An Evaluation of a "Time Tunnel" Display Format for the Presentation of Temporal Information

Kevin B. Bennett, Michael Payne, and **Brett Walters,** Wright State University, Dayton, Ohio

The time tunnel display design technique combines the benefits of configural displays (salient visual properties corresponding to critical domain semantics) with the benefits of temporal information (i.e., the value of variables and properties over time). In Experiment 1 a baseline configural display and a time tunnel display were evaluated using real-time measures of system control, fault detection, and state estimation in a simulated process control task. The results provided little evidence in support of the time tunnel format. In Experiment 2 access to the temporal context was limited: Participants performed the detection and estimation tasks with static "snapshots" of system states that had been generated in Experiment 1. The overall pattern of results indicates that the time tunnel display was more effective for state estimation tasks than was the baseline configural display and or a trend display. Issues in the design of temporal displays are discussed, including representational formats and the choice of temporal time frames. Issues in the evaluation of temporal displays are also discussed, including the role of temporal information and the critical nature of participants' access to this information. Actual or potential applications of this research include design techniques for improving graphical displays and methodological insights to guide future evaluations.

INTRODUCTION

Practitioners in virtually all domains of application need to be concerned with changes in resources, properties, and variables over time. The term *temporal information* will be used to refer to the historical trace of these changes. The primary benefit of temporal information is that "trends" are revealed (i.e., the fact that the value of variables have changed in particular ways, as opposed to others). It is difficult to imagine a domain in which temporal information is not at least potentially useful. For example, it is potentially useful when the behavior of the system is driven primarily by the goals and intentions of its users (e.g., to illustrate trends in the stock market). However, temporal information should be especially useful for domains in which the behavior of the system is driven by the laws of nature (e.g., process control). In these domains past system states will determine current and future system states, at least under normal circumstances. Therefore temporal information should be very useful in understanding current system states, predicting future system states, choosing appropriate control inputs, and detecting the presence of system faults.

Temporal information has traditionally been presented in trend or strip chart displays, which plot changes in the value of a variable or a resource as a function of time. These formats have been demonstrated to facilitate the detection of trends with the static presentation of data (e.g., Schutz, 1961). Despite their intuitive appeal and widespread use, there has been surprisingly little research conducted on trend displays in dynamic settings. At least one study has produced negative findings (Spenkelink, 1990). From a theoretical perspective, one potential drawback to trend displays is that they are essentially "separable" in nature: Each variable has its own unique representation. As a result the

Address correspondence to Kevin B. Bennett, Department of Psychology, 335 Fawcett Hall, Wright State University, Dayton, OH 45435; kevin.bennett@wright.edu. **HUMAN FACTORS,** Vol. 47, No. 2, Summer 2005, pp. 342–359.

relationships between variables are not emphasized or, at least, are not emphasized to the extent that is possible with alternative display formats. Because these relationships are critical in complex, dynamic domains this limitation is a potentially serious one.

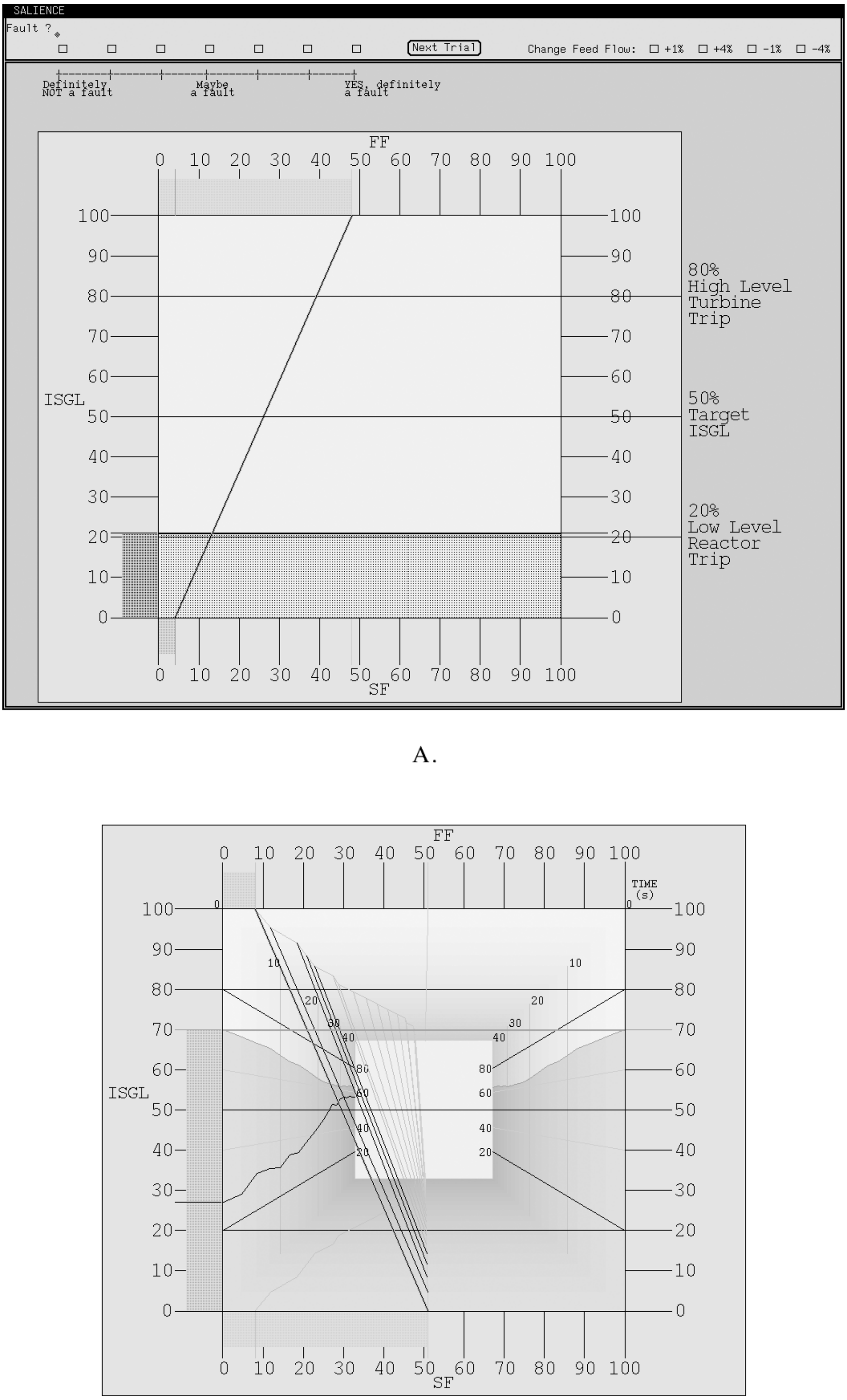
Configural displays explicitly emphasize the relationships between variables. Individual variables are arranged in spatial patterns (often connected with contour lines) to produce geometrical forms that change shape as a function of changes in the value of these variables. The salient, high-level visual properties that are produced (e.g., symmetry) are usually referred to as "emergent features" (Pomerantz, 1986). A substantial body of laboratory research indicates that configural displays can be effective when the consideration of relationships between variables is essential to the completion of domain tasks (Bennett & Flach, 1992). The degree of success is determined by the quality of the mapping between the visual properties of the display and the physical, functional, and goal-related properties of the domain (Bennett, Nagy, & Flach, 1997).

An example of a configural display is presented in Figure 1a (based on the "funnel" metaphor of Vicente, 1991). This display presents three variables that are critical to a process control task, the manual control of feedwater. The level of coolant in a steam generator (indicated level) is plotted as a vertical bar graph on the left side of the display. The flow rate of mass entering the steam generator (feed flow) is plotted as a horizontal bar graph at the top of the display. The flow rate of mass leaving the steam generator (steam flow) is plotted as a horizontal bar graph at the bottom of the display. The domain property of mass balance (the relationship between mass in and mass out) is represented directly by the line connecting the steam flow and the feed flow bar graphs (the mass balance indicator). The orientation of this line is an emergent feature that specifies mass balance.

From a theoretical perspective configural displays have a potential drawback, even when they are designed properly and are otherwise effective: There is no explicit representation of temporal information. As our previous analysis suggests, this could be a serious drawback, particularly for systems driven by the laws of nature. Pawlak and Vicente (1996) obtained some empirical evidence that supports these concerns in a process control setting. They evaluated an interface that contained configural displays (the P + F interface). Although this interface was generally effective, Pawlak and Vicente (1996) noted, "The masking problem experienced by P + F subjects during the second fault trial was probably due, in part, to the fact that there was no history [temporal] information available for reservoir volume" (p. 683).

In summary, configural displays and trend displays possess complementary strengths and weaknesses for data presentation requirements in complex sociotechnical systems. The next logical step is to consider if, and how, these two general formats might be combined. As Pawlak and Vicente (1996, p. 683) have observed, "the challenge lies in integrating historical [temporal] information with the existing emergent features of the display" (p. 683). Hansen (1995) has devised a technique that has the potential to meet this design challenge. The *time tunnels* display design technique allows variations in configural displays (i.e., geometric forms) to be seen over time. It accomplishes this goal by scaling multiple versions of a geometric form according to the laws of perspective geometry and presenting them in the depth plane of the display. This representational format combines the positive aspects of configural displays (direct representation of high-level domain properties) with the positive aspects of temporal displays (a trace of these properties over time).

Figure 1b illustrates a variation of Hansen's (1995) time tunnel technique applied to the funnel display of Figure 1a. A static framework, or *perspective grid,* is plotted in the depth plane. The outermost rectangle represents display axes that correspond to the current time frame. Each successive rectangle is scaled and plotted deeper in the depth plane to represent the display axes at a point more distant in time. Temporal information (individual variables, relationships, and goals over time) is presented within this perspective grid. *Perspective trends* are formed by connecting the values of individual variables in contiguous time frames. Similarly, mass balance relationships over time are represented by a series of mass balance indicator lines that are



A.

B.

Figure 1. The displays investigated in Experiment 1: (a) the baseline configural display and (b) the time tunnel display. FF = feed flow, SF = steam flow, ISGL = indicated (steam generator) level.

formed by connecting the values of steam and feed flow within a time frame.

The present study continues a line of research investigating time tunnel displays. Bennett and Zimmerman (2001) evaluated a previous implementation of the time tunnel display using the manual control of feedwater simulation. They found very little evidence to suggest that the time tunnel display improved performance at system control, fault estimation, or state estimation tasks. One potential explanation is related to issues in representation. The specific representational conventions used in their version of a time tunnel display (i.e., geometrical planes, gray scale shading, occlusion) produced a display that was highly complex. One goal of the present experiment was to determine if these representational issues were responsible for the lack of performance benefits. To examine this possibility, we developed an alternative version of the time tunnel display that drastically reduced the visual complexity (Figure 1b).

A second potential explanation involves the nature of the information that was present in the time tunnel display. Previous analyses of the manual control of feedwater task have indicated that the counterintuitive energy effects and time lags that are associated with indicated level (the primary variable to be controlled) are factors that contribute to its difficulty (Roth & Woods, 1988). These insights led to the development of a form of decision support referred to as *compensated level,* a calculated variable that provides an estimate of the value that indicated level will assume after these transitory effects have dissipated. Bennett and Zimmerman (2001) evaluated both compensated level and the time tunnel technique. In contrast to the results obtained for a time tunnel display, compensated level was extremely effective in improving performance at a variety of tasks. It is possible that the presence of this powerful, predictor-like variable may have deterred participants from becoming attuned to the temporal information that was present in the display. The present study explores this possibility by removing compensated level from the displays.

The baseline display (Figure 1a) and the redesigned time tunnel display (Figure 1b) were evaluated using several dependent measures. Performance at the real-time system control and fault detection tasks was measured using the same methodological procedures of the previous evaluation. Participants also performed two types of information estimation tasks. One task required participants to provide an estimate of compensated level (to allow us to assess their understanding of current system state). A second task required participants to estimate indicated level (the primary variable to be controlled) at three points in time (current, –20 s, or –40 s).

EXPERIMENT 1

Method

Participants. Five men and 3 women (20–32 years of age) were paid $5.00/hr for their participation. All participants had normal or normal-corrected vision with no color blindness deficiencies. All participants had completed at least three similar experiments and had a minimum of 30 hr of experience.

Apparatus. All experimental events were controlled by a general purpose laboratory computer (Sun Microsystem Workstation, Model 4-110) located in an enclosed experimental room. A color video monitor (Sony Trinitron, model GDM1604-15, 40.64 cm, 1152 × 900 resolution) and a standard keyboard were used.

Simulation model. For a more detailed description of the simulation model, see Bennett, Toms, and Woods (1993).

Stimuli. The primary axes of the baseline configural display (Figure 1a) formed a 12.70-cm square. These axes contained scale markers and labels (0%–100%, black). The bar graphs for three variables (steam flow, feed flow, indicated level) were 1.18 cm wide; had a maximum extension of 12.70 cm; were green, blue, or mustard in color; and were "stenciled" (every other pixel assuming the background color, medium gray). A fourth variable ("adjusted" indicated level = indicated level + [feed flow – steam flow]) appeared on the same axis as indicated level. The current mass balance indicator consisted of three lines (a black line bracketed by two lines the color of the associated bar graphs). Two trip set points (20% and 80%, red) and a target indicated level (50%, white) were present.

The time tunnel display (Figure 1b) retained these physical characteristics and presented temporal information over the last 40 s at 2-s intervals. It was constructed in the following manner. A monocular observer views a picture plane (100 cm square) that is located 500 cm away (centered in the observer's field of view). This picture plane corresponds to the current time slice of data (i.e., the outermost rectangle corresponding to the primary axes of the display). Nineteen additional picture planes were created to form the tunnel. The size of each successive picture plane was calculated by assuming a displacement of 100 cm away from the observer. These picture planes were centered in the display, and the area between them was filled with gray scale shading ranging from light gray (most recent) to dark gray (most distant). Redundant scales appeared on the inside of the tunnel (20%–80%, black). Perspective trends (inside the tunnel) traced the value of these variables over a 40-s time span. Twenty mass balance indicators highlighted steam and feed flow relationships (one for each time frame); historical mass balance indicators were represented by single lines ranging from dark gray (most recent) to light gray (most distant).

Procedure. Each participant completed one practice session and four experimental sessions lasting approximately 1 hr each. In the first practice session the participants were given descriptions of the displays, the tasks, and experimental instructions. No discussion of specific control strategies was provided. The participants were tested individually in an enclosed room. There were eight nonfault trials (two displays with four repetitions) and two fault trials (one leak, one stuck valve) per experimental session. The presentation order was random. Fault trials occurred within the first eight trials in a session; additional, nonfault trials for the corresponding displays were administered at the end of the session. Each trial lasted for 5 min. Steam flow and feed flow were 0% initially; indicated level and compensated level were 35% initially. Every 2 s a 1% increase could occur in steam flow (25% probability) as long as steam flow was less than 80%. The displays were also updated every 2 s. Participants changed feed flow by pointing and clicking on one of four boxes (increasing or decreasing feed flow by 1% or 4%; see Figure 1a). They were instructed to provide control inputs that moved indicated level to a target level (50%) quickly, to maintain indicated level close to this target level, and to avoid crossing set point boundaries. Auditory feedback (four tones) occurred when a boundary was crossed. Continuously updated root mean square error scores were provided.

Two faults could occur. One fault simulated a steam generator leak by decreasing the value of indicated level by 0.25% at 2-s intervals. The second fault simulated a stuck valve: Control input to feed flow caused the displayed value to change (commanded value) but did not change the simulation (actual value). When a fault was present, it began at a random starting point ranging from 30 to 90 s into a trial. Participants provided a confidence rating for the presence or absence of a fault at seven points in an experimental session (40, 80, 120, 160, 200, 240, and 280 s). Participants pointed and clicked at a 7-point scale (see Figure 1a). Feedback on the presence or absence of a fault was provided at the end of each trial. Each participant completed two fault trials (one leak, one stuck valve) in each of four experimental sessions for a total of eight fault trials. Each combination of fault type (2) and display type (2) occurred twice for each participant. The experiment-wide presentation order was counterbalanced across participants and days so that each combination of fault type and display type occurred exactly four times on each day.

Participants completed six information probes during each experimental trial. During a probe the simulation was paused, an auditory tone sounded, a textual description of the probe was presented, and participants entered a numeric value via the keyboard. The display remained visible at all times. The participants were instructed to respond to probes as accurately and quickly as possible. Feedback on both accuracy and latency was provided. A probe was readministered in the final 25 s of a trial if the participant entered an unacceptable value or changed his or her estimate.

Two categories of information probes were completed. The compensated level estimation task required participants to estimate the value of compensated level (produced by the simulation model but not presented on the screen). An indicated level estimation task required participants to estimate (a) the current value of indicated level, (b) its value 20 s in the past, or

(c) its value 40 s in the past. Six probes (three compensated level, each of the three indicated level) were completed in each trial. These probes occurred during six time windows (45–75, 85–115, 125–155, 165–195, 205–235, and 245–275 s) and were administered when the next screen update was scheduled to occur. The presentation order was determined by (a) randomizing the three indicated level probes, (b) pairing each of these probes with a compensated level probe, and (c) randomizing the order of the two probes within a pair. The color coding for the four variables was counterbalanced across participants.

Results

A similar procedure was followed for the majority of analyses. Outliers were identified using the test described in Lovie (1986, pp. 55–56), $T_1 = (x_{(n)}\ \bar{x})/s$, in which $x_{(n)}$ is a particular observation (one of n observations), $\bar{x}$ is the mean of those observations, and s is the standard deviation of those observations. The individual observations identified as outliers were not considered in subsequent analyses; all other observations from individual participants were retained. Nonparametric tests (Friedman analysis of variance [ANOVA]) were conducted to determine if the outlier distribution was random (none was significant). Only significant effects involving the display manipulation are reported.

Compensated level estimates. Accuracy (error magnitude) was measured by computing the absolute value of the difference between the participant's estimate of a variable and the actual value. Latency was measured from the appearance of the prompt until the first digit of the participant's response (1/100-s accuracy). Data were averaged across the 12 repetitions occurring in an experimental session (three probes in four trials) and a 2 (display) × 4 (session) repeated measures ANOVA was conducted for both accuracy and latency. No effects were significant.

Indicated level estimates. Data were averaged across the four repetitions occurring in an experimental session (four trials), and a 2 (display) × 4 (session) × 3 (time: current, –20 s, –40 s) repeated measures ANOVA was conducted for both accuracy and latency. No effects involving the display manipulation were significant for accuracy. For latency the main effect of display, $F(1, 7) = 16.10$, $p < .006$, and the Time × Display interaction, $F(2, 14) = 7.69$, $p < .006$, were significant. Fisher's least significant difference revealed that there were no significant differences in performance between displays for current estimates of indicated level but that the time tunnel display took significantly longer than the baseline display for estimates of indicated level in the past (both –20 and –40 s).

Control performance. Six measures of control performance were considered: acquisition time, settling time, root mean square error, constant position error, modulus mean error, and standard deviation of the error (see Bennett & Zimmerman, 2001, for formulas). Analyses of control performance for nonfault and reservoir leak fault trials were conducted (the stuck valve fault trials were not analyzed because control input had no effect). For nonfault trials, data were averaged across the four repetitions occurring in an experimental session, and a 2 (display) × 4 (session) repeated measures ANOVA was conducted for each of the six control measures. No significant effects involving the display manipulation were obtained. For the reservoir leak fault trials, data were averaged across the repetition factor and a one-way (display) repeated measures ANOVA was conducted for each control measure. The standard deviation of the error analysis revealed that the time tunnel display produced significantly less variable control performance than did the baseline display, $F(1, 7) = 5.94$, $p < .05$.

Fault estimates. The estimates of fault certainty and the corresponding latencies were averaged across repetition and session. Preplanned contrasts for both linear and quadratic trends were conducted to compare performance between displays. No significant effects involving the display manipulation were obtained.

Discussion

Experiment 1 reinvestigated the time tunnel display technique with several modifications (i.e., alternative display design, additional performance measures, removal of compensated level). The results of this follow-on study did not differ substantially from the initial evaluation (Bennett & Zimmerman, 2001): There was very little evidence that the time tunnel display was more effective than the baseline display. The standard deviation of error during reservoir leak trials did indicate that control was significantly less variable with the time tunnel display than with the baseline display. However, the response latencies obtained with the baseline display were significantly lower for estimates of indicated level in the past (–20 and –40 s). There were no significant differences between displays for fault detection confidence ratings or for estimates of compensated level.

These results suggest that the two potential explanations outlined in the Introduction (i.e., issues in representation, the presence of compensated level) were probably not responsible for the lack of performance benefits for the time tunnel display. After considerable deliberation, we considered an alternative hypothesis. In simple terms, it is possible that the time tunnel displays were not successful simply because they provided no information of which the participants were not already aware. Several factors may have contributed to this outcome. The participants were very familiar with both general system dynamics and specific behaviors associated with the start-up sequence as a result of their fairly extensive experience. In addition, no divided attention or time-sharing demands were imposed, given that only the one primary task was performed. Under these circumstances it is possible that the participants themselves had internally integrated the temporal information that was required for successful performance. Stated alternatively, the time tunnel displays may not have had an impact on performance because they provided an external representation of temporal information that participants had already internalized.

Experiment 2 investigated this hypothesis through substantial changes in methodology. These methodological changes were designed to eliminate (or at least drastically reduce) the opportunity for participants to internally integrate temporal information by cutting off their access to the unfolding temporal context. The participants did not control and monitor the part-task simulation in real time. Instead, they were presented with "snapshots" of actual system states that had been generated in Experiment 1; they were required to perform the same fault detection and state estimation tasks on these snapshots that they had completed in Experiment 1. The two displays evaluated in Experiment 1 (time tunnel and baseline) were also evaluated in Experiment 2 with no changes. A third display represented temporal information in a more traditional format: a strip chart containing the value of individual variables over time (Figure 2). This display will be referred to as the *trend display.*

EXPERIMENT 2

Method

Participants. Of the 8 participants in Experiment 1, 4 (all men) also participated in Experiment 2. Three of the remaining participants had moved from the area (this study was conducted 1 year and 9 months after the first), and the fourth was unable to participate because of medical problems. Participants were paid $5/hr.

Apparatus. This was the same as in Experiment 1.

Simulation model. The same simulation model was used in the training sessions.

Stimuli. The implementation of the baseline and time tunnel displays remained the same as in Experiment 1. The trend display (Figure 2) consisted of the baseline display plus a strip chart display, which measured 6.60 × 12.70 cm. The values of variables were plotted as a function of time (current time was located on the right; 40 s in the past was located on the left). Trends for individual variables were formed by connecting values occurring at contiguous 2-s intervals.

Procedure. Participants were provided with written and oral instructions and then completed eight sessions (approximately 1 hr). In the first two sessions (training/refresher sessions) participants controlled the simulation. During each training session participants completed four trials (two nonfault, one reservoir leak, and one stuck valve) with each of the three displays (12 total trials). Feedback regarding the presence and nature of faults were provided.

During the six experimental sessions participants initiated a trial by pointing and clicking on a button. A display appeared depicting a system state, and participants completed three experimental tasks in the following order. Participants indicated their certainty regarding the presence of a fault by pointing and clicking on a 6-point

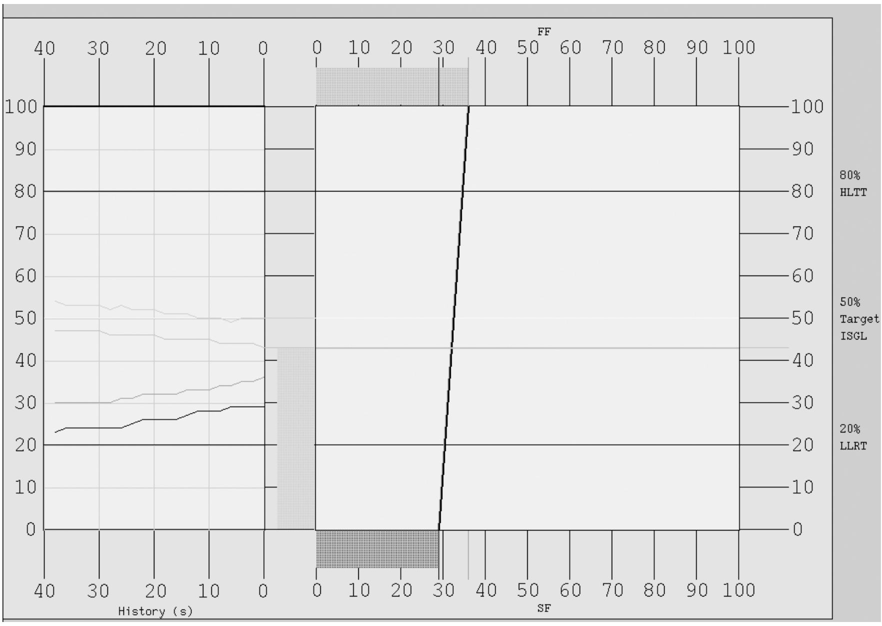


Figure 2. The trend display investigated in Experiment 2. FF = feed flow, SF = steam flow, HLTT = high level turbine trip, ISGL = indicated (steam generator) level, LLRT = low level reactor trip.

Likert scale (endpoints labeled *definitely no fault* and *definitely a fault*). Participants then completed a two-alternative, forced choice procedure that required them to indicate which of the two faults (reservoir leak or stuck valve) was more likely to have generated the configuration of system variables that was depicted. Finally, participants estimated the current value of compensated level. The participants were instructed to respond as accurately and as quickly as possible for all three tasks. Feedback for latency (fault certainty, compensated level) and accuracy (fault presence, fault type, and compensated level) was provided.

System states were chosen from simulation data generated during the last three sessions of Experiment 1. A total of 48 trials were used: 12 trials for each of the two fault types (randomly selected) and the 24 corresponding nonfault trials. Data were taken from 5 "onset" times (40, 80, 120, 160, and 200 s) measured relative to the beginning of a fault (for nonfault trials the onset time corresponded to the beginning of a fault in the matched fault trial). The resulting 240 system states (48 trials × 5 onsets) were randomly divided into six data subsets with the constraint that the 40 trials in a data subset contained a factorial combination of the four trial types (two faults, two matched nonfaults), the five onset times, and two repetitions. Participants completed three blocks (one for each display, random order) of 40 trials (random order) in each experimental session. Each of these blocks corresponded to one of the six data subsets. Each combination of data subset and display occurred exactly once for each participant during the course of the experiment.

Results

A similar procedure was followed for the majority of analyses. Outliers were identified using the same procedures outlined in Experiment 1. The compensated level estimate (absolute error, signed error), the compensated level latency, and the fault detection latency measures produced 64 (2.22%), 62 (2.15%), 58 (2.01%), and 60 (2.08%) outliers, respectively. Nonparametric tests were conducted to determine if the outlier distributions were random (none was significant). Scores were averaged across repetitions, sessions, and repetitions of nonfault trials (nonfault trials were not averaged for the signal detection and the fault identification analyses). Analytical comparisons were conducted: Four preplanned orthogonal contrasts were used to test display by onset time interactions. These contrasts were aimed at discovering whether or not temporal information (i.e., baseline vs. trend and tunnel displays) and the representation of temporal information (i.e., trend vs. tunnel display) had an impact on performance. Trend analysis (linear and quadratic components) was used to determine if the patterns of performance associated with the various displays were different across onset times. The four orthogonal contrasts resulted from the crossing of display (i.e., presence of temporal information, representation of temporal information) with onset time (linear, quadratic) components. These orthogonal contrasts were applied separately to the three trial types (normal, reservoir leak, stuck valve) for a total of 12 contrasts for each dependent variable. Additional analyses were performed only when a preplanned contrast was significant.

Table 1 lists significant preplanned orthogonal contrasts testing performance between the trend and tunnel displays. The supplemental analyses for significant contrasts are straightforward: A comparison to test the simple main effect of display was conducted for each of the five onset times. Table 2 lists the significant preplanned contrasts comparing performance between the baseline display and the two displays with temporal information (Contrasts 7–12). If a contrast was significant, then a separate interaction comparison was conducted for each temporal display relative to the baseline. These supplemental interaction comparisons are labeled in Table 2 with a number and a letter (e.g., 7a). Only when a supplemental interaction comparison was significant were tests for the simple main effects of display at onset levels conducted. A graphical summary of the significant simple main effects of display at onset time is presented in Figure 3. The display by onset means (along with linear best fits) for the signed error of compensated level estimates are illustrated in Figure 4.

Discussion

In contrast to the results of Experiment 1, those obtained in Experiment 2 revealed a number of performance differences that were related to both the presence of temporal information and the representational format that was used to present it (Tables 1 and 2; Figure 3). The majority of these effects involve the estimation of compensated level. The two displays containing temporal information (i.e., the tunnel and trend displays) produced significantly better estimates of compensated level than did the display without temporal information (i.e., the baseline display) during later onset times. As illustrated in Figure 3b, a total of 6 comparisons indicated that the trend display produced significantly better estimates than did the baseline display during the final two onset times (those contrasts outlined with thin borders). A total of 11 comparisons indicated that the time tunnel display produced significantly better estimates of compensated level than did the baseline display during the final three onset times (Figure 3c, those contrasts filled with gray).

In contrast, the pattern of performance was reversed during the initial onset times: The baseline display produced better estimates than did the two displays that contained temporal information. The significant performance decrements for the trend display (14 significant comparisons, Figure 3b, thick borders) were far more numerous than those for the tunnel display (6 significant comparisons, Figure 3c, thick borders). Finally, the direct comparisons between the two displays with temporal information revealed that the tunnel display produced substantially better estimates: 14 significant comparisons favored the tunnel display (Figure 3a, gray fill), whereas only 1 comparison favored the trend display (Figure 3a, thin outline).

Temporal information and onset time. The results indicate very clearly that the presence of temporal information (in the tunnel display and, to a lesser extent, in the trend display) improved the ability to estimate compensated level during later onset times but not during initial onset times. The interpretation of these findings requires an explicit consideration of the nature and the utility of the temporal information that was present during early and late onset times. Therefore, Figure 5 provides the average values of the system variables for normal and fault trials in Experiment 1. The average onset times are labeled on the horizontal axis; the corresponding temporal information for a particular onset time

TABLE 1: Preplanned Orthogonal Contrasts Comparing Performance for Displays With Alternative Representations of Temporal Information in Experiment 2

	Fault Detection: Estimates				Fault Detection: SDT *d′*		Estimates of Compensated Level					
	Fault Certainty		Latency				Absolute Error		Signed Error		Latency	
Contrast #, Verbal Description	$F(1, 3) =$	$p <$	$F(1, 3) =$	$p <$	$F(1, 3) =$	$p <$	$F(1, 3) =$	$p <$	$F(1 ,3) =$	$p <$	$F(1, 3) =$	$p <$
Trend vs. tunnel: nonfault trials												
1. Linear	93.65	.003							108.14	.002		
2. Quadratic							87.58	.003				
Reservoir leak trials												
3. Linear			11.69	.05	38.72	.009			69.44	.004		
4. Quadratic							64.34	.005	75.97	.004		
Stuck valve trials												
5. Linear	10.34	.05					186.31	.0009	70.12	.004		
6. Quadratic							236.98	.0006				

SDT = signal detection theory

TABLE 2: Preplanned Orthogonal Contrasts and Supplemental Interaction Contrasts for Experiment 2

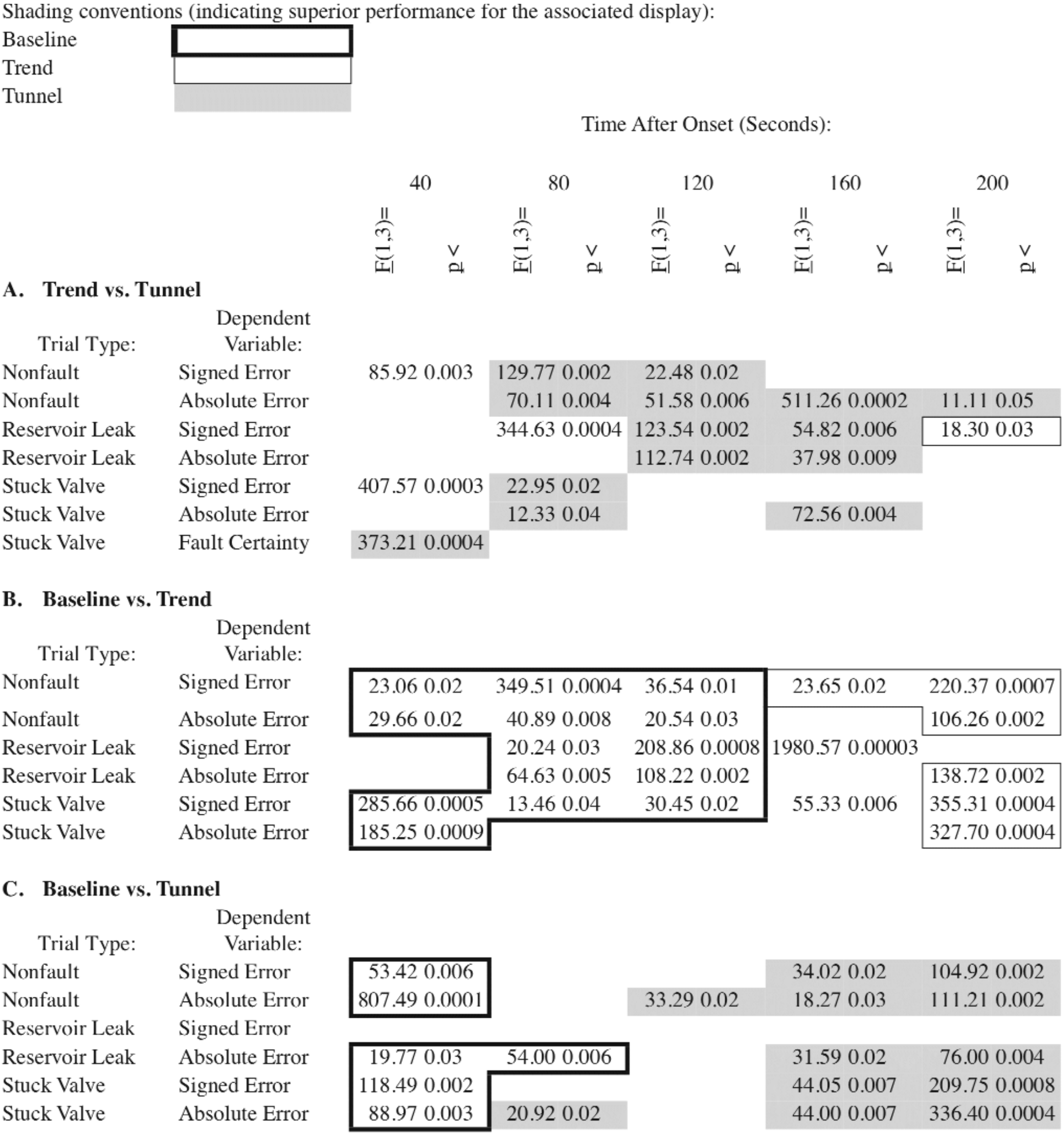
	Estimation of Compensated Level					
	Absolute Error		Signed Error		Latency	
Contrast #, Verbal Description	*F*(1, 3) =	*p* <	*F*(1, 3) =	*p* <	*F*(1, 3) =	*p* <
Baseline vs. Trend and Tunnel: Nonfault trials						
7. Linear	192.08	.0009	88.95	.003	11.82	.05
7a. Interaction comparison: baseline vs. trend	213.64	.0007	18.26	.03		
7b. Interaction comparison: baseline vs. tunnel	154.56	.002	205.34	.0008		
8. Quadratic	1170.79	.00006				
8a. Interaction comparison: baseline vs. trend	251.99	.0006				
8b. Interaction comparison: baseline vs. tunnel						
Reservoir Leak Trials						
9. Linear	1234.59	.00006				
9a. Interaction comparison: baseline vs. trend	112.30	.002				
9b. Interaction comparison: baseline vs. tunnel	155.82	.002				
10. Quadratic	356.23	.0004	27.21	.02		
10a. Interaction comparison: baseline vs. trend	561.39	.0002	445.79	.0003		
10b. Interaction comparison: baseline vs. tunnel	13.24	.04				
Stuck Valve Trials						
11. Linear	226.39	.0007	295.22	.0005	23.32	.02
11a. Interaction comparison: baseline vs. trend	210.63	.0008	67.38	.004		
11b. Interaction comparison: baseline vs. tunnel	235.28	.0007	459.57	.0003	12.80	.04
12. Quadratic	14.44	.04				
12a. Interaction comparison: baseline vs. trend	71.48	.004				
12b. Interaction comparison: baseline vs. tunnel						

are those data bracketed by the pair of vertical dotted lines above and to the left of an onset label. Please note that the values of compensated level are provided for illustrative purposes only and were never actually presented to participants.

The temporal information during the initial two onset times (40 and 80 s) did not vary substantially between normal and fault trials (see Figure 5). The similarities occurred because the effects of the two faults were being masked by the normal, but substantial, thermodynamic effects that resulted from the control inputs that were executed to move indicated level to the goal value. As a result, all three types of trials were characterized by large imbalances between mass input (feed flow) and output (steam flow) and/or substantial rates of changes for variables.

In contrast to the initial onset times, in the final two onset times (160 and 200 s) the temporal information present provided clear evidence that differentiated between normal and fault trials. The pattern of variables that appears during nonfault trials (Figure 5a) at later onset times specifies normal system dynamics and the absence of any transitory thermodynamic effects: mass input/output was balanced and the value of indicated level was constant over time. In contrast, the temporal information that is present during the reservoir leak trials (Figure 5b) at later onset times clearly specifies a break in system constraints. The value of indicated level was reasonably stable over time. However, this was attributable only to the fact that participants were successfully compensating for the reservoir leak by maintaining mass input (feed flow) at higher levels than mass output (steam flow).

The temporal information during stuck valve trials did indicate the presence of a fault, but the evidence was less clear than that for the reservoir leak. Participants were attempting to compensate for the decreases in indicated level by



Shading conventions (indicating superior performance for the associated display):
Baseline ▣
Trend □
Tunnel ▨

Time After Onset (Seconds):

A. Trend vs. Tunnel

Trial Type:	Dependent Variable:	40 F(1,3)=	40 p<	80 F(1,3)=	80 p<	120 F(1,3)=	120 p<	160 F(1,3)=	160 p<	200 F(1,3)=	200 p<
Nonfault	Signed Error	85.92	0.003	129.77	0.002	22.48	0.02				
Nonfault	Absolute Error			70.11	0.004	51.58	0.006	511.26	0.0002	11.11	0.05
Reservoir Leak	Signed Error			344.63	0.0004	123.54	0.002	54.82	0.006	18.30	0.03
Reservoir Leak	Absolute Error					112.74	0.002	37.98	0.009		
Stuck Valve	Signed Error	407.57	0.0003	22.95	0.02						
Stuck Valve	Absolute Error			12.33	0.04			72.56	0.004		
Stuck Valve	Fault Certainty	373.21	0.0004								

B. Baseline vs. Trend

Trial Type:	Dependent Variable:	40 F(1,3)=	40 p<	80 F(1,3)=	80 p<	120 F(1,3)=	120 p<	160 F(1,3)=	160 p<	200 F(1,3)=	200 p<
Nonfault	Signed Error	23.06	0.02	349.51	0.0004	36.54	0.01	23.65	0.02	220.37	0.0007
Nonfault	Absolute Error	29.66	0.02	40.89	0.008	20.54	0.03			106.26	0.002
Reservoir Leak	Signed Error			20.24	0.03	208.86	0.0008	1980.57	0.00003		
Reservoir Leak	Absolute Error			64.63	0.005	108.22	0.002			138.72	0.002
Stuck Valve	Signed Error	285.66	0.0005	13.46	0.04	30.45	0.02	55.33	0.006	355.31	0.0004
Stuck Valve	Absolute Error	185.25	0.0009							327.70	0.0004

C. Baseline vs. Tunnel

Trial Type:	Dependent Variable:	40 F(1,3)=	40 p<	80 F(1,3)=	80 p<	120 F(1,3)=	120 p<	160 F(1,3)=	160 p<	200 F(1,3)=	200 p<
Nonfault	Signed Error	53.42	0.006					34.02	0.02	104.92	0.002
Nonfault	Absolute Error	807.49	0.0001			33.29	0.02	18.27	0.03	111.21	0.002
Reservoir Leak	Signed Error										
Reservoir Leak	Absolute Error	19.77	0.03	54.00	0.006			31.59	0.02	76.00	0.004
Stuck Valve	Signed Error	118.49	0.002					44.05	0.007	209.75	0.0008
Stuck Valve	Absolute Error	88.97	0.003	20.92	0.02			44.00	0.007	336.40	0.0004

Figure 3. Simple main effects of display at various onset times for preplanned contrasts or significant interaction comparisons that were significant in Experiment 2: (a) trend versus tunnel, (b) baseline versus trend, (c), baseline versus tunnel.

increasing feed flow relative to steam flow (Figure 5c). The intended effect of these control inputs was accurately portrayed as the "commanded" value of feed flow in the displays. However, this commanded value is actually quite misleading: The "actual" value of feed flow did not change after the onset of the stuck valve fault (it remained at approximately 15% after fault onset).

The problem that this poses for participants is that they were viewing a configuration of variables that could very easily be confused with normal shrink effects: Increases in feed flow relative to steam flow normally produce thermodynamic "shrink" effects, which cause the value of indicated level to decrease initially, as illustrated in the beginning of all trials in Figure 5. The evidence that specifies abnormal system

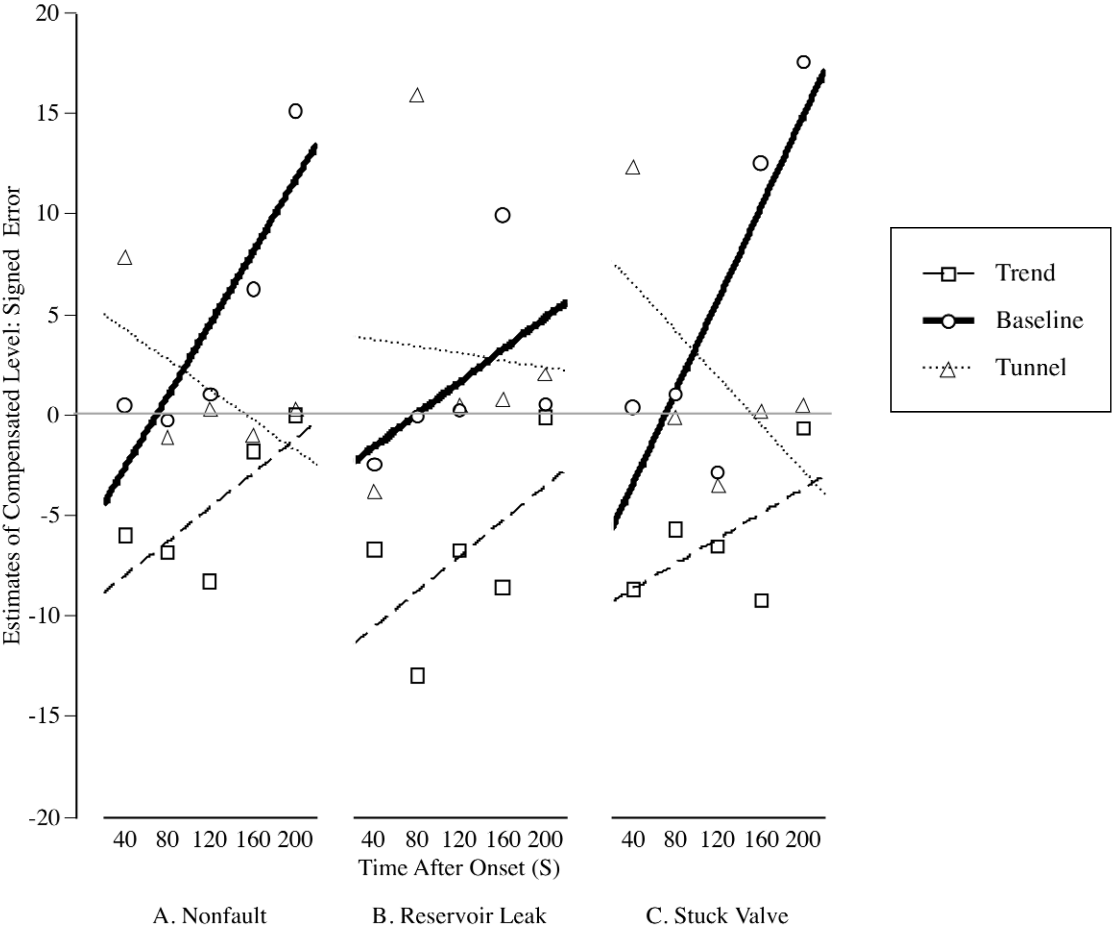


Figure 4. Average signed error of compensated level estimates for display and onset times during (a) nonfault, (b) reservoir leak, and (c) stuck valve trials.

dynamics is the rapid and sustained decreases in indicated level that occur over the course of the last 150 s: This behavior is clearly not consistent with normal system dynamics. However, this evidence accumulates over a time span that is substantially longer than the 40 s of temporal information that was portrayed in the displays. Thus the temporal information for the stuck valve trials also specifies a break in system constraints, although the evidence was less obvious than that for the reservoir leak because of the 40-s time frame that was chosen for presentation. These analyses form the basis for the ensuing interpretations.

Effects of temporal information on performance. The two displays with temporal information produced substantial performance advantages for the estimation of compensated level during the final two onset times. A total of 16 comparisons for the accuracy of compensated level estimates (absolute error, signed error) indicated significantly better performance for both the trend display (Figure 3b, thin outlines) and the tunnel display (Figure 3c, gray fill) relative to the baseline display; there were no significant comparisons favoring the baseline display.

These results are consistent with the conceptual analysis of temporal information that was outlined in the Introduction. The value of compensated level during the final two onset times was extremely context sensitive: It could be equal to (Figure 5a), greater than (Figure 5b), or less than (Figure 5c) the value of indicated level. The significant performance advantages for the time tunnel and trend displays suggest

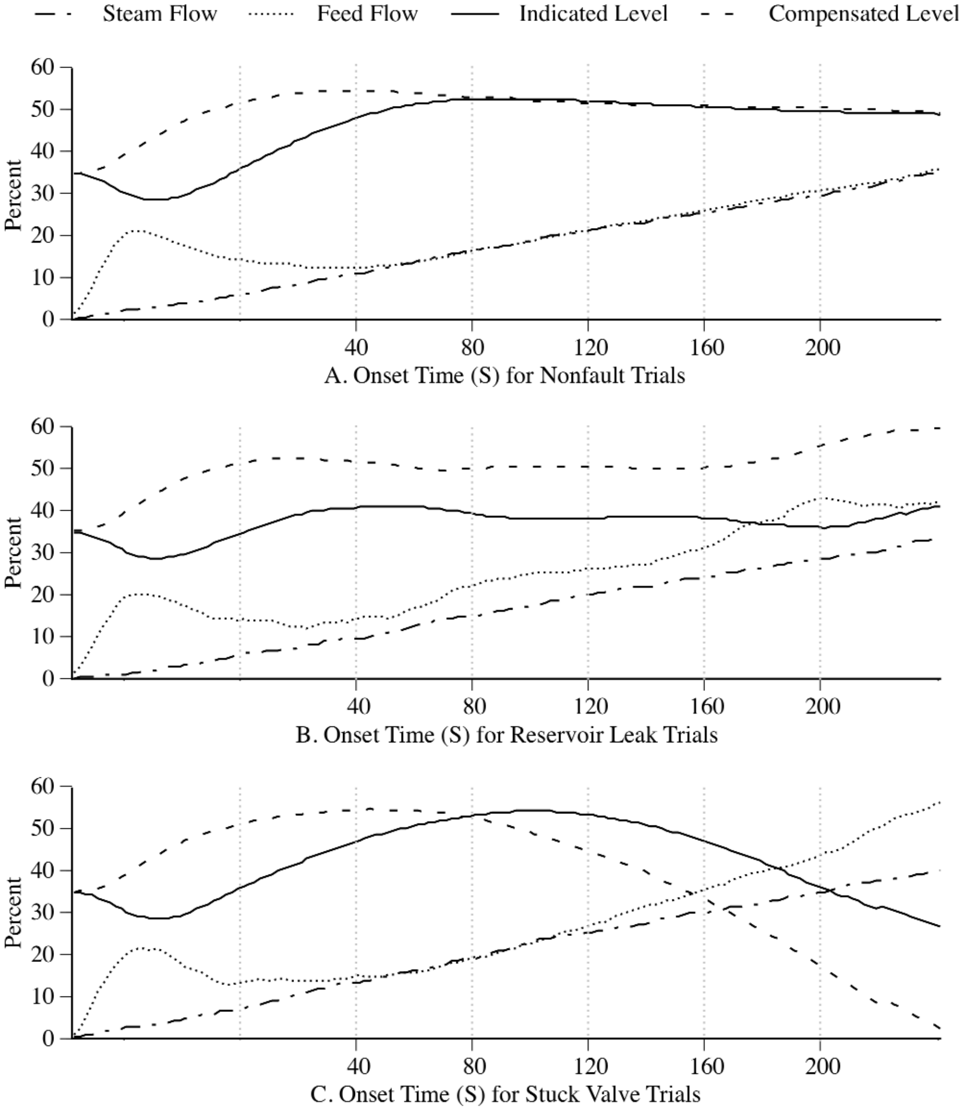


Figure 5. Average value of system variables for display and onset times during (a) nonfault, (b) reservoir leak, and (c) stuck valve trials in Experiment 1. The values of compensated level are for illustrative purposes only.

that participants used temporal information to fine-tune their estimates of compensated level during the final two onset times (Figures 3 and 4). The temporal information provided a context that was very useful in understanding current and future system states. Participants used this context, in combination with knowledge about system dynamics, to form more accurate estimates of compensated level.

Unlike the time tunnel and trend displays, the baseline display provided only momentary and punctate views of system states. This caused difficulties for estimating compensated level because the momentary configurations of the three variables could be quite ambiguous and uninformative (as outlined previously). The participants appear to have adopted a fairly simple and context-free strategy to deal with this uncertainty: “Take the value of indicated level and add some number of units to form an estimate of compensated level.” The evidence supporting this interpretation will be discussed in the following section.

In contrast to the final two onset times, during the initial onset times participants produced more accurate estimates of compensated level with the baseline display than with the temporal displays (Figures 3a and 3b). The positive results for the baseline display are somewhat misleading. During the initial two onset times the participants' control inputs (increases in feed flow) produced large positive net inflows of mass and therefore larger values of compensated level relative to indicated level (see Figure 5). Therefore the context-free strategy of simply adding some number of units to the value of indicated level would produce reasonable performance during the initial two onset times. This strategy would also produce poor performance during later onset times, as outlined previously. More specifically, participants should overestimate compensated level during nonfault trials (Figure 5a) and stuck valve trials (Figure 5c) when the value of indicated level was equal to or less than the value of compensated level. The results for signed error support this interpretation (Figure 4). In summary, observers had no contextual evidence for fine-tuning estimates of compensated level with the baseline display. They appear to have adopted a simple, context-free response strategy that was effective for the majority of experimental trials (including the initial onset times) but that also produced glaring errors for other conditions. For example, the overestimation of compensated level (Figure 4a) during the stable and balanced conditions that characterized nonfault trials at the final two onset times was completely at odds with system dynamics.

The poor performance in estimating compensated level for the temporal displays during the initial two onset times requires interpretation. The temporal information presented in these displays revealed the substantial mass imbalances and rates of changes that characterized the system during these onset times. Although inaccurate, the estimates of compensated level produced with the time tunnel display were consistent with system dynamics. The results for signed error (Figure 4) indicate that participants correctly predicted the direction of change in future system states but overestimated the magnitude. This suggests that the errors may have been the result of miscalibration, attributable perhaps to the unusual system dynamics present during the initial onset times. The pattern of results for the trend display is somewhat more puzzling. With the exception of the final onset time, there was a consistent tendency to underestimate the value of compensated level.

GENERAL DISCUSSION

Extensive analyses of the manual control of feedwater task have revealed that understanding and anticipating the impact of mass and energy effects on the value of indicated level is a key component of successful performance (Roth & Woods, 1988). The compensated level variable is a form of decision support that was explicitly designed to incorporate this knowledge (U.S. Patent No. 4,770,841, 1988 [Haley & Woods, 1988]). Empirical evaluations of compensated level have indicated that its presence significantly improved performance at a variety of control and fault detection tasks (e.g., Bennett & Zimmerman, 2001). These conceptual analyses and empirical results strongly suggest that the capability to accurately estimate values of compensated level constitutes evidence for an improved understanding of system dynamics. Similarly, the capability to estimate this variable more accurately with one display, relative to another, constitutes evidence for more effective design.

The time tunnel display produced the best overall performance for the estimation of compensated level in Experiment 2. The performance advantages for this display relative to the trend display were universal and unequivocal: The time tunnel display consistently produced significant performance advantages that were distributed across all onset times (Figure 3a). The time tunnel display also produced better overall levels of performance than did the baseline display (Figure 3c). The results obtained during the later onset times, when the value of compensated level was highly context dependent, revealed that the time tunnel display produced significantly better estimates on a consistent basis. The most likely interpretation of these results is that the presence of temporal information, in combination with the representational format employed in the time tunnel display, produced a better understanding of current system state (i.e., the contributions of mass and energy effects to current value of indicated level) and a corresponding increase in the capability to estimate values of compensated level (i.e., the value that indicated level that will assume once the transient thermodynamic effects have dissipated).

The presence of temporal information in the trend display also significantly improved estimation performance relative to the baseline display at late onset times during Experiment 2. However, it is clear that the same temporal information had a less positive impact on performance when it was presented in the trend display format than when it was presented in the time tunnel format. There are several potential explanations. One is fairly simple: Participants had less experience interacting with this display than with the time tunnel display. A second potential explanation is that this display violated a population stereotype by presenting current values on the right and historical values to the left. The alternative arrangement (i.e., data flowing from left to right over time) is more consistent with common information displays (e.g., written text) and may have proven more effective. A third possibility is that trend displays are simply not very effective representational formats. Spenkelink (1990) investigated the utility of trend displays for the detection of faults in a chemical plant simulation. He concluded that "the presence of historical [temporal] information, in the form of trends, seriously hampers the early detection of oncoming off-normals" (p. 199). Given that these displays are widely accepted and widely used, further investigations would appear to be in order.

The estimation of compensated level was least effective for the baseline display in Experiment 2. The lack of temporal information resulted in a great deal of uncertainty and ambiguity regarding future system states. To deal with this uncertainty, participants appear to have adopted a simple, context-free strategy (i.e., add a number of units to indicated level). This strategy also produced errors that were glaringly inconsistent with system dynamics at later onset times, when the potential utility of temporal information was at its highest: The faults had propagated through the system, and the corresponding evidence was not being masked by normal thermodynamic effects. It is precisely under these types of conditions that the interface and the associated displays should be providing the most effective decision support.

Thus the overall results for compensated level estimation in Experiment 2 suggest that temporal information improved participants' understanding of current and future system states. This understanding should have translated into an improved capability to detect and identify faults, especially with the time tunnel display. However, the results obtained for fault detection and identification in Experiment 2 were inconclusive. The time tunnel display did produce better average levels of performance for fault detection, particularly under circumstances in which the temporal information had the highest level of diagnostic value (i.e., reservoir leak trials at later onset times). However, only one of the statistical comparisons between displays reached significance (Figure 3a).

Establishing a clear link between temporal information and improved performance at these types of dynamic tasks, which are representative of the tasks that occur in the actual domain, is essential. Any number of factors may have contributed to the lack of performance benefits in the present experiment. However, one factor that is very likely to have had a negative impact on the detection and identification of faults is the amount of temporal information that was presented in the time tunnel and trend displays. In retrospect, the 40-s time frame used in the present evaluation is much smaller than the time frame that characterizes the system dynamics of the manual control of feedwater task. Consider the temporal information contained in each of the isolated individual 40-s time frames, relative to the entire 300-s time frame (see Figure 5). This short time frame is likely to have produced a "keyhole" effect that limited the utility of temporal information; it seems highly likely that a longer time frame would have improved fault detection and identification performance.

These observations underscore a central issue in the design of temporal displays: What is the most effective time frame to be displayed? One design strategy is to analyze the inherent time constants of a system and then pick a single, most appropriate time frame for the temporal information. An alternative strategy is to design the display so that it is capable of presenting a range of time frames and to allow the observer to choose the time frame that is most appropriate. For example, a user might manipulate a graphical slider that dynamically changes the time frame in real time. This would allow flexibility in the selection of a time span that matches current goals, needs, and the system context.

Methodological Insights for the Evaluation of Temporal Displays

The patterns of results obtained in the two evaluations of temporal displays were strikingly different, despite the fact that there was considerable overlap in the experimental manipulations. The results of Experiment 1 provided very little evidence that the time tunnel display improved performance at process control, fault detection, or state estimation tasks. Experiment 2 was conducted with the vast majority of experimental factors held constant, including the tasks (fault detection and estimation of compensated level), the displays (time tunnel and baseline), the participants (a subset of those participating in Experiment 1), and the system states (a subset of those generated in Experiment 1). In sharp contrast to Experiment 1, in Experiment 2 numerous and substantial differences in performance were obtained for the estimation of compensated level.

The most likely interpretation of these divergent empirical results is related to differences in the availability of temporal information in the two settings. In Experiment 1 the fault detection and the state estimate tasks were interleaved with real-time process control. Because no other substantial demands were being imposed on the participants, they were free to concentrate on the process control tasks and to internally integrate (or “absorb”) the unfolding temporal context. As a result, the presence (time tunnel display) or absence (baseline display) of an external representation of temporal information did not have an impact on performance. In Experiment 2 the participants were isolated from the unfolding temporal context: Static snapshots of system states were presented to participants prior to the completion of identical detection and state estimation tasks. This experimental manipulation forced participants to obtain temporal information from the external representations provided by displays. Under these conditions the presence or absence of temporal information did have an impact on performance.

As noted in the Introduction, there are surprisingly few empirical evaluations of temporal displays in dynamic settings, despite their widespread use and acceptance. The divergent results obtained in Experiments 1 and 2 provide some indication of why that might be. The capability to absorb or integrate temporal information will tend to negate performance benefits for external representations of the temporal information when participants can focus exclusively on these displays (as in Experiment 1). The utility of external representations of temporal information will be more apparent when observers do not have access to the unfolding temporal context (as in Experiment 2). This may explain the lack of empirical evidence in dynamic settings: Previous evaluations of temporal displays may have failed to restrict access to the temporal context and therefore failed to obtain supportive evidence.

These results indicate that developing dynamic evaluation settings that restrict access to the unfolding temporal context in a representative manner represents a formidable challenge. It is important to note that *not* having access to the temporal context may actually be more commonplace than having access to the temporal context in applied domains. For example, power plant operators are required to perform a variety of tasks that are interleaved across time; to perform these tasks they are required to consult a number of controls and displays. In addition, the primary displays and the associated displays of temporal information (i.e., strip charts) are presently located in different parts of the control room. Thus the operators cannot afford to focus exclusively on one task or one set of displays. Similar considerations related to temporal information are likely to be found in other domains of application.

Summary

Temporal information is critical in many complex, dynamic domains. However, there does not appear to be a solid empirical basis to assist designers in determining exactly how this information should be presented. The time tunnel display design technique is appealing on several dimensions. It combines the benefits of configural displays (i.e., the emphasis of critical domain properties through emergent features) with temporal information (an explicit representation how these properties have changed over time). In addition, the time tunnel design technique makes effective use of limited display "real estate" by incorporating current and temporal values in a single unified display. The results obtained with the time tunnel technique in the present evaluation are encouraging and complement those of Hansen (1995) with a substantially different implementation and a substantially different set of experimental tasks. Further investigation of the time tunnel display format, as well as alternative representations of temporal information, appears to be a reasonable course of action to pursue.

ACKNOWLEDGMENTS

The authors would like to thank Raja Parasuraman for his assistance in completing the signal detection analyses. The part-task simulation was originally developed at Westinghouse Corporation.

REFERENCES

Bennett, K. B., & Flach, J. M. (1992). Graphical displays: Implications for divided attention, focused attention, and problem solving. *Human Factors, 34,* 513–533.

Bennett, K. B., Nagy, A. L., & Flach, J. M. (1997). Visual displays. In G. Salvendy (Ed.), *Handbook of human factors and ergonomics* (2nd ed., pp. 659–696). New York: Wiley.

Bennett, K. B., Toms, M. L., & Woods, D. D. (1993). Emergent features and configural elements: Designing more effective configural displays. *Human Factors, 35,* 71–97.

Bennett, K. B., & Zimmerman, J. (2001). A preliminary investigation of the time tunnels display design technique. *Displays, 22,* 183–199.

Haley, P. H., & Woods, D. D. (1988). *U.S. Patent Number 4,770,841.*

Hansen, J. P. (1995). An experimental investigation of configural, digital, and temporal information on process displays. *Human Factors, 37,* 539–552.

Lovie, P. (1986). Identifying outliers. In A. D. Lovie (Ed.), *New developments in statistics for psychology and the social sciences* (pp. 44–69). London: British Psychological Society/Methuen.

Pawlak, W. S., & Vicente, K. J. (1996). Inducing effective operator control through ecological interface design. *International Journal of Human-Computer Studies, 44,* 653–688.

Pomerantz, J. R. (1986). Visual form perception: An overview. In H. C. Nusbaum & E. C. Schwab (Eds.), *Pattern recognition by humans and machines: Volume 2. Visual perception* (pp. 1–30). Orlando, FL: Academic Press.

Roth, E. M., & Woods, D. D. (1988). Aiding human performance: I. Cognitive analysis. *Le Travail Humain, 51,* 39–64.

Schutz, H. G. (1961). An evaluation of formats for graphic trend displays: Experiment II. *Human Factors, 3,* 99–107.

Spenkelink, G. P. J. (1990). Aiding the operator's anticipatory behaviour: The design of process state information. *Applied Ergonomics, 21,* 199–206.

Vicente, K. J. (1991). *Supporting knowledge-based behavior through ecological interface design* (EPRL-91-02). Urbana-Champaign: University of Illinois, Department of Mechanical Engineering, Engineering Psychology Research Laboratory.

Kevin B. Bennett is a professor in the Department of Psychology, Wright State University. He received his Ph.D. in applied experimental psychology from the Catholic University of America in 1984.

Michael Payne manages a research and design organization at Intel Corporation in Hillsboro, Oregon. He received an M.S. in human factors psychology from Wright State University in 1999.

Brett Walters is a human factors engineer at Micro Analysis and Design in Boulder, Colorado. He received a M.S. in human factors psychology from Wright State University in 1997.

Date received: August 5, 2002
Date accepted: November 2, 2004