taylor who developed the simulated process plant. We would also like to

thank the anonymous referees for helping to make our presentation of this work clearer.

References

- BABER, C. (1991) *Speech Technology in Control Room Systems: A Human factors perspective* (Ellis Horwood: Chichester)
- BABER, C., STAMMERS, R. B. and TAYLOR, R. B. (1991) An experimental assessment of ASR in high cognitive workload situations in control room systems, in Y. Quéinnec and F. Daniellou (eds), *Ergonomics for Everyone: Proceedings of the 11th Congress of the International Ergonomics Association, Paris 1991*, (Taylor & Francis: London), 833-835.
- BABER, C., STANTON, N. A. and STOCKELY, A. (1992) Can speech be used for alarm displays in 'process control' type tasks? *Behaviour and Information Technology*, **11**, 216-226.
- BABER, C., USHER, D. M., STAMMERS, R. B. and TAYLOR, R. G. (1992) Feedback requirements for automatic speech recognition in the process control room, *International Journal of Man-Machine Studies*, **17**, 703-719.
- COWLEY, C.K., MILES, D. AND JONES, D.M. (1990) The incorporation of synthetic speech into the human computer interface, in E.J. Lovesey (ed), *Contemporary Ergonomics 1990*, (Taylor & Francis: London), 438-444
- PAYNE, D.G., PETERS, L.J., BIRKMIRE, D.P., BONTO, M.A., ANASTASI, J.S. AND WENGER, M.J. (1994) Effects of speech intelligibility level of concurrent visual task performance, *Human Factors* **36**, 441-475
- ITO, N., INOME, S. ONKURA, M. and MASADA, W. (1989) The effect of voice messages on the interactive computer system, in F. Klix, N. A. Streitz, Y. Waern and H. Wankle (eds) *Macinter II*, (North-Holland: Amsterdam), 245-252.
- MORAY, N. and ROTENBERG, I. (1989) Fault management in process control: eye movements and action, *Ergonomics*, **32**, 1319-1342.
- ROBINSON, C. R. and EBERTS, R. E. (1987) Comparison of speech and pictorial displays in a cockpit environment, *Human Factors*, **29**, 31-44.
- SALAMÉ, P. and BADDELEY, A. D. (1982) Disruption of short-term memory by unattended speech: implications for the structure of working memory, *Journal of Verbal Learning and Verbal Behaviour*, **21**, 150-164.
- SELCON, S. J., TAYLOR, R. M. and MCKENNA, F. P. (1995) Integrating multiple information sources: using redundancy in the design of warnings, *Ergonomics*, **38**, 2362-2370.

- STANTON, N. A. (1991) Alarm information in fault diagnosis, IEE Digest number 1991/156, (IEE: London), 9/1-9/5.
- STANTON, N. A. (1994 a) Key topics in alarm design, in N. A. Stanton (ed) *Human Factors in Alarm Design*, (Taylor & Francis: London), 221-231.
- STANTON, N. A. (1994 b) Alarm initiated activities, in N. A. Stanton (ed) *Human Factors in Alarm Design*, (Taylor & Francis: London), 93-117.
- STANTON, N. A. and BABER, C. (1995) Alarm initiated activities: an analysis of alarm handling by operators using text-based alarm systems in supervisory control systems, *Ergonomics*, **38**, 2414-1431.
- STANTON, N. A., BOOTH, R. T. and STAMMERS, R. B. (1992) Alarms in human supervisory control: a human factors perspective. *International Journal of Computer Integrated Manufacturing* **5**, 81-93.
- SMITH, A. (1989) A review of the effects of noise on human performance, *Scandinavian Journal of Psychology*, **30**, 185-206
- USHER, D. M. (1994) The alarm matrix, in N. A. Stanton (ed) *Human Factors in Alarm Design*, (Taylor & Francis: London), 137-145
- WICKENS, C. D. (1992) *Engineering Psychology and Human Performance*, (Harper Collins: New York).

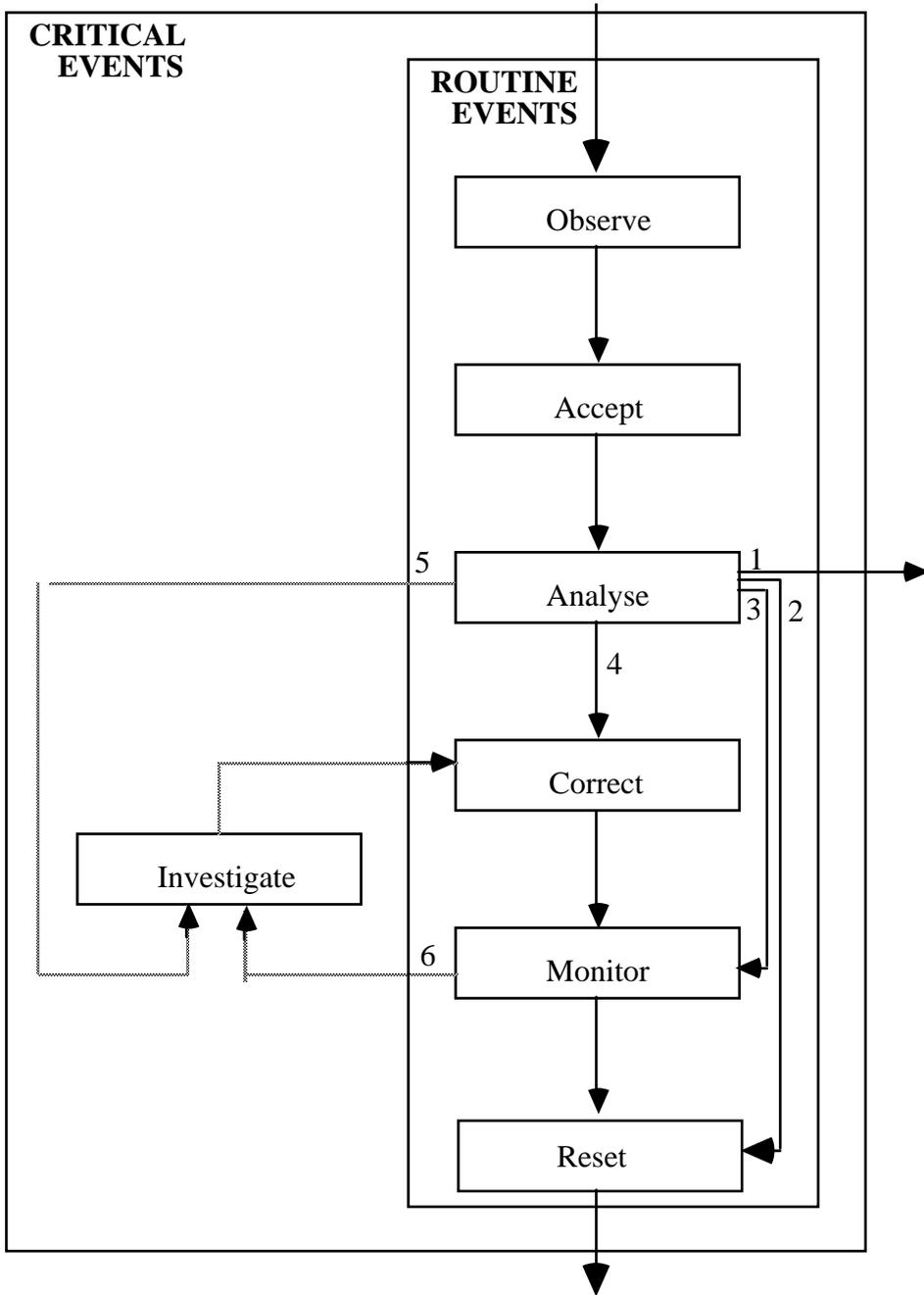


FIGURE ONE

FIGURE TWO

Source	Sum squares	d.f.	Variance	F	p
between groups	6842.85	2	3421.43	3.744	0.0367
within groups	24676.03	27	913.93		
total	3518.88	29			

Table 1. Analysis of variance table for output performance

Source	Sum squares	d.f.	Variance	F	p
between groups	659859779.57	2	329929889.78	7.418	0.003
within groups	1.112E9	25	44480000		
total	1.772E9	27			

Table 2. Analysis of variance table for alarm accept times

Source	Sum squares	d.f.	Variance	F	p
between groups	309917470.971	2	154958735.485	21.986	0.0001
within groups	169149704.214	24	7047904.342		
total	479067175.185	26			

Table 3. Analysis of variance table for fault accept times

Dependent variable	Condition		
	text vs. combined	speech vs. text	speech vs.combined
<i>Overall</i>			
Ouput level	ns	*	ns
<i>Alarm Handling</i>			
observe time	ns	*	*
investigate time	ns	ns	ns
correct time	ns	ns	ns
<i>Fault handling</i>			
observe time	ns	*	*
investigate time	ns	ns	ns
correct time	ns	ns	*
<i>Additional measures</i>			
inappropriate actions	ns	**	*
secondary task	ns	ns	ns
recall test	ns	***	**

Table 4. Summary of statistical comparisons of data