Effects of Control Automation on Operator Attention and Risk Probability in High Speed Trains

by

Helias John Marinakos

Diploma in Mechanical Engineering
National Technical University of Athens (1994)

Submitted to the Department of Mechanical Engineering
in partial fulfillment of the requirements for the degree of

Master of Science in Mechanical Engineering

at the

Massachusetts Institute of Technology

May 1997

Author________________________________
Department of Mechanical Engineering
May 15 1997

Certified by________________________________
Thomas B. Sheridan
Professor of Engineering and
Professor of Applied Psychology
Thesis Supervisor

Accepted by________________________________
Ain A. Sonin
Chairman, Department Committee on Graduate Studies

Effects of Control Automation on Operator Attention and Risk Probability in High Speed Trains

by

Helias John Marinakos

Submitted to the Department of Mechanical Engineering on April 25, 1997,
in partial fulfillment of the requirements for the degree
of Master of Science in Mechanical Engineering.

Abstract

In high speed rail systems automation is used as a means to compensate for the sensory, perceptual and cognitive limitations of the human operator. At the design level two approaches can be implemented for the operation of such vehicles that span the whole range of automation in use:

a. Fully manual operation of the train with decision aiding, such as preview and advisory displays, assisting in human decision. In such an approach the locomotive engineer actually controls the vehicle.

b. Fully automated operation of the train. In this approach the automated system drives the train and the human operator's role is shifted to that of the supervisor that monitors the proper functioning of the automation.

While the latter approach seems promising there are major safety concerns related to the so-called 'out-of-the-loop' problem, namely the human operator's inability to promptly intervene and act correctly in case of an unexpected event as a result of vigilance decrement, loss of situation awareness and over-reliance on automation.

Consequently one of the focal points of the current research was to determine the appropriate use of automation in high speed rail systems. For the purposes of the current research the Volpe High Speed Train Simulator was used. It exhibits three control modes: fully manual, cruise, and fully automated control. These three levels of automation were compared with regard to their effects on vigilance decrement and situation awareness of the human operator. Vigilance decrement was determined by measuring responses of the human operator in a set of emergency scenarios. Since each emergency scenario represented a different type of detection-response paradigm conclusions were drawn about system design. Situation awareness measurements were conducted using the "freezing the simulation" technique (Endsley, 1994).

The fully automated mode (autopilot) was found to have the best detection rate and least vigilance decrement over time compared with the rest of the control modes. It also had generally superior situation awareness except for speed awareness in which cruise and manual control outperformed the autopilot. Additionally while the autopilot was associated with less workload as far as system monitoring was concerned, it imposed greater stress when reacting to an obstruction.

High speeds reduce the allowable time of the vehicle operator to respond to an emergency, thus increasing the risk probability of the system. A probabilistic theory -known as Markov Renewal Theory- was utilized for the purpose of quantifying and tracing the time path of those probabilities. Exposition of this theory constitutes a separate contribution of the thesis.

Thesis Supervisor: Thomas B. Sheridan
Title:. Professor of Engineering and Applied Psychology and
Professor of Aeronautics and Astronautics

Acknowledgments

Upon the completion of this research I would like to acknowledge the contribution of many people who have been part of my endeavors before coming and during my stay at MIT.

First and foremost I feel the need to extend my gratitude to Professor Tom Sheridan, who has been my advisor, mentor and friend, for his guidance and support he offered me throughout the course of this research. He shared with enthusiasm many of my ideas, and encouraged me to purse them further always making critical remarks. His extensive knowledge on human machine systems, has been an invaluable source of intellectual wealth which has been instrumental in the development and completion of this work.

My association with the Human Machine Systems Lab has provided me with the opportunity to meet many bright people. I had the pleasure working along with Steven Villareal and Jacob Einhorn on the train simulator and really enjoyed the cooperation with both of them. Special thanks go to Dr. Edward Lanzilotta for providing his knowledge regarding software engineering issues and experimental procedures whenever I requested it. I would also like to acknowledge the support from the other group members as well: Jianjuen Hu for his companionship during the summer of 1996, Shinsuk Park who participated in my experiments, Mark Ottensmeyer, Nicholas Patrick, Suyeong Kim, Dave Schloerb, Shih-Ken Chen, Jie Ren, Steven Landry, Joe Conti, Shumei Yin Askey and Bernardo Aumond. Special thanks go to Michalis Kilaras. Our common interests helped forge a strong friendship.

This work was completed at the John A. Volpe National Transportation Systems Center who was the sponsor of the project. The opportunity to work in this facility lead to many productive relationships. I would like thank Dr. Jordan Multer for sharing his extensive knowledge on experimental design and providing insightful comments with regard to the present research, John K. Pollard for fixing the broken joystick in 'no time' when I had to run experiments and finally Dr. Donald Sussman for his support and interest in the project.

I owe a great deal of respect to my undergraduate advisor Professor Kyriakos Papailiou for the constant support and encouragement to purse graduate studies specifically at MIT. Thanks go as well to many friends that I made back in Greece.

Last but not least I want to acknowledge the love I have been given from my family, my father Ioannis, my mother Paraskevi and my sister Aggeliki - Eleni, during the course of my life. I wouldn't have made it all the way here, shouldn't have been there to support me in every aspect of my life. To those this work is dedicated.

Στον Πατερα μου Ιωαννη
την Μητερα μου Παρασκευη
και στην αδελφη μου Αγγελικη–Ελενη

Contents

CHAPTER 1

INTRODUCTION

1.1 Motivation and Objectives

Rail operation has played a central role in mass transportation incorporating advantages associated with operational efficiency, energy consumption and environmental cost. The aforementioned advantages become more salient in high speed trains like the ones that exist in the European countries and Japan. Examples include the Intercity Express (ICE) in Germany, the Train a Grand Vitesse (TGV) in France, and the Shinkansen - known as the "bullet train"- in Japan which are capable of achieving speeds at the range of 250 to 320 km/h. And the trend is towards even greater speeds: Experiments conducted with magnetically levitated trains (maglev), such as the German TR-07, have proven the technological feasibility of achieving speeds up to 500 km/h [3]. Of major concern though are the safety implications that arise from speed increase and are related to human physical and/or perceptual limitations as illustrated below:

1. The greater the speed the more difficult is for the human operator to see the wayside signals and act in time. A study conducted in France has shown that the maximum speed for accurate driver perception is 220 km/h [2].

2. The higher the speed the lesser the amount of time that is available for reaction by the locomotive engineer in a case of an emergency, due to the increased distance required for the vehicle to slow down or even come to a complete stop.

Contemporary technology offers one potential solution called *automation*. Automation has been introduced in many domains such the aviation industry, nuclear power plants and modern manufacturing systems - just to name a few - as a means of rendering their operation safer and more efficient, and the general consensus is that it has succeeded to a great extent. Unfortunately one cannot transfer the successful application of automation from one domain to another as each one is characterized by its own particular attributes. One such attribute is the fact that rail vehicles are constrained to move on fixed tracks and the view from the vehicle is often quite limited relative to the distance traveled. As a result the vehicle cannot change course should an emergency arise in stark contrast to an air or car vehicle. Another one is related to the *boredom* and *fatigue* that train driving conveys to the locomotive engineers. Trains cover large distances mostly through isolated areas and the locomotive engineer has to sustain prolonged periods of rail operation without human interaction and remain vigil and alert while "nothing happens". As a result the operator gets bored and becomes less attentive over the course of vehicle operation. There are several reports referring to locomotive engineers falling asleep while in duty and studies regarding effects of sleeping patterns on their performance [22]. Consequently one needs to be very careful as to the *degree* and *level of automation* that is most appropriate for application in rail vehicles and take into account potential side effects of the human-machine interface relative to the attributes of train driving like the ones outlined above.

At the design level two approaches can be implemented for the efficient operation of such vehicles that span the whole range of automation in use:

a. Fully manual operation of the train with various aids - such as preview and advisory aids [2] - assisting in human decision. In such an approach the locomotive engineer actually controls the vehicle and makes decisions regarding vehicle operation.

b. Fully automated operation of the train. In this approach - referred to as *automation in control* - all necessary information for the operation of the vehicle is passed to the automated system which is responsible for operating the train according to prescribed rules. The human operator's role is shifted to that of a supervisor which monitors the proper functioning of the automated system.

The level of automation used in the current operational high-speed rail systems (ICE, TGV, Shinkansen) is a blend of the two extreme approaches cited above. According to a study related to safety in high speed trains "the German philosophy (ICE) of rail development emphasizes automated control with use of the human as a system monitor, while the French (TGV) and Japanese (Shinkansen) depend more on the human for control decisions" [3]. A feature common to all of the above mentioned rail systems is the fact that the high speed train occupies a single track itself and no other rail vehicle can use it. Here in the U.S. the design plans for the introduction of high speed trains foresee that tracks should be shared by both high and low speed trains (passenger or freight) for the efficient utilization of the existing network of tracks. Hence even though fully automated rail operation is within the capabilities of current technology, the human operator is still necessary to be physically present at the cab in case the automation fails, to handle an unexpected event such an obstruction, a bad track condition, a discrepancy between an incab and a wayside signal, or even to reduce public anxiety. Questions regarding his capacity to take over control in case of an emergency are yet to be clarified and are often related to the philosophy behind the automation in use. Human operator inability to successfully and promptly intervene in case of an unexpected event deteriorates as a result of vigilance decrement, loss of situation awareness and manual skill and reliance on automation. Furthermore there exists a substantial body of evidence in the literature regarding vigilance, which suggests that "monitoring a system for potential failures is a role for which humans are poorly suited" [12].

Consequently one of the focal points of the current research has been to determine the appropriate use of automation in high speed rail systems with regard to attributes specific to train driving such as boredom and fatigue. For the purposes of the current research the Volpe High Speed Train Simulator was used to simulate a high speed train driving environment in which "nothing happens for long time". Towards this end the number of emergencies was kept to an extremely low level. The simulator exhibits three control modes for the operation of the train, namely manual, cruise, and fully automated control. Under manual control the human is solely responsible for controlling the speed of train, while in cruise control the human operator sets the desired speed (according to the signals or the civil speed limits) and the automation is responsible for achieving and maintaining it. Operation of the vehicle under fully automated control requires no human intervention as the automation is programed to recognize signals and civil speed limits and adhere to the lesser of the two. These three levels of automation were compared with regard to the vigilance decrement and situation awareness of the human operator. Vigilance decrement was determined with respect to responses by the human operator in a set of emergency scenarios. Since each emergency scenario represented a different type of detection-response paradigm conclusions where drawn related to system design. Situation awareness measurements were conducted using the "freezing the simulation" technique [14], [15].

The other major goal of this research was related to safety analysis. Both deterministic and stochastic models have been employed toward this end, the latter ones being the most promising for modeling human-machine systems due to the variability and unpredictability in the perception, decision and action that humans exhibit. As stated above, high speeds reduce the allowable time for the vehicle operator to respond to an emergency within certain time constraints, thus increasing the failure probability of the integrated human-machine system. The current research aimed at providing the theoritical foundations needed to model stochastic systems in such a way that the evolution over time of these "failure probabilities" can be quantified. Towards this end a probabilistic theory, known as Markov Renewal Theory, was utilized for the purpose of quantifying and tracing the time path of those probabilities. A simple model was then developed to capture these "failure probabilities" in the "high" versus the "low" speed regimes of train driving.

1.2 Thesis Overview

The goal of the thesis was to examine the effects of automation on human attention and fatigue in high speed trains and provide the means for modeling and quantifying the risk probability in the "high" versus the "low" speed regime. Chapter 2 gives the necessary background on Vigilance and Situation Awareness. Detection theory as applied to model human observers is presented along with a literature review on vigilance studies. The measurement of situation awareness is presented in the context of two recent articles. Chapter 3 introduces the reader to a powerful tool drawn from the field of Applied Probability: Markov Renewal Theory and the concept of semi-regeneration of stochastic processes are presented. These concepts are then taken further and developed to form the results used by the thesis. Chapter 4 gives an overview of the Volpe High Speed Train Simulator, addresses software engineering issues and gives a detailed description of the experiment. Chapter 5 summarizes and comments on the results of this work.

CHAPTER 2

BACKGROUND ON VIGILANCE AND SITUATION AWARENESS

2.1 Vigilance

Vigilance refers to the capacity of the human operator to sustain attention and remain alert to stimuli over prolonged periods of time. Vigilance studies have their origin in World War II when the Royal Air Force (RAF) commissioned Norman H. Mackworth to study the radar observers' decrement in detection rate of enemy submarines, which occured after only about 30 minutes. His studies marked the beginning of research aimed at discovering factors that might be related to what is known as the *vigilance decrement.*

2.1.1 Introduction to Detection Theory

Vigilance studies have been conducted mainly in the context of yes-no experiments: the observer is presented a set of stimuli some of which contain a "signal" to be distinguished from a "noisy" background and the observer responds "yes" to each particular stimulus that contains the signal and "no" to the ones that contain just the noise and not the signal. Correctly recognized stimuli that contain the signal, are termed *hits,* whereas mistakenly recognized ones *false alarms.* If the observer failed to recognize a signal stimulus and responded "no", then he is credited with a *miss,* while when he correctly didn't recognize any signal he has made a *correct rejection.* One can summarize the input stimuli and the responses using a table as the one presented below:

	Response	
Stimulus class	“yes”	“no”
Signal	Hits	Misses
Noise	False Alarms	Correct rejections

Table 2.1: Summary of Stimuli and Responses

The hit and false alarm rates can be written as probabilities of “yes” responses conditional on the possible stimuli:

Hit: $H = P(\text{“yes”} \mid \text{Signal})$

False Alarm: $F = P(\text{“yes”} \mid \text{Noise})$

Misses and correct rejection probabilities are (1- hit rate) and (1-false alarm rate) respectively.

One way to characterize the vigilance decrement is the decline in the percentage of correctly detected signals over time, and this decline is often regarded as resulting from a deterioration in the observer’s perceptual sensitivity. However the success in correctly detecting a signal might not always be a function of perceptual ability. Observers’ reports might depend on nonperceptual factors related to the anticipated gains or losses with regard to correct or incorrect responses. Such factors affect their willingness to respond “yes” rather than “no” and hence the vigilance decrement may or may not be attributed to decline in perceptual sensitivity.

For such reasons the Theory of Signal Detection [5], [6], [7], [8] has been adopted and used in vigilance research. The most commonly used measure to characterize the perceptual sensitivity of the observer is defined in terms of the inverse of the normal distribution function (figure 2-1):

$$d' = z(H) - z(F)$$

which converts the hit and false alarm rates to a z-score that is in standard deviation units. A proportion of 0.5 is converted into a z-score of 0, larger proportions into positive z-scores and smaller ones to negative z-scores. Index d' is an invariant measure of sensitivity and takes non-negative values with 0 representing inability to differentiate between critical and non-critical signals. Perfect accuracy, on the other hand, implies infinite d'. To avoid infinite values it is common to convert proportions of 0 and 1 to 1/(2N) and 1-1/(2N) where N is the total number of stimuli that were presented to the subject [8].

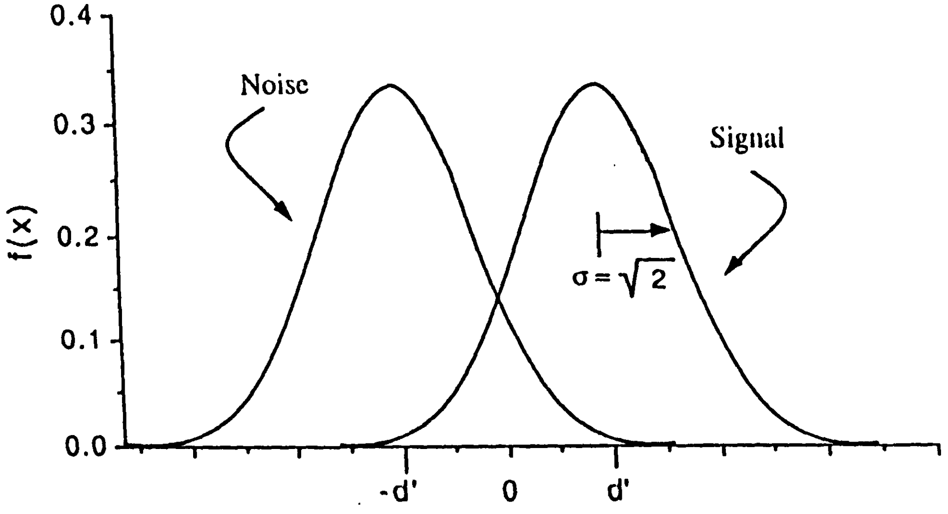


Figure 2-1: Underlying distributions of Signal and Noise

2.1.2 Literature Review on Vigilance Studies

Since the original work of N. Mackworth a variety of experiments have been conducted to study vigilance behavior and the vigilance decrement has been replicated consistently. The dominant characterization on vigilance performance has been the R. Parasuraman and D. R. Davies taxonomy of vigilance [9] according to which individual differences in vigilance performance are not so much task specific as task-type specific. Hence it was suggested that classification of the vigilance literature can lead to an improved specification of the types of tasks in which reliable decrements in efficiency occur in terms of a few dimensions of the viginalce task taxonomy.

In a later article [10] Parasuraman identified two such dimensions: the type of discrimination (successive versus simultaneous) and the event rate. Successive discrimination tasks are the ones that require the observer to distinguish a signal from a non-signal reference when these are presented successively, as opposed to simultaneous discrimination tasks in which signal and non-signal features are presented simultaneously within the same stimulus event. The successive discrimination tasks impose greater memory load since the observer has to distinguish a target from a non-target presented in recent memory. From his study he concluded that the vigilance decrement results from a decrement in perceptual sensitivity only if (i) target discrimination loads memory and (ii) the event rate is high. Otherwise the decrement reflects shifts in the response criterion over time. Another dimension on the vigilance task taxonomy that he added later was related to the level of signal discriminability. According to his studies for highly or moderately discriminable signals perceptual sensitivity declined for the successive discrimination tasks but not for simultaneous tasks when the event rate was high. For low discriminability signals perceptual sensitivity declined in both types of tasks.

While vigilance has been studied extensively in the context of laboratory tasks, relatively few studies have been performed in environments that closely simulate actual work settings, the main reason being the difficulty of applying signal detection theory. And from those the majority focuses on the aviation industry. Another feature that questions the extent to which data from laboratory tasks can be extrapolated to operational tasks is the event rate used. To the overwhelming majority of vigilance studies a target rate of 1 target per minute is considered low in sharp contrast with reality in which "events" that need to be distinguished and act upon, such as failures, are much more rare. Very few studies have been preformed with a very low target rate of the order of 1 target per 30 minutes [12], [13]. This is where one of the focal points was for the purposes of the current research: to measure vigilance decrement in an environment that replicates to a great extent an operational system using a very low target rate.

2.2 Situation Awareness

According to Endsley, Situation Awareness (SA) can be defined as a person's "perception of the elements of the environment within a volume of time and space, the comprehension of their meaning and the projection of their status in the near future". At the lower level of SA the person perceives the information from the operational system, extracts the necessary data and integrates them in a meaningful way at the second level, and at the third and higher level, based on the understanding he has, future events and states of the system can be predicted and hence decisions can be made.

The enhancement of situation awareness is a major goal for the human factors engineer, that must be taken into account during the development phase of operational systems. As a construct it is extremely useful since it provides an objective test-bed on which the engineer can evaluate different design concepts ranging from automation and displays to attention and performance.

Many methodologies for the empirical measurement of situation awareness have been proposed. The most promising one is the "*freeze technique*" whereby the simulation is frozen at random times and subjects are queried about their perception of the system at that time [14], [15]. With this technique the system displays are blanked and the simulation is suspended while the subjects answer questions related to the system. In this way data can be collected immediately, which reduces the problems incurred when collecting data after the task is over. Recent studies on situation awareness conducted by Endsley [15] using this technique provided evidence that it is not intrusive on normal subject behavior during the trials and doesn't suffer from limitations of human memory.

CHAPTER 3

MARKOV RENEWAL THEORY

This chapter introduces the main ideas of Markov Renewal Theory. Though the exposition draws heavily from [20] and [21] the notation used is somewhat different. The emphasis has been in the presentation of the theory rather than the mathematical rigor as the later can be found in the references cited above.

3.1 Markov Renewal Processes

Let N be the set of non-negative integers and R_+ the set of non-negative real numbers. Define X_n and T_n as random variables on a probability space $(\Omega, \Re, P)$ such that for each $n \in N$, X_n takes values in some fixed countable space E called the state space:

$$X_n\colon \Omega \rightarrow E \qquad \mathrm{T}_n\colon \Omega \rightarrow R_+$$

and $0 = T_0 \le T_1 \le T_2 \le \ldots$. The sequence of pairs (X_n, T_n) is said to form a *Markov Renewal Process* with state space E if:

$$Pr[X_{n+1} = j, T_{n+1} - T_n \le t | X_0, X_1, \ldots, X_n = i, T_0, T_1, \ldots T_n] = Pr[X_{n+1} = j, T_{n+1} - T_n \le t | X_n = i] \quad (3\text{-}1)$$

Note that X_n is the state of the process at the n^{th} transition and T_n is the time that the n^{th} transition occured. What the above equation states is just a generalization of the ordinary Markov Processes: Namely not only the state but also the time for the next transition are random variables. Throughout the discussion we will assume (without any loss of generality) that the process starts at $T_0 = 0$ and that the probabilities in (3-1) do not depend on n which implies that the process has stationary transition probabilities (the process is homogeneous). The probabilities:

$$Q_{ij}(t) = Pr[X_{n+1} = j, T_{n+1} - T_n \leq t | X_n = i] \quad (3\text{-}2)$$

are the *one step transition functions* of the process. The matrix $\underset{\sim}{Q}(t)$ for which $Q_{ij}(t)$ is the i,j element is called the *Semi-Markov kernel* of the process (X_n, T_n). For each pair (i, j) the function $t \to Q_{ij}(t)$ is non-decreasing right continuous, bounded and it is assumed that $Q_{ij}(0) = 0$. By taking the limit as time goes to infinity one can obtain:

$$\lim_{t \to \infty} Q_{ij}(t) = Q_{ij}(\infty) = Pr[X_{n+1} = j | X_n = i] = P_{ij}$$

That is, the P_{ij} are the one-step transition probabilities for some Markov chain with state space E. Hence the sequence X_n forms a Markov chain. On the other hand we can consider the state space E of consisting of a single point. In this case the times at which the process transitions $(T_n)_{n \in N}$ form a *renewal process*. Equation (3-2) can be written as:

$$Q_{ij}(t) = Pr[X_{n+1} = j, T_{n+1} - T_n \leq t | X_n = i] = P(i, j) \cdot Pr[T_{n+1} - T_n \leq t | X_{n+1} = j, X_n = i] = P_{ij} \cdot F_{ij}(t)$$

where $F_{ij}(t) = Pr[T_{n+1} - T_n \leq t | X_{n+1} = j, X_n = i]$ or $F_{ij}(t) = Q_{ij}(t)/P_{ij}, \quad i, j \in E, t \in R_+$ (3-3)

Equations (3-3) give the temporal distribution at which the process transitions given the current state of the process and the next state that will visit. From them one can compute the probability distribution function for the time that the process will spend in state *i* regardless of the next state that will be visited:

$$F_i(t) = Pr[T_{n+1} - T_n \leq t | X_n = i] = \sum_{j \in E} Q_{ij}(t) = \sum_{j \in E} Pr[T_{n+1} - T_n \leq t, X_{n+1} = j | X_n = i] \quad (3\text{-}4)$$

It should be noted that the random variables $T_{n+1} - T_n$, $n \in N$ which are the times between two successive transitions of the process are *not independent* (since they depend on the next state to be visited) but are independent conditionally on the path of the Markov chain X_n. That is, if one prescribes the states the process X_n, $n \in N$ has visited then the increments $T_{n+1} - T_n$, $n \in N$ will be independent.

3.2 The Markov Renewal Function

Throughout this section $(X_n, T_n)_{n \in N}$ will be a homogeneous Markov renewal process with a semi-Markov kernel Q over a finite space E starting at $T_0 = 0$. Then $(X_n)_{n \in N}$ is a Markov chain on E with one step transition probabilities $Q_{ij}(\infty) = P_{ij}$ underlying the Markov renewal process as mentioned above. We now introduce the two step transition functions for the Markov renewal process as:

$$Q^{(2)}{}_{ij}(t) = Pr[X_{n+2} = j, T_{n+2} - T_n \le t | X_n = i] = Pr[X_2 = j, T_2 \le t | X_0 = i] \quad (3\text{-}5)$$

where the last equality holds due to the homogeneity property. Using the standard renewal arguments we condition on the time of the first renewal:

$$Q^{(2)}{}_{ij}(t) = Pr[X_2 = j, T_2 \le t | X_0 = i, T_1 \le t] \cdot Pr(T_1 \le t) + Pr[X_2 = j, T_2 \le t | X_0 = i, T_1 > t] \cdot Pr(T_1 > t)$$

and noting that $Pr[X_2 = j, T_2 \le t | X_0 = i, T_1 > t] = 0$ the equation above can be written further as:

$$Q^{(2)}{}_{ij}(t) = Pr[X_2 = j, T_2 \le t, T_1 \le t | X_0 = i] = \sum_{k \in E} \int_0^t Pr[X_2 = j, T_2 \le t, X_1 = k, T_1 = ds | X_0 = i] \Rightarrow$$

$$Q^{(2)}{}_{ij}(t) = \sum_{k \in E} \int_0^t Pr[X_2 = j, T_2 - T_1 \le t - s, X_1 = k, T_1 = ds | X_0 = i] \quad (3\text{-}6)$$

The probability $Pr[X_2 = j, T_2 - T_1 \le t - s, X_1 = k, T_1 = ds | X_0 = i]$ can be written as:

$Pr[X_2 = j, T_2 - T_1 \le t - s | X_1 = k, T_1 = ds, X_0 = i] \cdot Pr[X_1 = k, T_1 = ds | X_0 = i]$ and using the definition of Markov renewal processes equation (3-6) takes the form: $Q^{(2)}{}_{ij}(t) = \sum_{k \in E} \int_0^t dQ_{ik}(s) \cdot Q_{kj}(t-s)$ (3-7) where the derivative is taken with respect to the time variable of the one step transition functions $Q_{ij}(t)$.

By induction one can show in that in general: $Q^{(n)}{}_{ij}(t) = \sum_{k \in E} \int_0^t dQ_{ik}(s) \cdot Q^{(n-1)}{}_{kj}(t-s)$ (3-8) which are the *n-step transition functions* of the process. Equations (3-8) can be put in matrix form as follows:

$$\underset{\sim}{Q}^{(n)}(t) = \int_0^t d\underset{\sim}{Q}(s) \cdot \underset{\sim}{Q}^{(n-1)}(t-s) = (\underset{\sim}{Q} \otimes \underset{\sim}{Q}^{(n-1)})(t)$$

where the again the matrix derivative is take with respect to the time variable of the *kernel* $\underset{\sim}{Q}(t)$.

Defining the indicator function: $1_A(X) = \begin{cases} 1 & \text{if } X \in A \\ 0 & \text{otherwise} \end{cases}$ we can define two random variables as follows:

$1_j(X_n) = \begin{cases} 1 & \text{if } X_n = j \\ 0 & \text{otherwise} \end{cases}$ which takes the value *1* whenever the process visits state *j* and zero otherwise

$1_{[0,t]}(T_n) = \begin{cases} 1 & \text{if } T_n \in [0, t] \\ 0 & \text{otherwise} \end{cases}$ which likewise takes the value *1* whenever the n^{th} transition of process occurs before *t* and zero otherwise.

It's worthy to note that the above two defined random variables are not independent and their product will be *1* only when the event $\{X_n = j, T_n \le t\}_{n \in N}$ occurs, namely only when the n^{th} transition is in state *j* **and** it occurs before time *t*. The sum:

$$\sum_{n=0}^{\infty} 1_j(X_n) \cdot 1_{[0,t]}(T_n)$$

is a random variable its self and gives the number of transitions that the process has made in *j* in the interval [0,t]. The expectation of the above random variable is found by applying the expectation operator $E_i(\ \cdot\ )$ which denotes the conditional expectation $E[\ \cdot\ | X_0 = i]$ with respect to the initial state *i*:

$$R_{ij}(t) = E_i\left[\sum_{n=0}^{\infty} 1_j(X_n) \cdot 1_{[0,t]}(T_n)\right]$$

By exchanging the summation with the expectation operation and noting that the expectation of the indicator function is just the probability that the event has occured the equation above can be written as:

$$R_{ij}(t) = \sum_{n=0}^{\infty} Q_{ij}^{(n)}(t) \quad \forall\ i, j \in E \text{ (3-9)}$$

where $R_{ij}(t)$ is the expected number of visits to state *j* in the time interval [0,t] for the process that started at $T_0 = 0$ in state *i*. For every *i, j* in the state space *E* we can define $\underset{\sim}{R}(t)$ to be the matrix whose *i, j* element is $R_{ij}(t)$:

$$\underset{\sim}{R}(t) = \sum_{n=0}^{\infty} \underset{\sim}{Q}^{(n)}(t) \text{ (3-10)}$$

This matrix is called the *Markov renewal function* and plays a very important role in the theory of *semi-Markovian processes* as we will see later on. In fact the main result of this thesis is that if state j is a *trapping state* for the *Markov renewal process* then the renewal function in (3-9) gives the probabilities that the associated *semi-Markovian processes* starting from each state i has moved in the trapping j state earlier than time t and so one by proper identification of states can trace the time path of those probabilities. For now we will write equation (3-10) as:

$\underset{\sim}{R}(t) = I + \underset{\sim}{Q}(t) + \underset{\sim}{Q}^{(2)}(t) + \ldots$ and observing that $(\underset{\sim}{Q} \otimes \underset{\sim}{R})(t) = \underset{\sim}{Q}(t) + \underset{\sim}{Q}^{(2)}(t) + \ldots$ ($\otimes$ is the convolution operator as defined above) we can write (3-10) as $\underset{\sim}{R}(t) = I + (\underset{\sim}{Q} \otimes \underset{\sim}{R})(t)$ (3-11) or equivalently in component form we get a set of integral equations as follows:

$$R_{ij}(t) = \begin{cases} 1 + \sum_{k \in E} \int\limits_0^t dQ_{ik}(s) \cdot R_{kj}(t-s), & i = j \\ \sum_{k \in E} \int\limits_0^t dQ_{ik}(s) \cdot R_{kj}(t-s), & i \neq j \end{cases} \qquad (3\text{-}12)$$

By using the Laplace transform of the time derivatives of the single step transition functions:

$$Q'_{ij}(s) = \int\limits_0^\infty e^{-st} dQ_{ij}(t)$$

and taking the Laplace transforms in equation (3-11) or (3-12) we obtain a closed form solution in the frequency domain:

$$\underset{\sim}{R}(s) = \underset{\sim}{I}(s)(\underset{\sim}{I} - \underset{\sim}{Q}'(s))^{-1} \quad (3\text{-}13)$$

where the $\underset{\sim}{I}(s)$ is the Laplace transform of the identity matrix and $\underset{\sim}{I}$ is the usual identity matrix. If the state space E is finite then the $R_{ij}(t)$ satisfying (3-11) or (3-12) are unique. However if the state space is infinite then a solution satisfying (3-11) or (3-12) might not be unique.

3.3 Markov Renewal Equations

The equations we will introduce form a system of integral equations which are often encountered in the study of renewal processes. We shall present the solution to such equations as this will be utilized later on in the study of *semi-Markov*ian processes.

Let's consider a homogeneous Markov renewal process $(X_n, T_n)_{n \in N}$ with a finite state space E and semi-Markov kernel $\underset{\sim}{Q}(t)$, and let $\underset{\sim}{R}(t)$ be the corresponding Markov renewal kernel. The class of functions we will be working with, to be denoted by ***B*** are the functions:

$$f: E \times R_+ \rightarrow R$$

continuous or right continuous and monotone in the second variable (well behaved probability functions).

A function $f \in$ ***B*** is said to satisfy the *Markov renewal equation* if for some function $g \in$ ***B***:

$$f_i(t) = g_i(t) + \sum_{k \in E} \int_0^t dQ_{ik}(s) \cdot f_k(t-s), \quad i, k \in E, t \in R_+ \quad (3\text{-}14)$$

The equations above has a unique solution which is given by:

$$f_i(t) = \sum_{k \in E} \int_0^t dR_k(s) \cdot g_k(t-s), \quad i, k \in E, t \in R_+ \qquad (3\text{-}15)$$

Proof of the existence and uniqueness of the above solutions in the case of finite state space E are provided in the references [20] and [21] among others.

3.4 Semi-Markov Processes

Let's consider again a homogeneous Markov renewal process $(X_n, T_n)_{n \in N}$ with a finite state space E and semi-Markov kernel $\underset{\sim}{Q}(t)$. We can define a continuous-time parameter process $Y = (Y_t)_{t \in R_+}$ on E such that:

$$Y_t = X_n \quad \text{if } T_n \leq t < T_{n+1} \ (3\text{-}16)$$

The stochastic process defined by (3-16) is called a semi-Markov process associated with $(X_n, T_n)_{n \in N}$. The times $0 = T_0 \leq T_1 \leq T_2 \leq \ldots$ are the successive times the process Y_t transitions and $X_0, X_1, X_{2, \ldots}$ are the successive states visited. The word *semi-Markov* comes from the somewhat limited Markov property that Y_t enjoys, namely the present and the past of the process are independent given the present state **only** at the times T_n the process transitions. If the times between transitions are exponentially distributed and depend only on the current state and not to the next one visited, then the semi-Markov process becomes a temporally homogeneous Markov process.

In such a case the Markov transition functions are of the form: $Q_{ij}(t) = P_{ij}(1 - e^{-\lambda(i)t})$. This means that at **any** time instant on the real line the Markov process is regenerated continuously in accordance with the memoryless property of the exponential distributions. In other words every time instant the future and the past of the process are independent. In contrast the same is true for semi-Markov processes **only at discrete time instants** in the real line. These **discrete time instants** are the times T_n at which the process transitions. To put is in other words in the case of a Markov process one can shift the time origin of the process at any point in the real line whereas in a semi-Markovian process the time origin can be shifted only at the time instants T_n at which the process transitions. The transition probabilities of a semi-Markov processes can be studied by the usual renewal arguments as done in [20] and [21]. However to give more insight on such processes we will derive the same result using the notion of semi-regeneration. But first we introduce the idea of semi-regeneration.

3.5 Semi-Regenerative Processes and Stopping Times

A random variable $T: \Omega \rightarrow R_+$ is called a stopping time for the process Y_t if the event $\{T \leq t\}$ can be determined from the history of the process Y_t up to time *t*. To give a feel of what stopping times are we present some examples:

- Consider a Markov chain. The time at which the process visits a fixed state for the first time is a stopping time for the process because one can observe the process and be able to tell whether the process has visited the fixed state by time *t*.

- Consider a Markov chain that has some transient states. The random variable defined as "time at which the process visits a fixed state for the last time" is **not** a stopping time since one is not able to tell whether the process has visited the fixed state for the last time just by looking at the history of the process up to time *t*. He needs to observe the whole history of the process in order to able to tell which time the process exited the fixed state for ever. But this contradicts the definition a stopping time.

The stochastic process $Y = (Y_t)_{t \in R_+}$ is said to be *semi-regenerative* if there exists a Markov renewal process $(X_n, T_n)_{n \in N}$ with finite state space E imbedded in Y such that:

a) for each $n \in N$, T_n is a stopping time for Y

b) for each $n \in N$, X_n is determined by $\{Y_s : s \le T_n\}$

c) for each $n \in N$, $m \ge 1, 0 = t_0 \le t_1 \le t_2 \le \ldots$ and any bounded function f in R^m:

$$E_i[f(Y_{T_n+t_1}, Y_{T_n+t_2}, \ldots, Y_{T_n+t_m})|Y_s : s \le T_n] = E_j[f(Y_{t_1}, Y_{t_2}, \ldots, Y_{t_m})] \text{ given that } \{X_n = j\}$$

The semi-Markovian process introduced earlier can be defined in an alternative way [20] to be semi-regenerative and using this fact we will obtain the sample paths for the transition probabilities. Using the usual renewal argument we condition on the time of the first renewal T_1:

$$P_{ij}(t) = Pr[Y_t = j|X_0 = i, T_1 \le t] \cdot Pr(T_1 \le t) + Pr[Y_t = j|X_0 = i, T_1 > t] \cdot Pr(T_1 > t) \quad (3\text{-}17)$$

The second term in the above equation can be written as:

$$Pr[Y_t = j|X_0 = i, T_1 > t] = I(i, j)\Big(1 - \sum_{k \in E} Q_{jk}(t)\Big) \quad (3\text{-}18)$$

while to compute the first term we shall use the notion of semi-regeneration. Since the process Y is semi-regenerative it satisfies the above conditions. We apply condition c) by choosing for function f the indicator function which is bounded in R: $f(Y_t) = 1_j(Y_t)$

Statement c) above for any time $t \ge T_1$ after the first transition becomes:

$$Pr[Y_t = j|X_0 = i, T_1 = s] = E_i[1_j(Y_t)\ |Y_s : s \le T_1] = E_k[1_j(Y_{t-s})|] = Pr[Y_{t-s} = j|X_0 = k, T_1 = s] \quad (3\text{-}19)$$

Plugging (3-17) and (3-18) into equation (3-16) we obtain:

$$P_{ij}(t) = I(i, j)\Big(1 - \sum_{k \in E} Q_{jk}(t)\Big) + \sum_{k \in E_0} \int^t dQ_{ik}(s) P_{kj}(t - s) \quad \forall\, i, k \in E, t \in R_+$$

But the above equation is just the Markov renewal equation (3-14) satisfied by the transition probabilities. The solution to this equation was given in (3-15) and thus we have:

$$P_{ij}(t) = \int_0^t dR_{ij}(s) \cdot I(i, j)\Big(1 - \sum_{k \in E} Q_{jk}(t)\Big) \quad (3\text{-}20)$$

3.6 Semi-Markov Processes with an Absorbing State

An absorbing or trapping state in a stochastic process is a state which once the process has entered it never leaves. Hence transitions to another state are impossible. Transition from a state to itself is possible; however this happens after an infinite time. For any such state, call it *j* equations (3-3) take the form: $F_{ij}(t) = 0$ for all finite times $t \geq 0$ and $F_{ij}(\infty) = 1$. Consequently the one step transition functions from equation (3-3) are zero for all finite times $t \geq 0$ and hence the Markov renewal kernel will have zeroes in the line corresponding to the trapping state:

$$\underset{\sim}{Q}(t) = \begin{bmatrix} \times & \times & \times \ldots & \times \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 \ldots & 0 \\ \times & \times & \times \ldots & \times \\ \times & \times & \times \ldots & \times \end{bmatrix} \leftarrow j^{th} - line \quad (3\text{-}21)$$

Looking at equation (3-20) the sum $\sum_{k \in E} Q_{jk}(t)$ will equal to zero when *j* is a trapping state and the integral in (3-20) yields just the $R_{ij}(t)$ that were introduced in equations (3-10) and (3-12)

One point to remark is that in the case of a semi-Markov process with transient and absorbing states the above analysis can still be used. In such a case the $R_{ij}(t)$ will yield the average number of times that the process starting from the transient state *i* has visited state *j* by time *t*, for every transient state $j \in E$ in the state space of the process, and will yield the probability that the process starting from the transient state *i* has been trapped in state *j* by time *t*, for every trapping state $j \in E$. So in essence once the process has visited the absorbing state it ends.

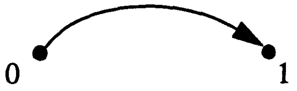
In the next section we give a simple example so as to make the ideas presented thus far more tangible and demonstrate their validity in modeling simple stochastic processes.

3.7 A Simple Application of Markov Renewal Theory

Let's consider the simplest stochastic process which models the lifetime of a machine component, such as light bulb. The process begins at the transient state (the machine component functions, the light bulb is on) and after some random period of time it fails. Assume for simplicity that the lifetime of such a component follows an exponential distribution with rate b and let's ask what is the probability that the light bulb has failed by time t and what is the average number of times that the bulb will be on?

Since the lifetime of the bulb follows an exponential distribution with rate b the probability that it has failed by time t is just $1 - e^{-bt}$ and the average number of times that the bulb will be on in any time interval *[0,t]* is just *1* since after it fails it is assumed not to function again (it is "thrown away").

The above stochastic process can be modeled as a semi-Markov process with two states only: One transient (the bulb functions) that we will call state *0* and one absorbing or trapping state (the bulb has stopped functioning) that we will call state *1*:



For this simple process matrices $\underset{\sim}{P}$ and $\underset{\sim}{Q}(t)$ are: $\underset{\sim}{P} = \begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix}$, $\underset{\sim}{Q}(t) = \begin{bmatrix} 0 & F_{01}(t) \\ 0 & F_{01}(t) \end{bmatrix}$ where $F_{01}(t) = 1 - e^{-bt}$, $F_{11}(t) = 0,\ \forall t \in [0, \infty)$ and $F_{11}(\infty) = 1$. We take the matrix derivative of $\underset{\sim}{Q}(t)$:

$\underset{\sim}{Q}'(t) = \begin{bmatrix} 0 & be^{-bt} \\ 0 & 0 \end{bmatrix}$ and then its Laplace transform which is: $\underset{\sim}{Q}'(a) = \begin{bmatrix} 0 & \frac{b}{b+a} \\ 0 & 0 \end{bmatrix}$ and we apply equation (3-13) to obtain the $R_{ij}(a)$ in the frequency domain:

$$\underset{\sim}{R}(a) = \underset{\sim}{I}(a)(\underset{\sim}{I} - \underset{\sim}{Q}'(a))^{-1} \Rightarrow \quad \underset{\sim}{R}(a) = \frac{1}{a} \cdot \begin{bmatrix} 0 & -\frac{b}{b+a} \\ 0 & 1 \end{bmatrix}^{-1} \quad \Rightarrow \quad \underset{\sim}{R}(a) = \frac{1}{a} \cdot \begin{bmatrix} 0 & \frac{b}{b+a} \\ 0 & 1 \end{bmatrix}$$

Taking the inverse Laplace transform we obtain:

$$\underset{\sim}{R}(t) = \begin{bmatrix} u(t) & 1-e^{-bt} \\ 0 & u(t) \end{bmatrix}$$

where *u(t)* is the step function at the origin.

Note that *0* is not a trapping state so $R_{00}(t)$ gives the average number of times that the process will visit state *0* by time *t* which of course is always equal to 1. However state 1 is a trapping state so the last column of the above matrix gives the probabilities that process by time *t* has stopped when starting from either the transient state *0* or the trapping state *1* which of course agrees with the results presented earlier. Alternatively one could have obtained the same results directly by using equations (3-12).

CHAPTER 4

EXPERIMENTS

4.1 Experiment Design

The experiment described here is focused on exploring human factors issues that have been deemed to be a critical component of safe rail operation. Rail vehicles cover large distances mostly through isolated areas and as a result train operators find it difficult to remain vigil, alert or even awake over prolonged periods of time. The attenuation in their attention has been attributed to *boredom* and *fatigue* as after of several hours of vehicle operation "nothing has happened". Several reports on accidents within the rail transportation community have identified engineer's *inattention* stemming from boredom and fatigue, as the main cause of signal violations and accidents. And the picture is projected to worsen as train speed increases since the available latency time for perception, intervention and actuation is reduced.

Automation offers one potential solution for improving the performance of the integrated human-machine system. However one needs to investigate the effects of control automation on the train operator's attention. Automation's side-effects such as the so-called "out of the loop problem" and the consequent over-reliance on automation might contribute to an operator's fatigue instead of reducing it. Humans tend to rely on automation and the concern is that a "bored" locomotive engineer might become even less attentive when the functionality of the vehicle is assumed by the automated system.

Given the above it was decided in the earlier stages of this research to concentrate on the underload case of the train operator by simulating a high speed train driving environment in which "nothing happens" for a "long period of time" hence inducing boredom to the human operator. Toward this end the number of emergency scenarios was kept to a minimum as discussed later on. The next step was to use this environment as a test-bed for exploring how different levels of automation affect human performance in terms of the vigilance decrement over time and situation awareness. Further issues such as the risk probability at high speeds were explored.

Vigilance decrement over time was evaluated using signal detection theory, which requires instances of signals to be presented over a noisy background to the observer who has to distinguish them. Accordingly emergency scenarios which represented failures of the system were introduced. The failures remained active only for a certain period of time and then, unless they had been noticed and reset by the subjects, they were automatically reset. The amount of time to complete a round trip (45 minutes) was broken into 3 time blocks and the vigilance decrement was evaluated in each one of them. When conducting experiments in a simulated environment that replicates an actual setting it is difficult to estimate the percentage of false alarms. Hence the percentage of failures reset by the subjects (detection rate) in each time block was used as a proxy for the vigilance decrement [12].

Situation awareness was decided to be conducted using the query technique for the reasons outlined in Chapter 2. The queries chosen to be asked were related to the overall state of the rail system ranging from speed and position to display indications. Provision was taken to familiarize the subjects with simulation suspensions during the training sessions of the experiment.

Emergencies were presented within the cab (instrument panel failures) and out the window (obstructions) to ensure that subjects' attention wouldn't be biased towards monitoring a particular display. Finally a bonus/penalty scheme was used to ensure consistency of operator behavior with the operational rules of rail systems and the purposes of the experiment. This scheme is presented in the tutorial (appendix A).

4.2 Facilities - Interactive High Speed Train Simulator

4.2.1 General System Description

The High Speed Train Simulator is a real-time interactive and distributed system that has been developed for the Volpe National Transportation Systems Center for studies associated with human factors research in high speed rail transportation. Below is given a brief outline of the simulator. (A detailed description is out of the scope of this document and can be found elsewhere [1]). The system is comprised of the two SGI personal Iris Workstations, one SGI Indigo-2 workstation, a personal 486 computer, a projector with a projection screen, the cab and control lever. A rough layout is presented in the figure below:

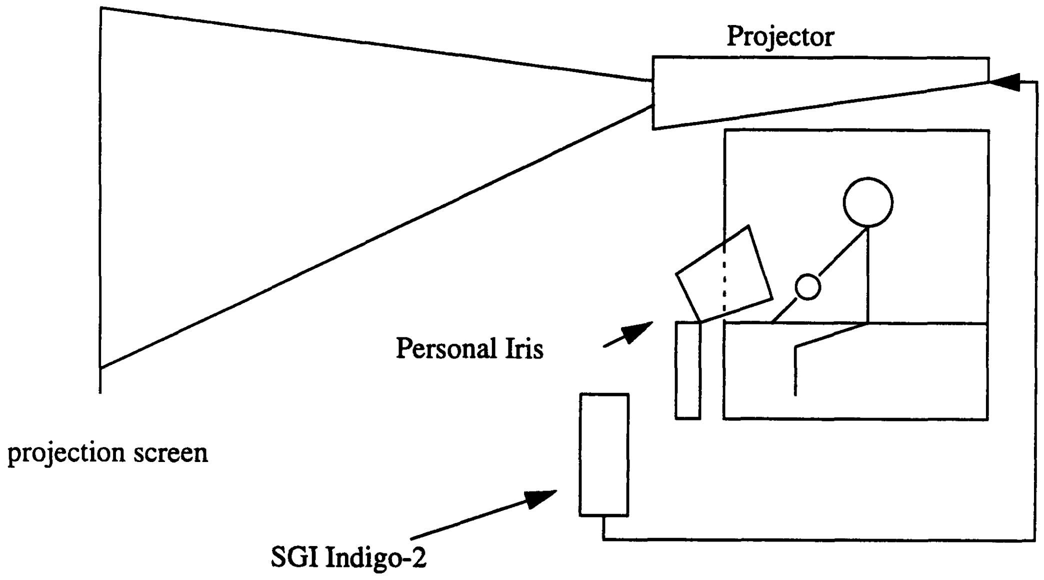


Figure 4-1: Simulator layout

The SGI Indigo-II workstation (with Extreme graphics) is used to generate high resolution graphics for the out-the-window view display and its output is fed into the projector which is used to project the image on the projection screen. Vehicle dynamics computations are performed in the same workstation as well. One personal Iris is responsible for displaying the instrument panel (figure A-3) while the other one is used as the Central Traffic Control workstation. All three machines communicate with each other over a Local Area Network (LAN).

The throttle represents an integral part of the Simulator and is used for controlling the speed of the train. It is connected to the serial port of the Indigo-II workstation via an Analog to Digital (A/D) converter. This lever is used to control both the thrust and braking commands, with the forward (up) direction for thrust and the backward (down) direction for braking. The center position (coast) is notched for reference and there neither thrust nor brakes are applied.

The simulator exhibits three control modes, namely full manual control, cruise control, and autopilot control. Under manual control the vehicle operator is solely responsible for controlling the speed and the position of the vehicle using the combined control lever. The cruise control system is designed to maintain a constant speed which is set by the operator. The functionality is generally similar to the cruise control systems found in automobiles. The autopilot system is able to fully control the speed of the vehicle without human intervention. In such a case the human's role is shifted to that of the supervisor who monitors the system for proper functioning. This means that the operator must be prepared to take over control should the situation warrant.

The rail system used in the experiment is comprised of two stations, named the East and the West station (figure A-2), separated by 50km of single track. At the ends of each station loops are used to reverse the vehicle in the main track. The one way travel time from station to station depends on the signals set and is approximately 19 minutes if no signals are set while the travel time around the loops takes a bit more than 4 minutes. Hence the round trip lasts roughly 43 minutes. To this time one must add the time the vehicle remains in the station which was about 30 seconds.

Rail systems have traditionally used a system known as ***block signaling*** for the control of trains in the rail system. With block signaling, the track is divided into chunks known as blocks. While the length of each block is fixed (does not change over time), different blocks are not necessarily of equal length. Typically, shorter block lengths are used near the vicinity of stations, while longer block lengths are used in regions away from the stations. Block lengths are generally of the order of one mile. In the road system used in the simulation, all blocks between stations are 2 km, and all blocks in the loop sections are 1 km in length. At the boundaries of each block there is a ***signal light***. This signal light displays a color-coded signal (details can be found in the tutorial), which indicates the maximum speed permitted throughout the block. The signal acts as a ***dynamic speed limit***, and it is the responsibility of the vehicle operator to identify the signal as the block boundary is approached and set the vehicle speed accordingly. A fundamental rule in block signaling is that no more than one train can occupy a block at any given time.

A red signal is used to indicate that the block is currently occupied by another train, and the approaching train is not permitted to enter that block. The blocks that precede the occupied block have signal levels which ensure that the train can be slowed in time to stop before entering the occupied block.

In addition to the speed limits imposed by the block signal system, there are also ***civil speed limits***, which are static. These are dictated in urban areas due to the increased likelihood of having accidents. In the simulation the blocks that have civil speed limits have grade crossings as well. In all cases, the prevailing speed limit is the ***lesser*** of the block signal limit and the civil speed limit. The exact specification of signals used and speed limits associated with those signals is a design parameter for a rail system, and varies from system to system. In the simulation system, a five-aspect signaling system is used. This means that there are five color codes used in the system, with the codes defined as shown in table A.1.

4.2.2 Simulator Upgrades

For the purposes of the present research several upgrades of the train simulator had to be performed. These included:

1. The addition of suspend/resume capabilities in the train simulator for the purpose of gathering situation awareness data. Any time the experimenter - acting as the CTC operator - can suspend the simulation and ask questions relevant to the state of the vehicle. There have been designed and implemented two versions: in one, which is the default and used in the experiments, the out of the window view is blanked and the dashboard is blanked except for the communications window which still remains active so that the subject can respond. In the other version both screens are just frozen but not blanked. One can invoke the later version by using the `-noblank` flag in the command line argument when running the train simulation. In any case provision has been taken for the simulator to record everything pertaining to the state of the train as well as the subject's response. Additional provision has been taken to reset the simulation clock for purposes discussed below.

2. Options for the experimenter to activate/disactivate the ATP (Automatic Train Protection) system from the CTC screen. This option can be invoked using the `-atp` flag when running the CTC simulation.

The network programming for both of the above options was done using the BSD socket interface and the communications protocol used was the UDP (User Datagram Protocol) which is the protocol used originally to build the simulator and is layered above the Internet Protocol (IP). For the particular application the UDP protocol though connectionless has advantages over the connection oriented TCP, since in case of a network failure the data packet sent from one machine to another is guaranteed to be delivered when using the TCP protocol. This means that the packet will be retransmitted until it reaches its destination, which might result in a delay of several seconds. As a result the data received and displayed by the destination machine will not correspond to the real time data. In contrast the UDP has no mechanism of checking whether the data sent are lost, making it ideal for a real time simulation system: by keeping the rate at which data are sent to the destination machine high a lost data packet will not even be perceived by the simulation since another one - with the accurate simulation data - will be sent shortly.

3. Development of a new version for the autopilot. In the previous version of the simulator the autopilot once invoked would control the speed of the train adhering only to civil speed limits. Hence in the case that a signal was present the autopilot would fail to recognize it and the vehicle operator needed to assume manual control to adjust the vehicle speed. In the current version - under which the present research was carried out - the autopilot would recognize civil and signal speed limits simultaneously and adhere to the lesser of the two, thus replicating existing train autopilots such the one used in the German ICE. Hence under normal conditions the operator need not interfere with speed control of the train but only monitor that the autopilot adheres to the speed limits of each block. This version of the autopilot can be realized using the `-full_auto` flag in the command line argument when running the train simulation.

4. The addition of bearing failure. In this scenario the temperature of a faulty bearing rises as a result of increased friction in its rollers. The rate at which the temperature goes up under failure conditions is higher than it is under normal conditions.

5. The addition of failure reset mode for the purpose of gathering vigilance data. In older versions of the simulator once the failures are set they remain active until the subject notices them. In the current version one can have the failures reset according to the time that has elapsed or the distance that the train has traveled since the onset of the failure, in the event that the subject has not been able to spot and reset them in time. To use this mode one must specify the `-fail_reset` option in the command line argument of the main simulation and in the failure input file has to enter the keywords `time` followed by the time in milliseconds one wants the failure to remain active, or `distance` followed by the distance in meters that

the vehicle will travel before the failure is reset. Needless to add that the previous mode (whereby the failures are set and remain active until reset form the subject) remains yet an additional option.

As stated above, provision has been taken to reset the simulation clock. Hence in case one has suspended the simulation while a failure is active, the failure will remain active for the residual amount of time when the simulation is resumed.

4.3 Experiment Description

4.3.1 Subject's Task

The subjects were required to operate the High Speed Train Simulator according to the operational rules and operating requirements of such a system as outlined in the training tutorial (appendix A) that they were given to review. Since speed and position control of the vehicle are the most important aspects of train driving the task was to control the vehicle speed according to the signal and civil speed limits present in each block, monitor the system for potential hazards posed by failures and respond to them in a proper manner should they be present.

To explore the effects of different levels of automation the experiment consisted of driving the simulated train from West Station to East Station (figure A-2) and back (one shift) using one particular control mode. For each subject there were a total of three runs, lasting about 45 minutes each and completed using a different control mode namely manual, cruise control, and autopilot (or fully automatic control). Between successive runs periods of rest longer than 10 minutes were mandatory so that the subjects could refresh themselves before proceeding in the next trial. Under manual control mode the subjects used the lever to control the vehicle speed by applying thrust or braking. When driving the train under cruise control mode the subjects were responsible for setting the desired speed according to the speed limits of each block and let the system apply the proper amount of thrust to achieve and maintain the set speed. Cruise control would not apply brakes, so the subjects had to apply the brakes should the situation warrant this as dictated by either emergencies (obstructions, failures) or the presence of either signal or civil speed limits. Under autopilot mode speed control of the vehicle was exclusively assumed by the automated system which was responsible for setting the speed in accordance to the speed limits of each block. In this case the subject's task was to monitor the proper functioning of the automation.

4.3.2 Emergency Scenarios

For the purposes of assessing performance the following failure scenarios were introduced, each used for a different purpose:

1. Motor failure. Under normal operating conditions when thrust was applied (either manually by using the control lever or by the system) the ammeters, which show the amount of current that goes through each motor, would be lit on the instrument panel (figure A-3). A motor failure resulted in an absence of current flow through one of the motors which was indicated on the motor ammeters. In this case the correct response was to a) first remove power from all the motors by pulling the control lever back to braking position b) press the appropriate key for resetting the failure and c) resume operation of the vehicle using the previous control mode. The failure remained active for 12 seconds and unless reset by the subject was automatically reset after that amount of time.

2. Bearing failure. The bearing temperature would normally follow the speed of the train with some time lag and never rise above 55 ^{o}C for safety reasons. So in essence the bearing temperature fluctuated around a mean value, depending on the average speed of the vehicle during the course of the trip, with a maximum rate of 1.53 mHz. In a failed bearing the temperature rose with a rate almost double the maximum rate regardless of the vehicle speed. The subjects had to notice this change in the rate and press a key to reset it. The failure remained active for 20 secs before it would reset automatically.

The failures presented above were used to assess the vigilance decrement. In essence normal train driving conditions provided the “noisy background” and the above two types of failures were the “signals” to be detected by the “observer” (train operator). The time that the failures remained active was chosen to represent the average reaction time of the population and was inferred by experiments [1] conducted with the same type of failures. Bonus was given for spotting and resetting the failures whereas penalties were applied for hitting the reset keys when there was no failure present. The scheme is presented in the tutorial in appendix A.

3. Brake failure. The brake pressure gauges indicated the pressure of the compressed air in the brake tanks. Their indication ranged from 98 psi when no brakes were applied to 22 psi when full service brakes were applied and would fall further to 0 psi under emergency braking. When a failure was present one of the brake tanks would loose pressure. The pressure loss was indicated in the corresponding brake tank

gauge. Once the failure was detected the correct response was to switch to an alternate compressor by depressing the appropriate key on the computer keyboard. This failure differentiates from the previous ones in that a certain distance of 800 meters was given to the subject to reset the failure. Now the reaction time depends on the speed of the train. Data gathered from this type of failure were used by the Markov renewal theory to provide the time paths of the risk probability in the high versus the low speed regime. The above distance was again chosen as the average distance traveled before a brake failure was reset as indicated by experiments conducted with the same type of failures.

Additional perils to the safe operation of the system were posed by obstruction hazards. These were present at points where highway roads and rail tracks intersect. In total, there were five grade crossings. At each grade crossing, highway vehicles (cars) could cross in front of the train from either direction. These vehicles were visible at over a half kilometer distance. As stated each of the blocks containing grade crossings, had a civil speed limit of 100 km/hr. This means that the maximum speed of a vehicle in the block was 100 km/hr unless a lower speed limit was imposed by a signal in that block. Traffic at the grade crossings arrived according to a probabilistic process. A car would proceed across the crossing only if there was sufficient distance to clear the crossing before the train. (In other words, a car would not proceed if there was not enough "room"). However, it was possible for a car to become disabled as it was crossing the tracks, which would result in an obstruction for the train. In this event, the train operator should bring the train to a complete stop before the intersection. If the train was not stopped in time, a collision would occur, which would be indicated by a cracked windscreen. It was important for the train operator to be able to quickly assess the crossing traffic and determine whether the train should be stopped. On one hand, a collision was a major event, and would result in a significant delay in operation. On the other hand, stopping the train unnecessarily would also cause delays in service. It was up to the vehicle operator to evaluate the situation and determine the best course of action under these constraints.

This type of emergency was used to assess the effects of control automation when the operator had to react under stress and to estimate the feel of control the subjects had over the vehicle.

4.3.3 Experiment Control

In the course of the experiments for the purpose of comparing the results and experimental control the failures were set to occur at fixed positions along the track according to the distance that the vehicle had traveled. However from the perspective of the subjects the failures were set randomly. The average number of failures (motor, bearing) for the whole round trip was only ***two***, since our interest was to simulate a train driving environment that induced boredom and fatigue, and this number was given to the subjects during the training sessions. Additionally, the motor and bearing failures where spread across the different levels of automation and time blocks evenly (counterbalancing principle). Thus, for instance, each time block would have on average 5 motor failures and 4 bearing failures or the other way around, depending on the control mode and time block. Detailed positions of where the failures were set to occur are given in appendix D while the emergencies each subject experienced along with the subject's performance are given in appendix E. The counterbalancing design of Table 4.1 was used both in the training sessions and actual trials using the "shift" principle. According to this, subject 1 would be introduced to the control modes according to line 1 of that table but would experience them according to line 2, subject 2 would be introduced to the control modes according to line 2 of that table but would experience them according to line 3 and so on for the rest of the subjects.

4.3.4 Performance Incentives

To ensure that the operator performed according to the objectives of the experiments a bonus system was used which provided monetary rewards for good performance. However penalties where also assessed if operator's performance didn't fall within certain criteria. The precise specification of the bonus/penalty used in the experiment can be found in the Training Tutorial in appendix A. Bonus or penalty points where converted to monetary rewards at a rate of $1 per 1000 points and the subjects were paid after the experiment.

4.4 Training Procedures

4.4.1 Instruction Material

Due to the complexity of the system and to familiarize the subjects with rail operation a written tutorial was prepared. Its backbone as far as rail operation and system description is concerned was the original version developed by Dr. Edward Lanzilotta modified accordingly to highlight and address issues pertinent to the current research. It included the general operational rules that govern rail operation, operating requirements and control modes of the train simulator, and it provided details on the experiment.

The use of the tutorial served many purposes: It provided with an adequate background all of the candidates and gave them ample time to assimilate the material. Hence they had the opportunity to get a feel of what the purpose of the experiment was and could have any questions answered before they had hands on exposure to the system. On the other hand it acted as a filter to gauge interest of the participants in the experiment since it required a substantial commitment of time to read and comprehend. In this sense review of the training tutorial was the first step of their participation is the experiment. The training tutorial is included in appendix A.

4.4.2 Training Sessions

Once the subjects had read the training tutorial they had to take a written quiz. This was given to ensure that all of them had read the tutorial before coming in the training sessions and to identify potential problems on the understanding of the material that needed to be clarified during the training sessions that followed. The quiz consisted of 17 multiple choice questions, required approximately 10 minutes to complete, and was graded upon completion. There was a minimum performance requirement of 50% on the quiz to continue on the next phase which was the training session. The quiz is included in appendix B.

The training session lasted three hours and gave hands on exposure to the system. During the first hour the experimenter demonstrated the proper vehicle operation and introduced the subjects to the emergency scenarios so that they would get familiar with them and learn the proper response to such situations. The remaining two hours were devoted to preliminary rides both with and without the emergency scenarios that would further familiarize them with the system and its operating modes. During this period the subjects completed a run from station to station using one particular control mode. To counterbalance possible learning effects across subjects each subject was introduced to the control modes according to the counterbalancing design presented in Table 4.1. During these runs the experimenter was physically present in the cab monitoring the way the subjects controlled the vehicle and making recommendations related to the proper actions under each control mode. Performance was monitored and suggestions regarding speed compliance were given. These runs lasted about one hour and a half. During the final half hour of the training session another run from station to station was given in which the subjects had the freedom to experiment with the simulator in any way they liked and ask questions that they might have. During this run the experimenter shared his time between the cab and the CTC room to allow more freedom on vehicle operation by the subjects. Additionally the simulation was suspended three to four times during the last hour of training to familiarize the subjects with the freezing technique and queries related to the system were asked.

After the training session was over the experimenter would give a short break to the subjects and ask for their opinion regarding the level of control they felt they had over the system. At this point the subject could pick to have additional run(s) that would either expose them to more emergency scenarios or would replicate an actual trial with a very few failures and so on. Never did the experimenter proceed in the actual trial unless he had the "green light" from the subjects.

Subject Number	Presentation order of control modes
1	Autopilot
	Manual
	Cruise
2	Autopilot
	Cruise
	Manual
3	Manual
	Cruise
	Autopilot
4	Manual
	Autopilot
	Cruise
5	Cruise
	Autopilot
	Manual
6	Cruise
	Manual
	Autopilot

Table 4.1: Counterbalancing design of the mode presentation ordering

4.5 Subjects

A total of 15 subjects participated in the experiment. No specific criteria were used to filter out the subjects. A consent form was required to be signed prior to their participation in the experiment and is included in appendix C. The majority (12) where undergraduate or graduate students at the Massachusetts Institute of Technology. Two students where used in the initial phase for fine tuning the experiment while the remaining 13 ran in the actual trials. The data from one subject (non-MIT) had to be thrown since upon inspection data showed that he had repeatedly failed to comply with speed limits. No particular problems were noticed in the behavior of the remaining 12 the subjects.

CHAPTER 5

RESULTS AND DISCUSSION

5.1 Human Behavior Results and Discussion

For the vigilance study, time -particularly time block- was the *independent variable* and the vigilance decrement was the *dependent variable* with *parameter* being the level of automation. Detailed presentation on the performance of each subject is given in appendix E. The results from the three practice blocks across the different levels of automation are summarized in tables 5.1, 5.2, and 5.3. Figure 5-1 shows the percentage of correctly detected signals in each time block for each different control mode.

Evidence of vigilance decrement is present for all three modes. However the detection performance under autopilot mode was consistently higher than for the rest of the control modes in all three time blocks. Additionally the detection rate showed less decline over time as evidenced from figure 5-1. From the two remaining control modes, namely the manual control and cruise control, vigilance performance was worst for the cruise control mode but relatively stable over time in contrast with the manual control mode. The latter had a high detection rate in the initial stages of the task (first time block) but declined as time elapsed. The vigilance results regarding human efficiency of automation monitoring are in accordance with the ones found in the literature [12]: When humans have to monitor the automation and report 'failures' (as was the case in the experiment) performance is equally good under either manual or automatic control. The problems with automation arise when humans are involved in multiple tasks. Such were the cases under cruise and manual control in which the subjects had to continuously control the speed of the vehicle in addition to system monitoring. And as a result vigilance performance declined over time. Below we will see how interesting the vigilance results are in regard to situation awareness and subject comments.

It is worth mentioning that despite the different discrimination category for the two signals (the motor failures required simultaneous discrimination as opposed to the bearing failures which required successive discrimination) subjects detected equally well both signals in all time blocks. This is attributed to the very low event rate which didn't impose memory load on the subjects and on the extensive and thorough training the subjects went through before the trials [10] as evidenced by their subjective ratings in figure 5-6.

During the trials the simulation was suspended at random times, and the subjects were asked to report their speed, the block number and kilometer post where they were, the brake pressure, the bearing temperature and the current through the motors. Results from the brake pressure reports were not significant since the brake pressure was fairly stable over time. To analyze the rest of the data the following procedure was used. For the speed and bearing temperature responses, the mean of the absolute error (*MAE*) was computed. The subject response was subtracted from the actual value and the absolute value of the result was taken. The average was taken across all subjects: $MAE = \sum_{i=1}^{n} \frac{|x-\hat{x}|}{n}$ where x was the subject response, $\hat{x}$ the actual value and n the number of subjects. For the motor current the block number and the kilometer post the answer was classified either as correct or wrong since a) the deviation from the actual value was of the order of one or two and b) the response could be given with the required accuracy. The *independent* variable was the system state and the *dependent variable* either the *MAE* or the *percentage of correctly answered questions,* depending on the query category. *Parameter* was the level of automation in use. The results gathered using the "freeze technique" for assessing situation awareness for the different control modes are presented in figures 5-2. Boxplots are used for the speed and bearing temperature results and display the mean, the standard deviation and the range of the data, whereas for the motor current and the position data the percentage of correct responses is given.

Additionally during each trial the subject was presented with one obstruction. This type of emergency was used to assess the impact of control automation when the operator had to react under stress and to estimate the feel of control the subjects had over the vehicle. The *independent variable* used here was the control mode and the *dependent variable* the number of collisions and the distance stopped before the obstruction if the collision was avoided. The distance that the vehicle stopped before the obstruction is shown in figure 5-3.

After the experiment was concluded the subjects were asked the following questions:

1. Which control mode (manual, cruise, autopilot) imposed less workload on you during the each trial, in the sense of having ample time to process information regarding the system (such as speed, block number and kilometer post where you were, brake pressure, bearing temperature, etc.)?

2. Which control mode (manual, cruise, autopilot) imposed greater stress when you were about to react to an emergency such as an obstruction?

3. Which control mode did you prefer?

4. Comment on level of your training using a scale from 1 to 5 with 1: worse and 5: best. You may use one decimal digit.

Responses to the first and second questions are presented in figure 5-4, while the response to the third and fourth questions are presented in figures 5-5, 5-6 respectively.

The results of the situation awareness study are quite interesting. The autopilot mode presented the higher level of situation awareness for most categories but not for vehicle speed, which is the most important parameter in train driving. In fact autopilot had the worst performance in this category. To explain how this occured I'll quote exactly from what one subject said when asked to respond to question 1 above: "Auto pilot imposed the least workload since I could keep an eye on what was happening, yet I didn't have to worry (as much) about watching the speed - i.e. I had to check to make sure that autopilot was following the speed limits, but I didn't actually have to adjust the speed myself, so I could spend my time systematically checking all the information". Very similar were the comments of the majority of the subjects. The other interesting result is that the cruise control mode had the best speed awareness (lower *MAE*) from the rest two modes. Under this control mode subjects had to set the speed using the keyboard keys or the control lever for braking. However this hybrid control mode of the train wasn't preferred by any subject as indicated in figure 5-5. They complained that they didn't have a good feel of speed control when setting the speed and accordingly they had to pay more attention (compared to the rest of the control modes) to the speedometer to make sure that the automation was achieving the set speed. This is probably the reason that the detection rate was low from the beginning of the trials (figure 5-1) but had the best level of situation awareness of speed (figure 5-2A).

As far as the rest of the categories are concerned, autopilot had overall higher level of situation awareness. For the bearing temperature that might seem a contradiction according to figure 5-2B. However there is only one data point (the one in the upper range) that increases the variance of the *MAE*. If that point is considered 'noisy' and thrown away then the autopilot mode has both lower mean and variance compared to cruise control and of course the manual control mode. Manual control had generally worse situation awareness (but best speed awareness from the autopilot) according to the data. However it is interesting to see that this was the control mode that gave greater level of control and confidence over the vehicle. At a first glance this statement seems contradictory since one can see that this control mode had higher *MAE* in the speed category than cruise control. However if one looks at figure 5-3, one can see the reasoning of the argument. Under manual control the subjects had to regulate the speed using only the combined control lever. The angular position gave the subjects a cue as to the level of thrust or braking that they were applying.

Figure 5-3 shows the distance from the obstruction that the vehicle was bought to a complete stop under different control modes. It doesn't include data from five subjects, two who collided, one using cruise control and the other using autopilot, two that didn't get an obstruction and one who avoided collision without using emergency brakes. The results indicate that all subjects responded properly. Once they had spotted the obstruction they applied full service brakes and then they used their judgement as to whether and when to apply emergency brakes. No subject applied emergency brakes immediately once the obstruction was spotted. However every subject (except for the one mentioned above) under every control mode resort to the use the emergency brakes to avoid collision, and hence the results are comparable. Response time data to obstructions were collected. The analysis showed no significant difference in the variance of the response time data in accordance to [1]. As can be seen the average distance that the vehicle was stopped prior to the obstruction was greater under manual control, with cruise control and autopilot following. However the difference between the cruise control and autopilot modes was not statistically significant in contrast with the differences to the manual control modes which are statistically significant at the 3% level.

It is important to see how the results presented above correlate with how subjects responded to questions 1 and 2. The results are shown in graphs 5-4. The conclusion drawn is that while the autopilot mode was the most convenient to use no subject felt comfortable with either cruise or autopilot control modes when he or she was about to react to an obstruction. Again I'll use the remarks of one of the subjects which are very illustrative and representative of what the subjects answered in question 2:
"I believe that the cruise and autopilot controls imposed the greatest amount of stress when an obstruction was pending because after setting the speed, I would remove my hands from the control and begin to focus on other potential problems and information that I had to know. Thus, when an obstruction occurred, I had to jump from passively driving to actively driving and grab the controls in order to slow down".

This feel of control over the vehicle along with the ‘actively driving’ feel (the so-called in-the-loop-feel) was cited as the reason for ranking the manual control mode first in their preferences by the subjects that did so (figure 5-5) while the subjects who preferred the autopilot did so because it imposed less workload on them. Finally, as shown in figure 5-6, training was more than adequate in the subjects’ opinion.

	Time block 1	Time block 2	Time block 3
Failures	4 motor, 5 bearing	4 motor, 5 bearing	5 motor, 4 bearing
Hits	8	7	7
Misses	1(motor)	2 (motor, bearing)	2 (motor, bearing)
False Alarms	3	3	0
Detection rate (%)	89	78	78

Table 5.1: Performance under Autopilot control mode

	Time block 1	Time block 2	Time block 3
Failures	6 motor, 3 bearing	5 motor, 4 bearing	5 motor, 5 bearing
Hits	6	6	6
Misses	3 (1 bearing, 2 motor)	3 (2 bearing, 1motor)	4 (2 motor, 2 bearing)
False Alarms	2	1	0
Detection rate (%)	67	67	60

Table 5.2: Performance under Cruise control mode

	Time block 1	Time block 2	Time block 3
Failures	4 motor, 5 bearing	5 motor, 5 bearing	5 motor, 4 bearing
Hits	8	7	6
Misses	1 (motor)	3 (1 bearing, 2motor)	3 (1 motor, 2 bearing)
False Alarms	0	2	0
Detection rate (%)	89	70	67

Table 5.3: Performance under Manual control mode

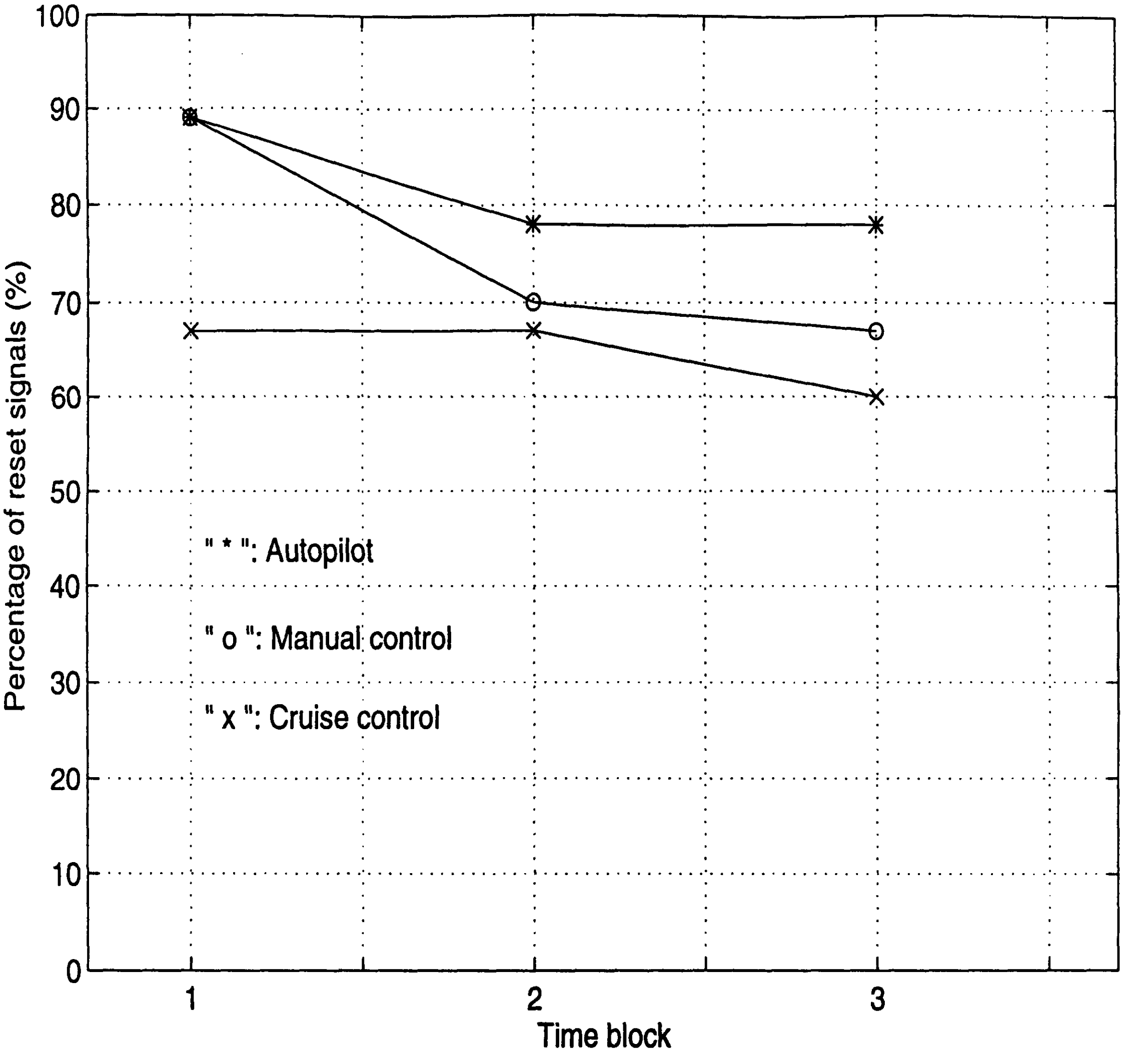


Figure 5-1: Effects of control automation on vigilance performance

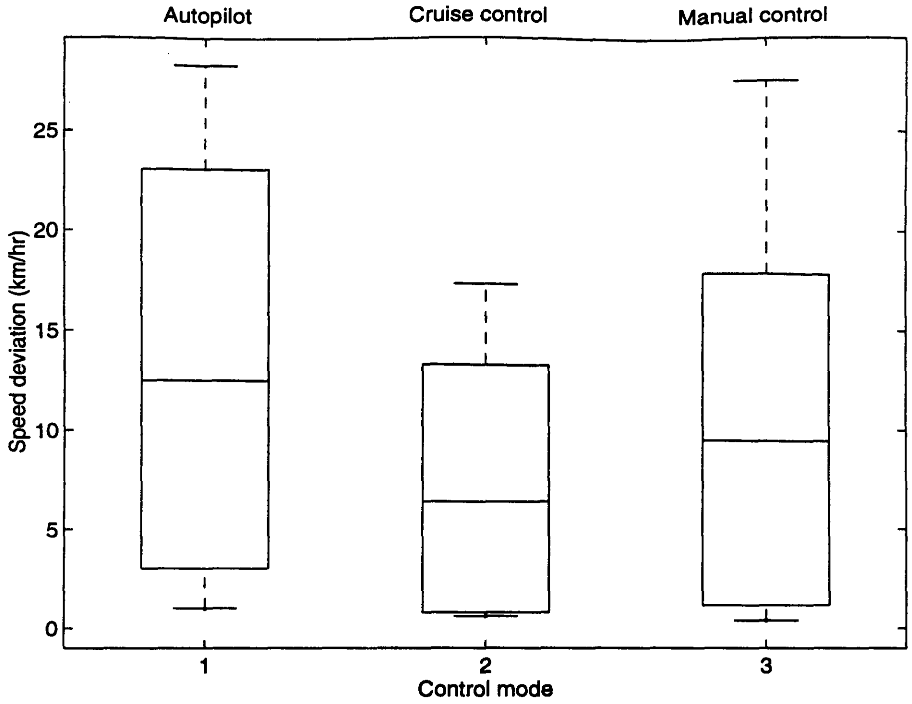


Figure 5-2A: Effects of control automation on speed awareness

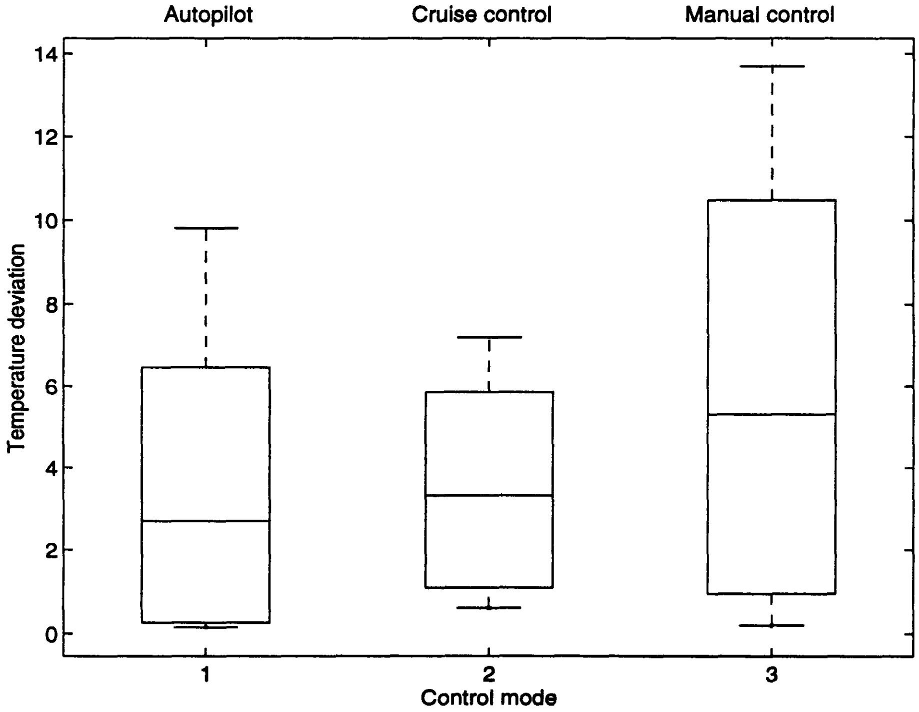


Figure 5-2B: Effects of control automation on bearing temperature awareness

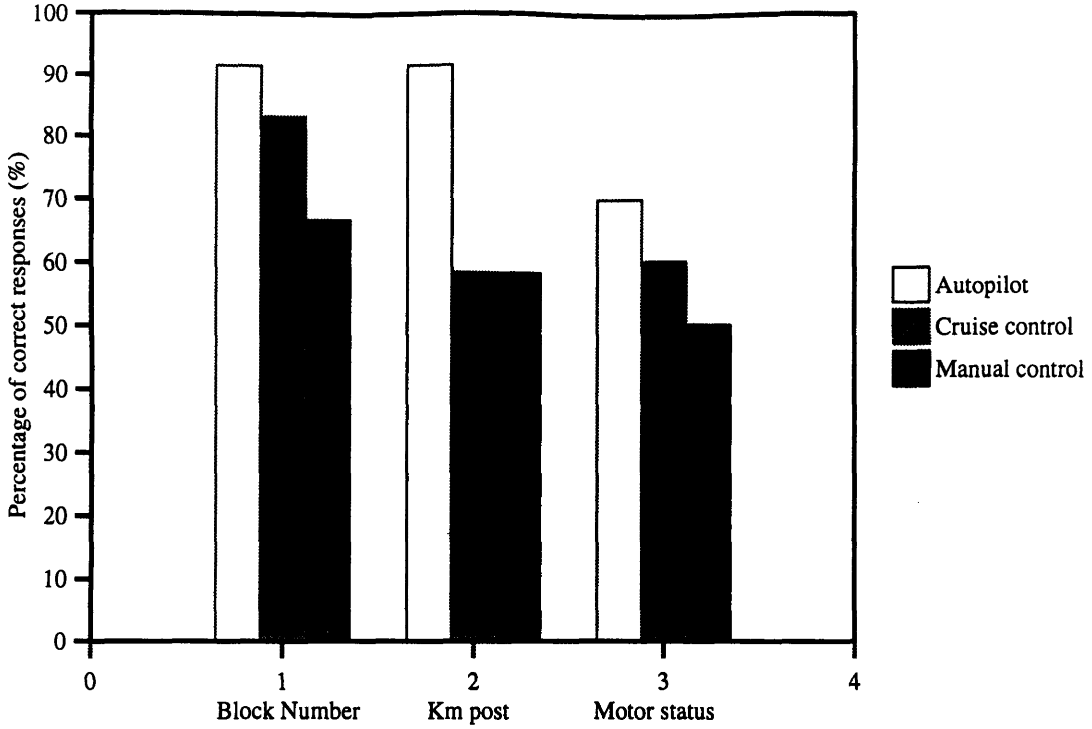


Figure 5-2C: Effects of control automation on position and motor awareness

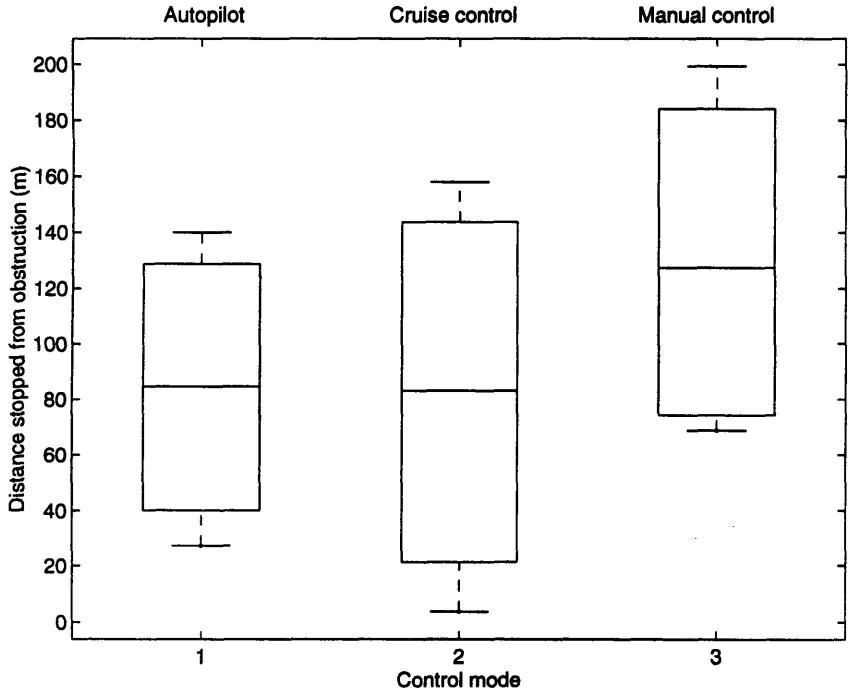


Figure 5-3: Distance stopped prior to a vehicle obstruction

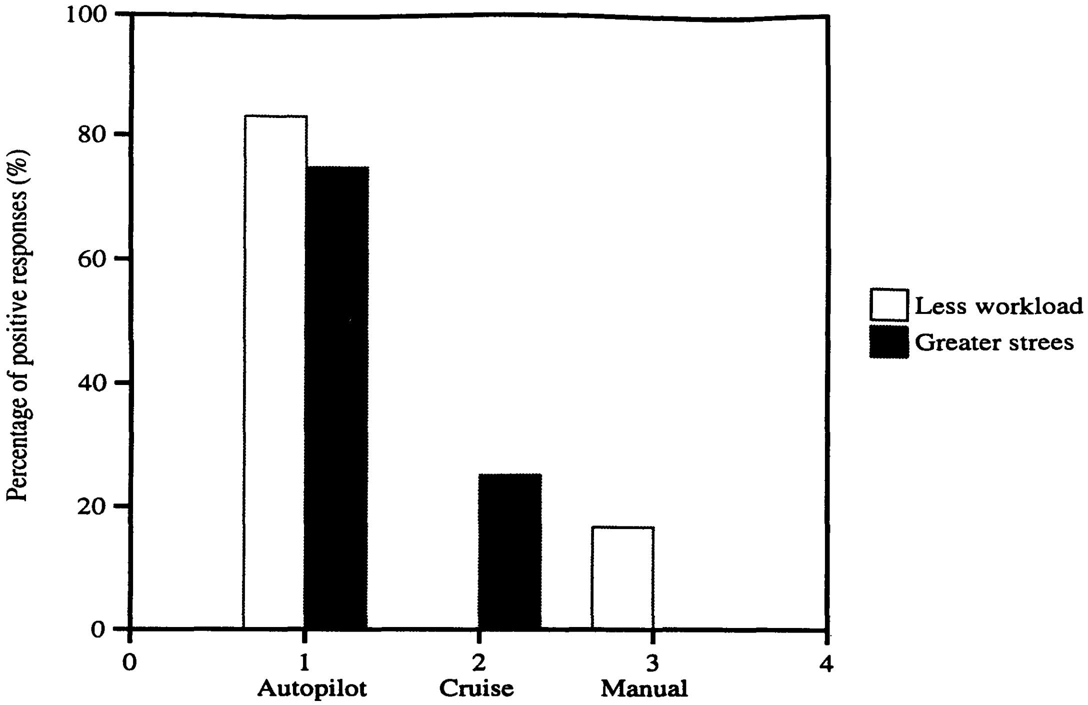


Figure 5-4: Subject rating of control automation with regard to workload - situation awareness and stress.

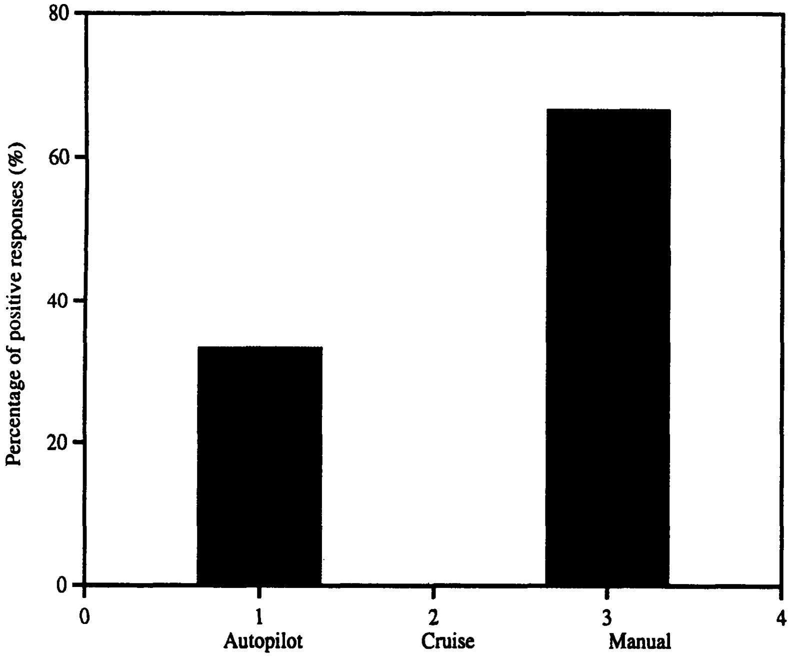


Figure 5-5: Control mode preference by the subjects

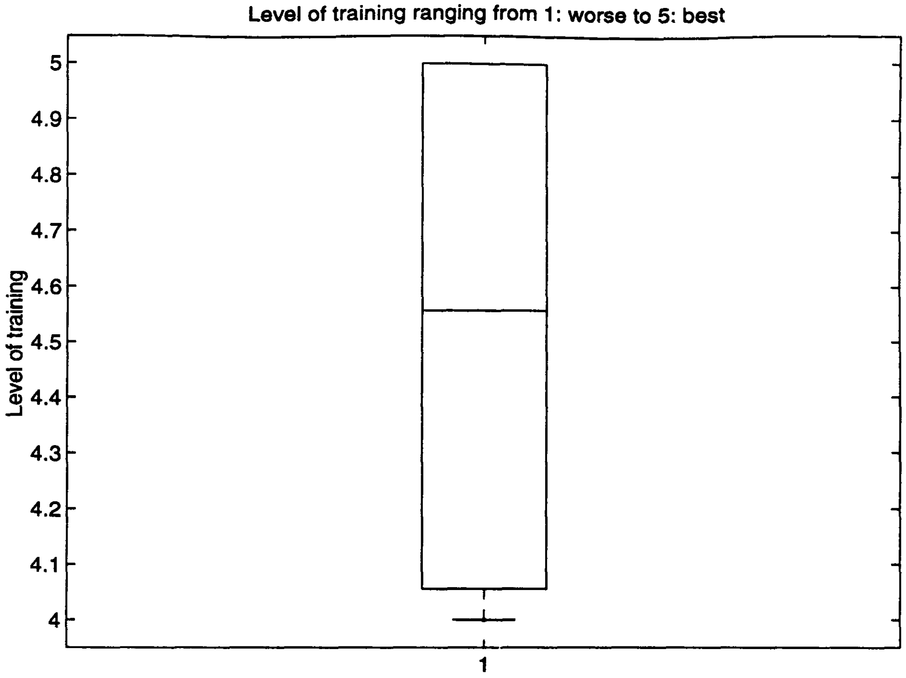
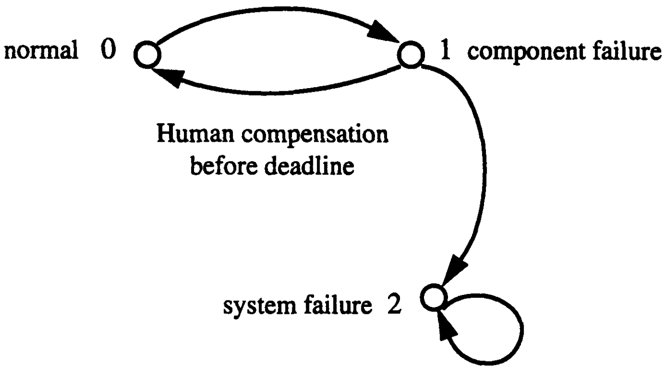


Figure 5-6: Ratings on the level of training

5.2 Risk Probability Results

5.2.1 Model Presentation

To explore the effects of high speed in the ability of the human to react before a 'deadline' a simple model was developed. The human-machine system is considered to be in one of three possible states: State 0 is the operational state. Component failures are assumed to "arrive" according to a Poisson process. In such an event the system moves into state 1 which represents the state of the system in which a component failure is present and requires human compensatory reaction before the 'deadline'. If the human operator responds properly to it, then the system moves back to the operational state 0. Otherwise moves to state 2 in which the system is considered inoperational and is modeled as a trapping state. The following diagram depicts the idea behind the model:



Matrix *P* which defines the transitions probabilities of the underlying Markov chain will be:

$$\underset{\sim}{P} = \begin{bmatrix} 0 & 1 & 0 \\ P_{10} & 0 & P_{12} \\ 0 & 0 & 1 \end{bmatrix}$$

and given the above functions for *F(t)* we can compute matrix $\underset{\sim}{Q}(t)$ as the elementwise product:

$$\underset{\sim}{Q}(t) = \begin{bmatrix} 0 & F_{01}(t) & 0 \\ P_{10}F_{10}(t) & 0 & P_{12}F_{12}(t) \\ 0 & 0 & 0 \end{bmatrix}$$

Assuming that matrix $\underset{\sim}{Q}(t)$ is known, we are interested in the time evolution of the probability that the system will be failing after an onset of a component failure. Namely we are interested to track $P_{12}(t)$ in high and the low speeds. This probability can be evaluated directly by applying equations (3-12) since state 2 is a trapping state:

$$R_{ij}(t) = \begin{cases} 1 + \sum_{k \in E} \int_0^t dQ_{ik}(s) \cdot R_{kj}(t-s), & i = j \\ \sum_{k \in E} \int_0^t dQ_{ik}(s) \cdot R_{kj}(t-s), & i \neq j \end{cases}$$

First note that $R_{22}(t) = 1 \ \forall t \geq 0$ from the first of the equations above since $Q_{2i}(s) = 0 \ \forall \, i$. Hence the above equations reduce to:

$$R_{12}(t) = \int_0^t dQ_{10}(s) \cdot R_{02}(t-s) + \int_0^t dQ_{12}(s) \cdot R_{22}(t-s) \text{ and}$$

$$R_{02}(t) = \int_0^t dQ_{01}(s) \cdot R_{12}(t-s)$$

Now taking the Laplace transform of the above equations we reach to a closed form solution for $R_{12}(t)$ in the frequency domain:

$$R_{12}(a) = \frac{Q'_{12}(a)}{1 - Q'_{10}(a) \cdot Q'_{01}(a)} \cdot R_{12}(a) \quad (5\text{-}1)$$

which in essence is the $P_{12}(t)$ since state 2 is a trapping state. The $Q'_{ij}(a)$ in the above equation are the Laplace transforms of the derivatives of the transition functions.

5.2.2 Model Application

The first step in applying the model is system calibration. The operational system was the high speed train. To study the effects of speed in the human's ability to respond to an emergency (component failure), a distance deadline was used. The human operator had to notice and react in less than 800 meters or the system would be considered inoperational (system failure). In the simulation brake (component) failures were assumed to arrive according to a Poisson probabilistic model at a rate of 1 failure per round trip (or 1 failure per 45 minutes). Hence the interarrival times had exponential distributions of the form:

$F_{01}(t) = 1 - e^{-bt}$, where $b = \frac{1}{45}min^{-1} = \frac{1}{2700}seconds^{-1} = 3.7 \cdot 10^{-4}$ was the rate of the Poisson process.

The information required by the $\underset{\sim}{Q}(t)$ were taken by system observation. The train speed range was partitioned into to regions: The "low" speed region with velocities from 0 to 150 km/hr and the "high" speed region from 150km /hr to 300km/hr. Eighteen failures where presented to the subjects 10 in the high speed and 8 in the low speed region. The probabilities P_{ij} where:

Low speed region: $P_{10} = 0.625$ and $P_{12} = 0.375$

High speed region: $P_{10} = 0.4$ and $P_{12} = 0.6$

while to obtain the transition functions (3-3) curve fitting was applied to the data given in the high and low speed regions. These data points are presented in figures 5-8, 5-9 for the low speed region and 5-10, 5-11 for the high speed region.

Exponential least squares fit was found to be pretty good:

$F_{10}(t) = 1 - e^{-ct}$, where c is the reset rate of the brake failure prior to the deadline and

$F_{12}(t) = 1 - e^{-d(t-t_0)}$, where again d is the rate at which the system fails. It is interesting to notice that this distribution function has a time shift of t_0 since even when the train travels with its maximum speed of 300 km/hr and a failure is present the system will not move to the fail state prior to $t_0 = 9.6$ seconds. The parameters of the exponential were found by the fit and were:

Low speed region: $c = 0.089 seconds^{-1}$, $d = 0.333 seconds^{-1}$, and $t_0 = 18 seconds$

High speed region: $c = 0.14 seconds^{-1}$, $d = 0.5 seconds^{-1}$ and $t_0 = 9.2 seconds$

Taking the derivatives of the above functions and plugging them into (5-1), one can obtain a closed form solution for the time paths of the failure probabilities in the high and low speed region. The expression is:

$$R_{12}(a) = dP_{12} \cdot \frac{a^2 + (c+b)a + cb}{a^3 + (c+b+d)a^2 + [cbP_{12} + d(c+b)]a + dcbP_{12}} \cdot \frac{e^{-t_0}}{a} \qquad (5\text{-}2)$$

where the fact that the Laplace transform of $R_{22}(t) = 1$ is $L\{1\} = \frac{1}{a}$.

The solution for the time paths of the probabilities was obtained taking the inverse Laplace transform. The results are presented in figures 5-12. Before commenting on the results we should note that a stochastic system with transient states and a trapping state will eventually be trapped as time tends to infinity. This can be quickly seen from equation (5-2) by applying the final value theorem in the frequency domain. This is evident in figure 5-13.

Figure 5-12 and 5-13 present interesting information regarding the failure probabilistic structure of the system given the assumptions stated above. Figure 5-12 presents the failure probability immediately after the onset of a failure; we shall call this the transient behavior. This is dominated by the failure probabilities P_{12}. One can infer that by doubling the operational speed of the train the probability of failing to react to a failure before a deadline, immediately after its onset, goes up about 50%. However the long run behavior is dominated by the rates c and d of the distribution functions. According to figure 5-13 the system operating under high speeds will fail in the best case, in half the time than it would otherwise under the low speed regime. Additionally by using this curves one can predict the mean time to failure.

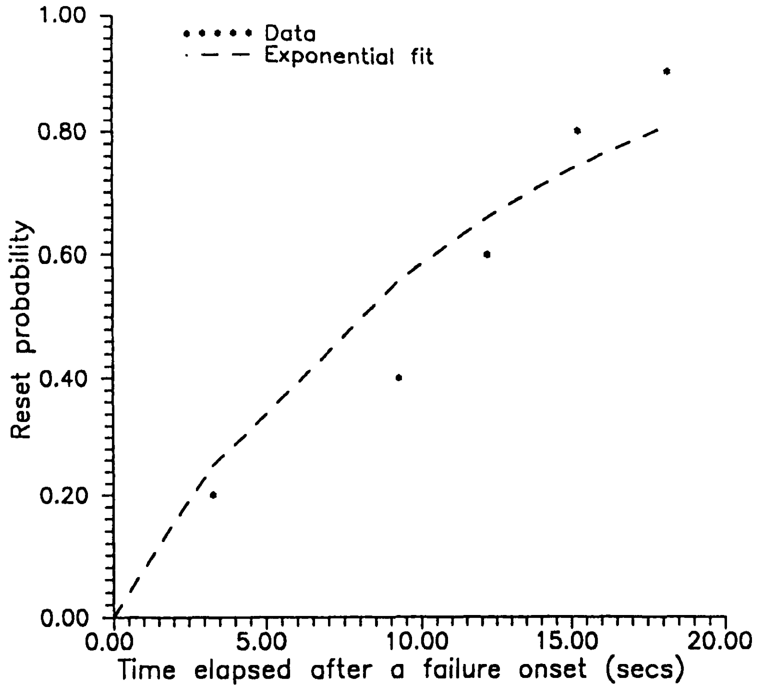


Figure 5-8: Transition function $F_{10}(t)$ for the low speed regime

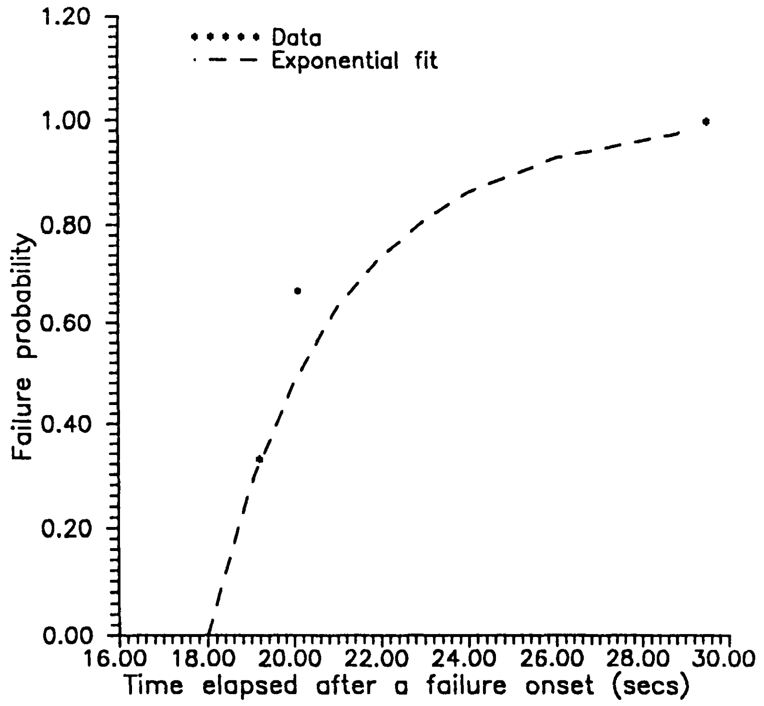


Figure 5-9: Transition function $F_{12}(t)$ for the low speed regime

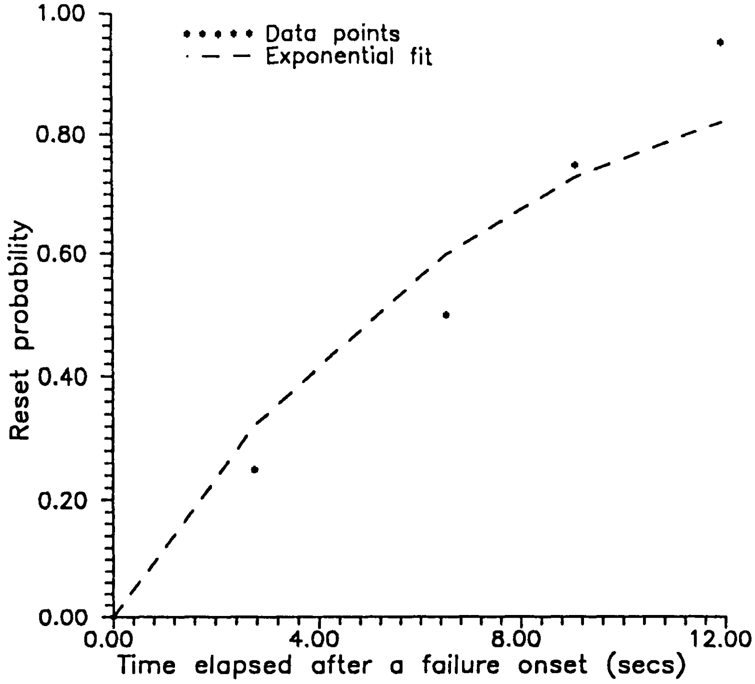


Figure 5-10: Transition functions $F_{10}(t)$ for the high speed regime

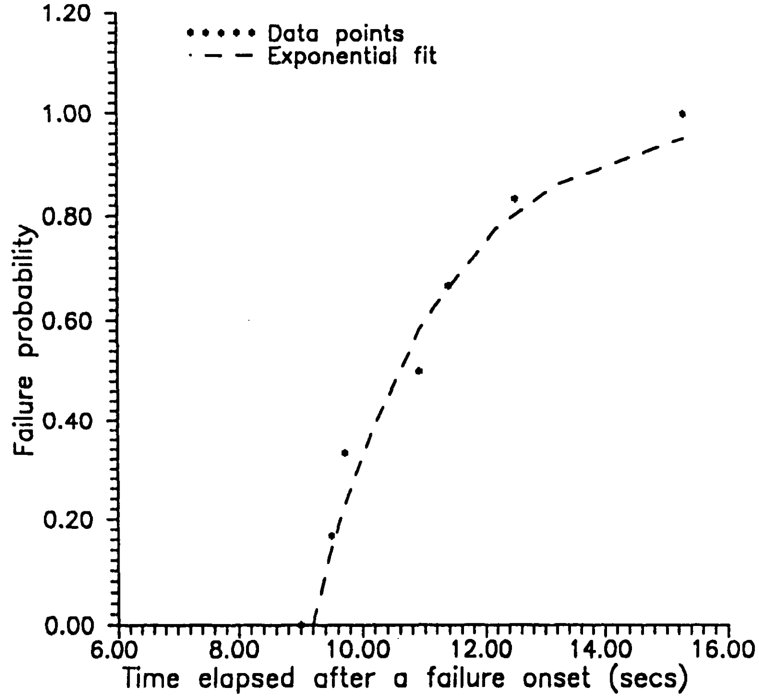


Figure 5-11: Transition functions $F_{12}(t)$ for the high speed regime

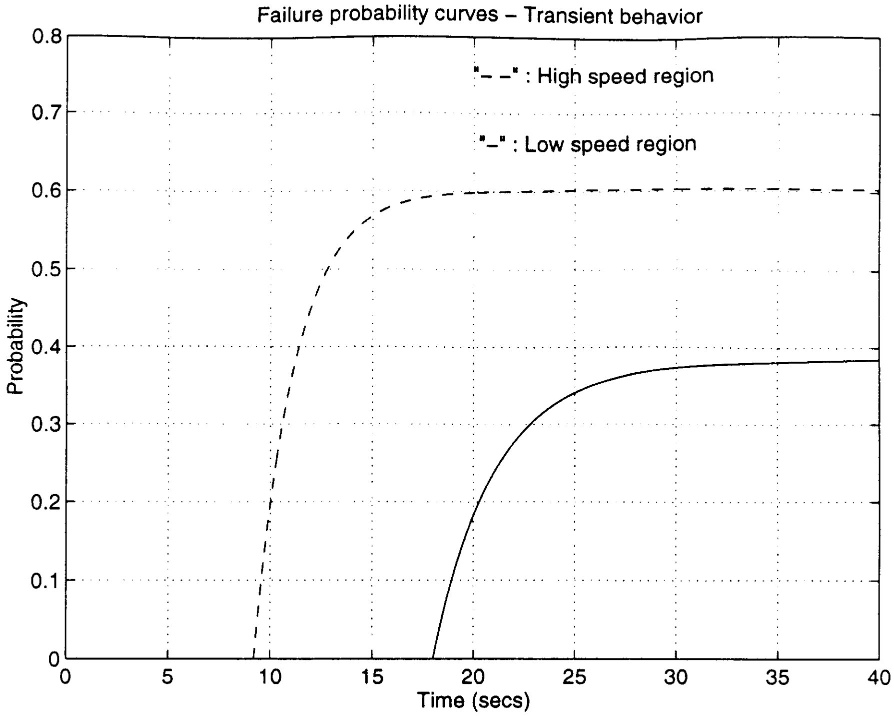


Figure 5-12: Failure probability curves - Transient behavior

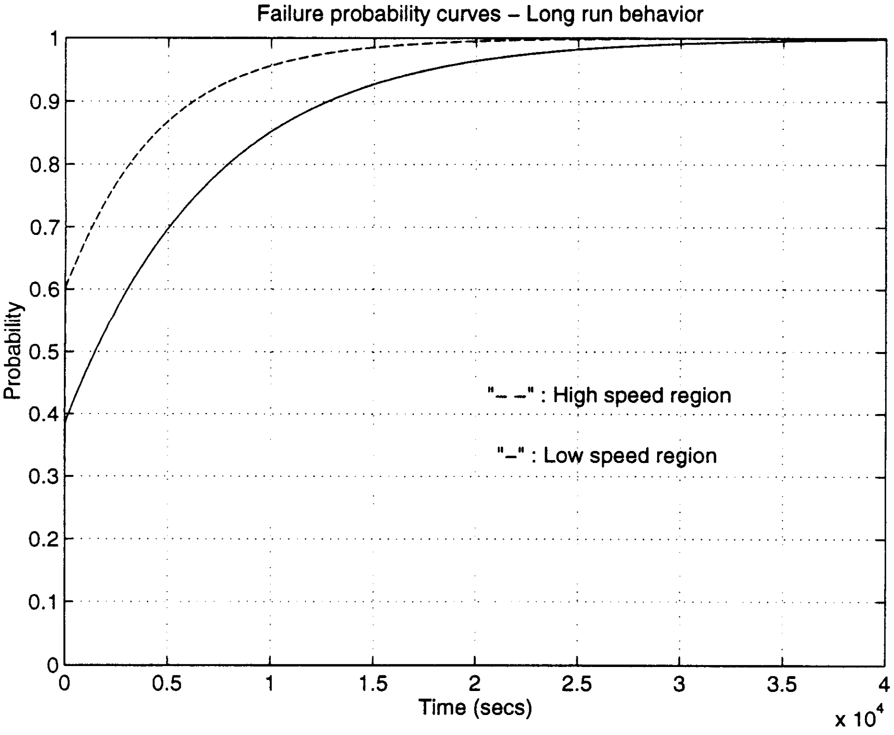


Figure 5-13: Failure probability curves - Long run behavior

5.3 Conclusions

The results that were just presented indicate that the type of control automation definitely affects human performance in many aspects. A good example is the cruise control mode used in this research, where because of the element of uncertainty with regard to speed control, subjects were more attentive to the speedometer than the other instruments, and had very good performance at the expense of proper monitoring of the rest of the 'system' as the vigilance data indicated. Additionally the cruise control mode imposed greater workload on the subjects and hence no one preferred it. Generally this hybrid control mode was disliked by all subjects.

Manual control on the other hand gave confidence in train control. Situation awareness of speed, though worse than cruise control, was better than autopilot. This seems a bit of a paradox, since manual control had a lower detection rate, as far as vigilance was concerned, and worse performance than autopilot in the rest of the situation awareness categories. This can be attributed to the feel of control over the vehicle the subjects felt they had. When the simulation was suspended the subjects answered mainly based on the feel of the speed they had and rather than the indication of the speedometer. This is similar to car driving: Very rarely do we attend to the speedometer, yet we "know" what our speed is.

Full automatic control, on the other hand, imposed the least workload on the subjects, in the sense that they had ample time to monitor the system. However two potential disadvantages must be mentioned. First is the diminished feel of control over the vehicle speed, a perceived loss of manual skill which was evident when an obstruction was pending. Second is the fact that overreliance on automation makes humans less attentive, luring them in higher level of complacency over the course of time.

From the human behavior experiment data were obtained for a safety analysis in high speed trains. A simple model based on Markov renewal theory was developed and demonstrated to show how one can trace the time changes of the failure probabilities of a stochastic system. As a suggestion for further research one can relax the assumptions of the model by using the actual distribution functions instead of the assumed exponentials, and by using the actual 'rate' at which incidents arrive can model and more accurately predict failures in an actual rail system.

Bibliography

1. Lanzilotta, Edward J. Dynamic Risk Estimation: *Developement of the Safety State Model and Application to high speed rail operation.* Ph.D. thesis, Massachusetts Institute of Technology, 1995.

2. Askey Shumei Y. *Design and Evaluation of Decision Aids for Control of High-Speed Trains: Experiments and Model.* Ph.D. thesis, Massachusetts Institute of Technology, 1995.

3. USDOT/FRA (U.S. Department of Transportation/Federal Railroad Administration) Safety of High Speed Guided Ground Transportation Systems. Human Factors Phase I: Functional Analyses and Theoritical Considerations. Final Report by T. Sheridan, E. Lanzilotta, S.Y. Askey 1994.

4. Sheridan, T. *Telerobotics. Automation, and Supervisory Control.* MIT Press, 1974.

5. Sheridan, T. and Ferrell, W. *Man -machine systems.* MIT Press, 1974.

6. Green, D. M. and Swets, John A. *Signal detection theory and psychophysics.* New York: Wiley, 1996.

7. Swets, John A. editor. *Signal Detection and Recognition by Human Observers.* New York:Wiley, 1964.

8. Macmillan, Neil A. and Creelman, C. Douglas. *Detection Theory: A user's guide.* Cambridge, England, Cambridge University Press, 1991.

9. Parasuraman, Raja and Davies, D. R. A taxinomic analysis of vigilance performance, pages 559-574 in Mackie, R. R. Ed., *Vigilance: Theory, Operational Performance and Physiological Correlates,* Plenum, New York 1977.

10. Parasuraman, Raja. Memory load and event rate control sensitivity decrements in sustained attention. *Science* 1979, Volume 205, 724-727.

11. Parasuraman, Raja and Mouloua, Mustapha. Interaction of signal discriminability and task type in vigilance decrement. *Perception and Psychophysics* 1987, 41 (1),17-22.

12. Molloy, Robert and Parasuraman, Raja. Monitoring an Automated System for a Single Failure: Vigilance and Task Complexity Effects. *Human Factors* 1996, 38 (2) 311-322.

13. Loeb, M. and Binford, J. Examination of some factors influencing performance on an auditory monitoring task with one signal per session. *Journal of Experimental Psychology*, 83 (1), 40-44, 1970.

14. Endsley, Mica. Toward a Theory of Situation Awareness in Dynamic Systems. *Human Factors*, Volume 37 Number 1, 1995.

15. Endsley, Mica. Measurement of Situation Awareness in Dynamic Systems. *Human Factors*, Volume 37 Number 1, 1995.

16. Drake, Alvin W. *Fundamentals of Applied Probability Theory*. McGraw-Hill, New York, 1967.

17. Gallager, Robert G. *Discrete Stochastic Processes*. Kluwer Academic Publishers, 1996.

18. Ross, Sheldon. *Stochastic Processes*. John Wiley, New York 1983.

19. Medhi, Jyotiprasad. *Stochastic Processes*. John Wiley, New York 1982.

20. Cinlar, Erhan. Markov Renewal Theory: A Survey. *Management Science* 1975, 21, 727-752.

21. Cinlar, Erhan. *Introduction to Stochastic Processes*. Prentice-Hall 1975, Englewood Cliffs, New Jersey.

22. USDOT/FRA (U.S. Department of Transportation/Federal Railroad Administration) Engineman Stress and Fatigue: Pilot Tests. Final Report by G. I. Kuehn 1992.

Appendix A

Training Tutorial

1. OUTLINE

The current document describes an experiment that aims at addressing vigilance, situation awareness issues and their relationship to automation in high speed rail systems as well as to model risk probability in the "high" versus the "low" speed regime. To explore these effects the High-Speed Rail Simulator located in the Laboratory for Human Factors Research in Transportation at the Volpe Center will be used.

This document is conceptually divided into two parts. The first part is a tutorial on how to operate the simulator. It draws heavily from Dr.Lanzilotta's (Ph.D. ME 1995) system description tutorial as he the one that built the Simulator it differentiates however substantially from it at parts that are important to the present research and features that where added to the simulator by the author. It provides a general description on how rail systems operate and then goes on to present details regarding vehicle operation, speed and position control,automation modes and so on. The second part addresses operator performance requirements and summarizes the bonus-penalty system in regard to the present experimental study.

Participation in the experiment consists of following phases: First *review* this document before coming to the Volpe Center. The experimenter will give you a short 10 min multiple choice written quiz once you've read it. This is intended to gauge your understanding of the system thus enabling the experimenter to spot deficiencies that need to be covered in the training session that follows. Note that there is a minimum performance requirement of 50% on the quiz to continue on the next phase which is the *training session.* This is a three hour session that gives hands on exposure to the system. The first hour you will be acquainted with vehicle operation and introduced to emergency scenarios allowing you to learn the proper response to such situations. The remaining two hours are devoted to tentative rides both with and without the emergency scenarios that will further familiarize you with the system and its operating modes. Unsatisfactory performance in this 2 hour session -non compliance with speed limits, ignorance of the dead man alerter e.t.c - will result in your dismissal from the rest of the experiment.In such a case you will be eligible for payment for the training phase only. The final phase consists of the *experimental trials* which will last overall 3 hours approximately and take place the same day. You will be eligible for payment for the training phase upon completion the practice and test session unless disqualified earlier in the training session.In such a case you will be eligible for payment only for the training phase completed. The payment rate is $10 per hour. There additional opportunities for gain or loss depending on your performance with regard to the experiment.

2. RAIL SYSTEMS OPERATION-BLOCK SIGNALING

Rail systems have traditionally used a system known as ***block signaling*** for the control of trains in the rail system. With block signaling, the track is divided into fixed length chunks known as blocks. While the length of each block does not change, different blocks are not necessarily of equal length. Typically, shorter block lengths are used in the near vicinity of stations, while longer block lengths are used in regions away from the stations. Block lengths are generally of the order of one mile. In the road system used in the simulation, all blocks between stations are 2 km, and all blocks in the loop sections are 1 km in length.

At the boundaries of each block there is a ***signal light***. This signal light displays a color-coded signal, which indicates the maximum speed permitted throughout the block. The signal acts as a ***dynamic speed limit***, and it is the responsibility of the vehicle operator to identify the signal as the block boundary is approached and set the vehicle speed accordingly. A fundamental rule in block signaling is that no more than one train can occupy a block at any given time. A red signal is used to indicate that the block is currently occupied by another train, and the approaching train is not permitted to enter that block. The blocks that precede the occupied block have signal levels which ensure that the train can be slowed in time to stop before entering the occupied block.

In addition to the speed limits imposed by the block signal system, there are also ***civil speed limits***, which are static. These limits are either memorized or written down by the operator. In all cases, the prevailing speed limit is the ***lesser*** of the block signal limit and the civil speed limit. The exact specification of signals used and speed limits associated with those signals is a design parameter for a rail system, and varies from system to system. In the simulation system, a five-aspect signaling system is used. This means that there are five color codes used in the system, with the codes defined as shown in table A.1.

If a train was occupying block 157, then the signal at the entrance to block 157 would show STOP (red), the signal at the entrance to block 156 would show RESTRICTED (red/yellow), the signal at the entrance to block 155 would show APPROACH (yellow), the signal at the entrance to block 154 would show APPROACH MEDIUM (green/yellow), and the signals at blocks prior to 154 would show CLEAR (green). The speed limits apply to the entire block, which means an approaching train must reduce speed to the limit before reaching the entrance of the block. So, in this example, another train approaching the train in block 157 must be going slower than 230 km/her before entering block 154, slower than 150 km/hr before entering block 155, and slower than 80 km/hr before entering block 156 (figure A-1).

Table A.1: Rail Signal Codes

COLOR	CODE	ACTION
Red	STOP	Not permitted to enter the block
Red/Yellow	RESTRICTED	maximum speed of 80 km/hr in this block
Yellow	APPROACH	maximum speed of 150 km/hr in this block
Green/Yellow	APPROACH MEDIUM	maximum speed of 230 km/hr in this block
Green	CLEAR	maximum speed of 300 km/hr in this block

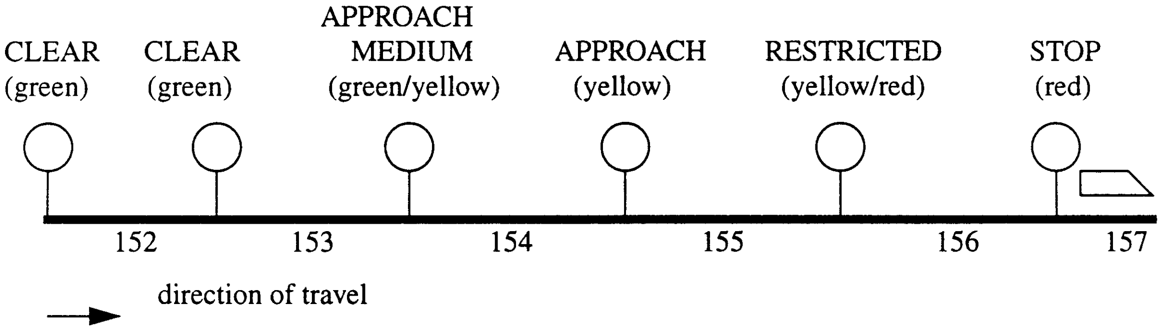


Figure A-1: Block Signaling System

Located throughout the system are position markers known as kilometer posts. The use of these by vehicle operators is discussed in detail in the next section. It is important to note the difference between block signals and kilometer posts. At the entrance to each block, there is a signal board which identifies the block number and displays the current signal level for that block. Because block boundaries occur at 2 km intervals on the main line in this system, there is also typically a kilometer post at the block boundaries. So, for example, block 13 comprises the distance of track between kilometer posts 26 and 28. This provides opportunity for confusion: The entrance to block 13 is marked by kilometerpost 26 in one direction of travel, but when traveling in the opposite direction, the entrance to the same block occurs at kmpost 28. Operators must take care to differentiate between block identifiers and kilometer posts, as the relationship between them is not as simple as it might at first appear.

3. CENTRAL TRAFFIC CONTROL

The main element in a rail system operation is the **Central Traffic Control** (CTC) located at a fixed position in the system. CTC operators -known as the dispatchers- have the task of monitoring and coordinating the operation of several vehicles that must share resources (such as the track system). To carry out this task, the CTC operators have control over the switches in the system (thus determining which train has priority of using a certain portion of the track in the event of a conflict) and can set signal levels manually thus regulating the speed of the trains. Given this perspective, the CTC operators represent the utmost level of authority which is higher than the wayside signals, the operating rules, and any other influence. Any conflict or anomaly regarding vehicle operation spotted by the locomotive engineers should be reported to the CTC which in turn is responsible for giving instructions on how to proceed.

The rail system used in the simulation is a fictitious rail system connecting two stations, named West Station and East Station (figure A-2). The two stations are connected via a single track which is 50 km in length. At each end of system, beyond the stations, there is a loop of track which is used to turn the vehicles around for the return trip. The system is operated as a high-speed shuttle between these two stations. Although the CTC simulation element is capable of handling multiple vehicles in this experiment there will be only one vehicle in operation. That vehicle will travel from one station to the other, discharge passengers, loop around to reverse direction, board new passengers, and proceed to the other station. This procedure is followed throughout the duration of the shift.

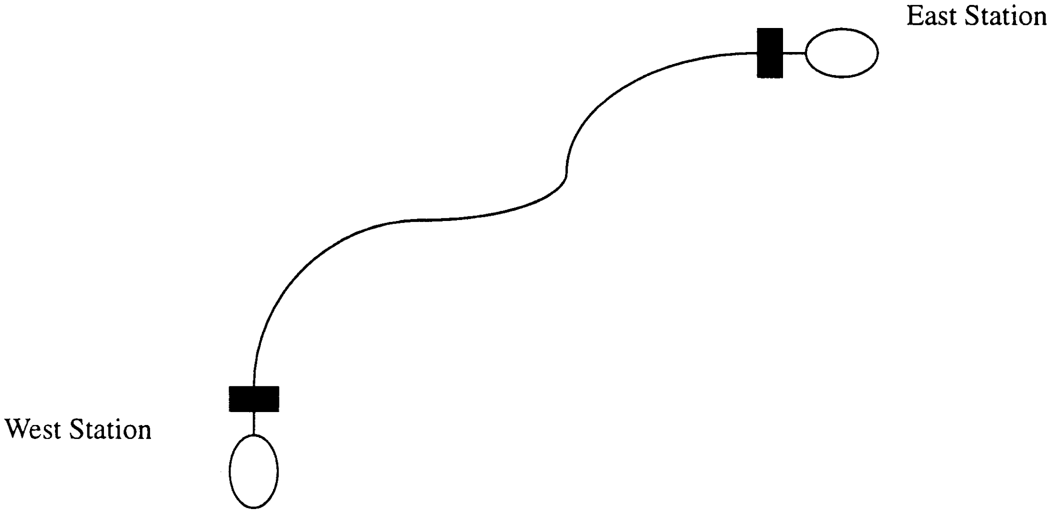


Figure A-2: Track Layout, Simulated Rail System

4. VEHICLE OPERATION

Instrument panel layout

The instrument panel contains gauges that display information regarding the state of the vehicle. A schematic representation of the instrument panel is shown in figure A-3 below. The largest gauge in the middle is the speedometer used for controlling the vehicle speed. Secondary gauges display the brake tank pressure, the bearing temperature, and the trolley voltage. There are also light indicators regarding the operating control mode of the train, the motors current, the Automatic Train Protection (ATP) and the Alerter systems (discussed subsequently) the door and the emergency stop status. The signals of the current and the next block are displayed above the speedometer. In the simulation the communication between the CTC and the vehicle operator is done textually via the communication area of the instrument panel by using the keyboard. While a message is being composed, it appears in the lower portion of the communications display, and is visible to only the operator. When the return key is pressed, the message is "sent" to all other operators on the system, including the CTC operator. The message then appears in the top portion of the communications area and is visible to all operators in the system.

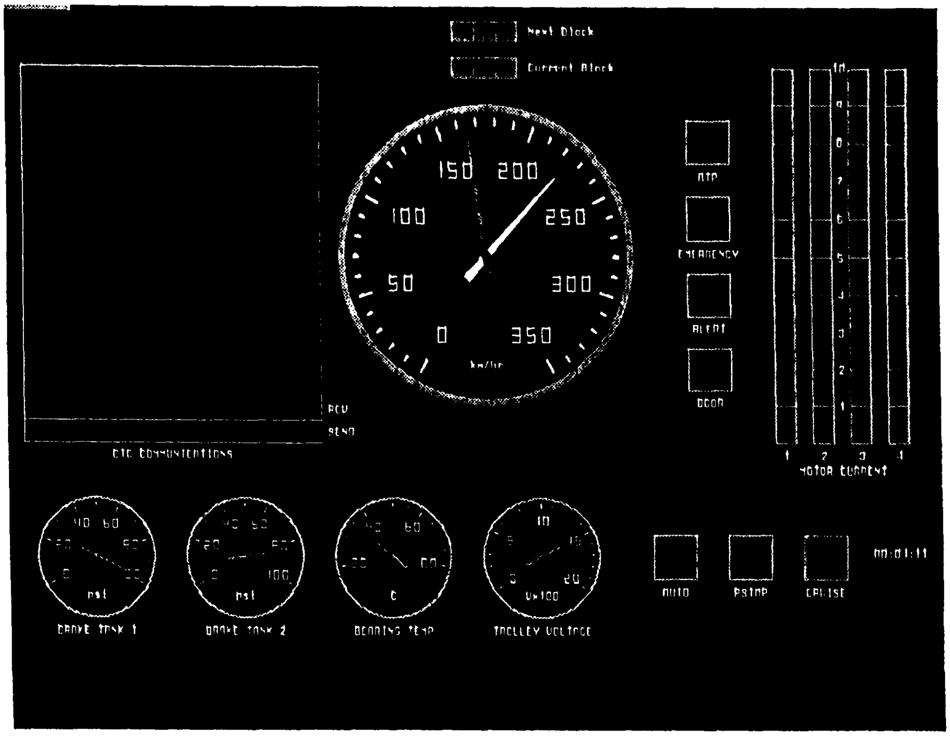


Figure A-3: Instrument panel layout

a. Speed and Position Control

The most important aspect of train driving is the control of vehicle speed. This is not an easy task as might appear at first. Due to the large inertia of the vehicle (typically in the range of hundreds of tones) the distances required for the acceleration and decceleration of the vehicle are very large. As a result the operator does not get immediate feedback in regard with the outcome of his/her action. The picture becomes gloomier as speed increases since distances required for braking increase whereas the allowable response time to unexpected dangerous situations (such as a sudden detection of an unexpected obstacle) decreases. A reasonable solution that is being implemented is rail systems is the utilization of civil speed limits at risk prone areas such as urban areas and grade crossings.

In the simulation ***blocks 0 1 11 12 and 13 have civil speed limits*** due to grade crossings present is these blocks. Grade crossings areas are the only places where car/track obstructions can occur and thus the civil speed limit has been set to a maximum speed of 100km/h. Consequently the train's speed on a block is governed by stepwise civil speed limits and signals. Observation of these limits is absolutely necessary for the safe and efficient operation of all vehicles because of potential risks otherwise involved such as collision between trains or train and vehicle in gradecrossings, derailment at switches or curves and so on. This implies that if a train is approaching a block with a particular speed limit, its speed must be reduced to below that speed limit before -not after- it enters the block. The operator uses the position of the vehicle to obtain the current speed limit, through a combination of civil speed limits and block signal states (which are observed on the wayside and provided in the instrument panel). The operator then uses the control lever to adjust the speed of the vehicle accordingly. The vehicle operator gets information about vehicle speed through the speedometer, which is located on the instrument panel in the locomotive cab. In the train simulation, this speedometer is implemented as a round dial gauge. The units shown are kilometers per hour (km/hr), and the available range of speeds is from 0 to 350 km/hr. The major increments of the gauge display are 50 km/hr, with minor increments each 10 km/hr. The red pointer indicates the current speed, while the smaller yellow pointer (underneath the red pointer) indicates the set speed (used by the automation systems).Another important task of the vehicle operator is monitoring the position of the vehicle in the rail system. This is done by monitoring the out-the-window view. Typically, at one mile intervals (or kilometer intervals depending on the metric system used), a post is placed on the wayside with numbers indicating the mile (kilometer) marker known as "milepost" ("kilometer post"). The simulation provides options for using either mile or kilometer posts. In the current experiment we will be using kilometer posts. In order for the operator to control the speed appropriately he needs to become familiar with the braking and accelerating characteristics of the vehicle. Figure A-4 shows the full service braking profile and Figure A-5 the emergency braking profile at ground level.

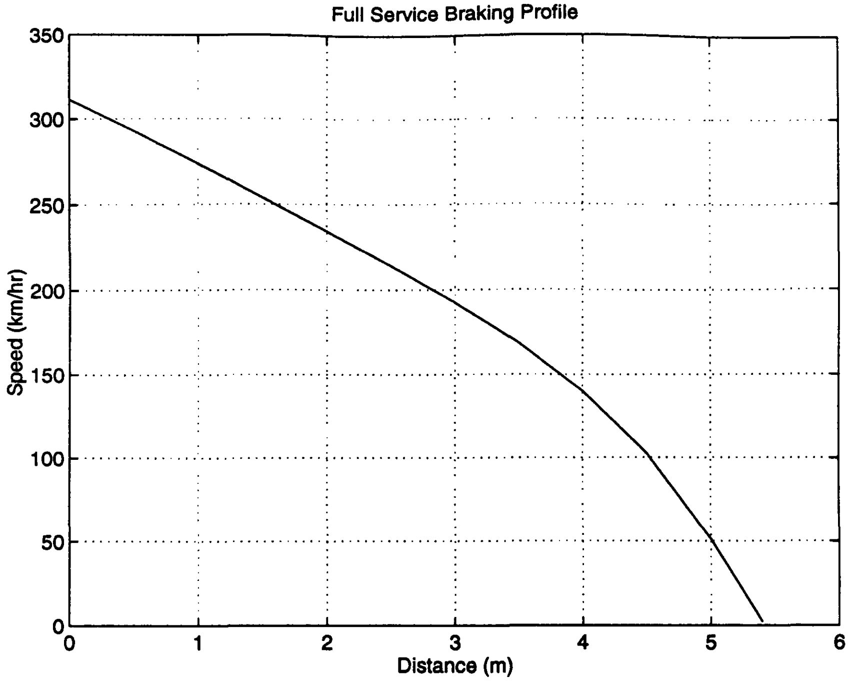


Figure A-4: Full-Service Braking Profile

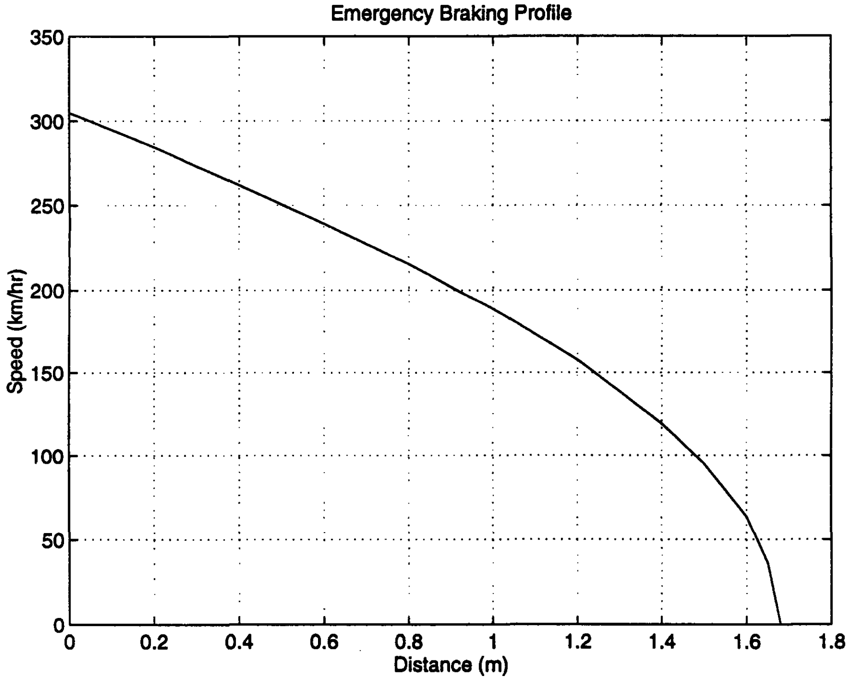


Figure A-5: Emergency Braking Profile

The information provided by these curves is ***crucial*** and can be used in many different ways: The full-service braking profile can be used to estimate appropriate braking points under manual control. For instance assuming that the vehicle is traveling at 300 km/h Eastbound and is currently at block 7. Remembering that block 11 has a civil speed limit the operator must apply full service brakes before entering block 9 since as can be seen from the ***full service profile the braking distance from 300km/h to 100km/h is 4.5 kilometers*** (roughly the distance of two blocks) approximately. Similar considerations should be taken while approaching the Stations. In another case the above profiles can be used to determine the appropriate braking action is a case of an obstruction. Obstructions in the grade crossings are visible from the vehicle for a distance of almost one kilometer therefore this braking performance provides adequate opportunity for the vehicle operator to detect and react properly to an obstruction in a grade crossing. The ***braking distance from 100 km/hr is around.8 kilometers using full service braking***. However in the event that the grade crossing obstruction isn't spotted in time the operator may have to use (depending on the distance to the obstruction) ***emergency braking which will result in a complete stop from 100km/h in less than 200 meters***. Emergency braking has a high cost both in equipment and service quality and should be considered the last resort as discussed in the later part of this tutorial.

b. Control Modes

The simulation has ***four control modes*** available for vehicle operation. These are:
a) manual mode, b) cruise control mode, c) programmed stop mode, and d) autopilot mode. The latter three are considered automatic control modes because part of the vehicle control task is performed by a computer-based control system.

In ***manual mode*** the locomotive engineer is actively involved in train driving using the combination control lever to provide all thrust and brake commands required to achieve speed and position control of the vehicle. He is ***operating*** the train in all aspects.

In ***cruise control mode***, the automatic control system applies the appropriate level of thrust force to maintain a constant speed setting. The vehicle operator can invoke the cruise control mode, by depressing the cruise control enable switch (F5 key). When the cruise control mode is enabled, the cruise control indicator light (green) is illuminated, and the yellow pointer on the speedometer indicates the set speed. The vehicle operator can alter the set speed by depressing the "up-arrow" and "down-arrow" keys on the keyboard. With the depression of each key, the set speed is adjusted up (or down) by 1 km/hr. This feature allows the operator to "tune" the set speed. When the operator adjusts the set speed down to a lower speed, the vehicle will coast down to the lower speed. The level of control in this mode is somehow split between the human and the machine. By depressing the cruise control switch (F5 key) while in cruise control mode the operator returns to manual control (toggle key).

In ***programmed stop mode***, the automatic control system applies the appropriate level of brake force to stop the vehicle at a specific position. We will not be using this mode in the experiments.

In the ***autopilot mode (or automatic control mode)***, the automatic control system applies the appropriate level of thrust and brake forces to follow a predetermined speed trajectory. The vehicle operator invokes the autopilot function by depressing the autopilot enable switch (F7 key) while the vehicle is in motion. For best performance, the vehicle must be traveling at a speed greater than 10 km/hr when the autopilot is activated. When the autopilot mode is enabled, the autopilot indicator light (blue) is illuminated. In addition, the pre-determined speed setting is indicated by the yellow pointer on the speedometer. Once vehicle control has been assumed by the automatic control system, all necessary vehicle control commands are provided by the control system. The automatic mode is programmed to abide to the set speed limits (civil or signal). F7 key is toggle as well and can be used form within the autopilot mode to return to the manual control mode.The task of the operator is reduced to monitoring the vehicle and wayside signals, looking for potential problems in operation. The engineer's role is shifted from the operating to the supervising level.

For reasons of safety, application of the brakes will always disengage any automatic control system. As a result, it is not possible to engage an automatic control mode when the brakes are in use. If the operator attempts to engage an automatic control system while the brakes are applied, the system will not respond to that mode command, and the vehicle will remain in manual mode. Obviously one can switch to manual mode regardless of the automation mode she/he is using just by moving the control lever backwards to its center (coast) or braking position.

c. In-Cab Signal System

This system displays the signal of the current and the next block inside the cab. Bringing the signal indications inside the cab has several advantages: Signals cannot be obscured by bad wheather conditions (fog, snow) and they provide a form of preview to the locomotive engineer since at high speeds (in excess of 220 km/h) it is virtually impossible to detect a wayside signal and act appropriately in time. It is noteworthy that in case of a discrepancy between the wayside signal and the in cab signal the locomotive engineer is required to follow the in cab signal indication. The simulation uses such a system. Two horizontal light indicators above the speedometer indicate the current and the next block signal.

d. Traction System

The simulation uses four traction motors as this is the number usually encountered in practice. The engineman's principal guide to locomotive performance is the ***ammeter*** which shows the current actually going through a traction motor. By moving the control lever the locomotive engineer governs the amount of current through the motor windings thus the tractive force (if any) provided by the motors. At the center position no tractive force is applied whereas moving the lever forward form the center position the tractive force is increased proportionally to the amount the lever is moved. Moving the lever backwards from its center position all traction power is removed from the motors and brakes are applied. The automatic modes available to the simulation have the ability of determining automatically the amount of tractive force required to operate the train.

The dashboard display includes four current meters (ammeters), which display the level of current through each of the four traction motors. In manual mode, these displays will respond directly to the input at the combination control lever, while in automatic mode, they provide a mechanism for observing the operation of the automatic systems. The traction motors are protected by circuit breakers, which will interrupt the flow of electrical power to the motors if a failure condition is detected. Each of the four motors has a separate circuit breaker. Under certain circumstances one traction motor will fail. The occurrence of this event can be observed through the ammeters---when one (or more) of the ammeters does not respond with the others. The procedure for resetting the failed motor is as follows: remove all power from the other traction motors, by moving the combination control lever to a coast or brake position. b) Depress the appropriate traction motor reset switch (Fl through F4 on the control panel). c) Apply tractive power manually, using the combination control lever. d) Resume the control mode previously in use. If any of the traction motor reset switches are depressed while power is applied, a safety system causes all of the traction motor circuit breakers to trip, preventing motor overload. In this event ***all*** of the circuit breakers must be reset to resume proper operation.

e. Brake System

There exist several different systems for braking train vehicles ranging from conventional such as ***air braking*** to the more advanced such ***dynamic*** and ***regenerative*** braking. The simulation utilizes the conventional braking system namely air brakes. Compressors are used to compress air which is stored in tanks within the locomotive. When the brakes are applied, pressure is released from the tanks, causing the brakes shoes to contact the rotating surfaces resulting in a friction force which deccelerates the vehicle. All trains -the simulation included- have two modes of braking: ***service*** and ***emergency*** braking. Under normal operating conditions the locomotive engineer uses the control lever to deccelerate the vehicle by pulling it backwards from its center position. Application of the maximum available braking force is known as full service braking. However there might be circumstances in which the full service braking decceleration is inadequate (collision avoidance). In such a case the operator uses the emergency braking by depressing the appropriate control switch on the instrument panel, which results in the release of all of the pressure in the brake system. This provides the maximum possible brake force. Once the emergency brakes have been applied, the emergency brake indicator on the instrument panel becomes red and the brakes cannot be released until the vehicle comes to a complete stop. After that the operator can disengage the emergency brakes by pulling back the control lever to a position which results in application of the service brake and then depressing the appropriate switch. At that point, the emergency brake indicator light will be extinguished, and the vehicle will be ready to continue with normal operation.

As shown in the dashboard there exist two round gauges which display the pressure in the brake tanks. In the train simulation, there are two brake tanks. The corresponding gauges are round dial gauges, calibrated in units of pounds per square inch (psi), with a range from 0 to 100 psi. The normal reading when the brakes are not applied (i.e., the nominal high pressure) is approximately 98 psi. When full service braking is applied, the pressure drops to approximately 22 psi, and the pressure further drops to 0 psi under emergency braking. If there is a failure in the braking system, one or both of the tanks may show a reduction in tank pressure. This situation will result in the brakes being applied without being commanded by either the operator or the control system. The procedure for rectifying this situation is to switch to an alternate brake compressor. This is accomplished by depressing the brake compressor switch (F10 key). The pressure in the faulty tank will then rise to the appropriate level.

f. Bearing Temperature Display

Under normal conditions the bearings temperature rises (falls) as speed increases (decreases). In a faulty bearing the temperature will rise regardless of the vehicle speed due to increased friction and its temperature will gradually rise to 70 ^{0}C. Even when the train comes to a complete stop the temperature will remain high due to the wear of its metal parts. In such a condition the operator is required to depress the F11 key which will drive cooling air at the bearing thus restoring the temperature back to normal.

g. Trolley Voltage Display

The right most dial gauge on the instrument panel displays the voltage available from the power supply grid to the traction motors of the vehicle. Under normal operating conditions that will always be 1500 V DC.

h. Alerter System

The Alerter System (some times called the "deadman" alert system) is an onboard safety system whose purpose is to ensure that the driver is alive and functional at the controls. The principle behind the system is the requirement for periodic input to it - by depressing a button - from the vehicle operator.In the case that the system does not receive an input from the operator within a period of 42 seconds from the last depression it generates a warning reminding the operator to do so. The warning consists of a flashing yellow indicator light and an audible chime. If the operator does not respond within 10 seconds of the onset of the warning, the system applies the emergency brakes. In this scenario, both the alerter warning light and emergency brake light will be illuminated. The operator resets the alerter used in the simulation by depressing the Esc key on the keyboard.

i. Automatic Train Protection (ATP) System

The ATP is another safety related system designed to prevent a train from overspeeding within a block, thus reducing the risk of an accident. As has been stated the effective speed limit in a block is the lesser of the speed limits dictated by the block signal and the civil speed limit (if any) present in the block. This limit is calculated from the ATP based on the position of the train and if its speed is larger than it by more than 15km/h emergency brakes will be applied. However if the trains speed is larger the effective speed limit but by an amount not grater than 15km/h then a warning is issued that consists of an audible chime and the yellow indicator light on the dashboard. The operator then has 20 secs to reduce the speed of the vehicle at an acceptable level or emergency brakes will be applied while the corresponding dashboard light will remain lit.

j. Door Control

The vehicle operator is responsible for controlling the state of the passenger doors. The doors are to be opened when the vehicle is stopped in the station. In principle, the doors must not be opened at any other point in the system, for the protection of the passengers. Door control is accomplished through the door control button (F8 button). The state of the doors is indicated by the door indicator light (red) on the instrument panel. When the light is illuminated, the doors are open. Depressing the door control button while the vehicle is stopped will cause the state of the doors to change---if the doors are open, they will be closed, and if the doors are closed, they will be opened.

A safety system prevents the doors from being opened while the vehicle is in motion. Door control commands while the vehicle is in motion will be ignored. If the vehicle is stopped with the doors open, any attempt to move the vehicle will cause a penalty application of the emergency brakes.

4.EXPERIMENT DESCRIPTION

The experiment consists of driving the simulated train from West Station to East Station and back (one shift) using one particular control mode. There will be a total of three runs each completed using a different automation mode namely manual, cruise control, and autopilot modes. Note while these will be the primary operating modes during each shift you are allowed to switch modes for brief periods only as dictated by emergencies posing danger to the safe operation of the vehicle (obstructions, speed limit violations). Since high speeds are desirable, bonus points are given for early arrivals as shown in figure A-6

Under average performance you should be able to complete the trip within 45 minutes in which case no bonus/penalties apply. In any other case the scheme at figure A-6 will apply. However this does not imply that you should neglect or try to circumvent safety systems such as the ATP. In fact penalties regarding inappropriate application of the emergency brakes will result as shown in the following table:

Emergency Stop braking	Penalties
Triggered by ATP :	-2000
Triggered by Alerter:	-2000
Triggered by the operator -collision avoidance	-100
Triggered by the operator- No reason	-300

Table A-2: Penalties resulting from application of the emergency brakes

Your self as test subject is required to operate a virtual high speed train and comply with the operational requirements and operating rules of such a system as described in this tutorial. Speed and signal compliance for instance are considered key performance items regarding the safe operation of the vehicle. As stated in the beginning of this document one of the goals the current experiment is to investigate and assess vigilance performance and situation awareness of the operator in the context of high speed train driving. This is done by monitoring and recording each operator's (test subject) actions throughout the test sessions, to a set of emergency scenarios which are presented below. In addition as an incentive, there is a bonus system which provides monetary rewards for good performance. If operator performance does not fall within certain minimum criteria, penalties may be assessed.

Emergency Scenarios

Response to abnormalities in the operation of the rail system where introduced earlier while describing the system. However for the purposes of this research we will be using a later version of the simulator in which failures are ***reset automatically*** according to the time that has elapsed or the distance that the train has traveled since the onset of the failure, in the event that the subject has not been able to spot them and reset them earlier. There will be 4 kinds of emergency scenarios used for a different purpose each:

1. *Motor failure*: In this scenario the circuit braker for one motor is tripped requiring reset. The failure results in an absence of current flow through one of the motors indicated on the motor ammeters located in the instrument panel. In this case the correct response is to first remove power from all the motors by pulling back to braking position the control lever then pressing the appropriate key for reset and finally resuming control of the vehicle using the previous control mode. The ***failure remains active for 12 seconds*** and unless reset is ***automatically reset***. Figure A-7 summarizes the bonus incentive system used for motor failures.

2. *Bearing failure:* The bearing temperature fluctuates around a value depending on the speed of the train.Under no condition should the temperature rise above 55 ^{o}C for safety reasons. If it does the reset procedure is to press a key (F11) that will drive cooling air to the bearings thus lowering their temperature. The ***failure will remain active for 20 secs*** before it ***resets automatically***. The bonus incentive system is presented at Figure A-7.

3. *Brake failure*: One of the brake tanks loses pressure. The pressure loss is indicated in the brake tank gauge. Once the failure is detected the correct response is to switch to an alternate compressor by depressing the appropriate key on the computer keyboard (F10). This ***failure though differentiates*** from the previous 2 in that a ***certain distance of 800 m*** is given to the subject to reset the failure. Now the reaction time depends on the speed of the train. The bonus given for noticing and resetting the failure is 800 points.

4. *Obstruction Hazards:* There are points in the rail system where highway roads and rail tracks intersect. In total, there are five grade crossings, one each located in blocks 0, 1, 11, 12, and 13. At each grade crossing, highway vehicles (cars) can cross in front of the train from either direction. These vehicles are visible at over a half kilometer distance.In each of the blocks containing grade crossings, there is a civil speed limit of 100 km/hr. This means that the maximum speed of a vehicle in the block is 100 km/hr unless a lower speed limit is imposed by a signal in that block. Traffic at the grade crossings arrives according to a probabilistic process. A car will proceed across the crossing only if there is sufficient distance to clear the

crossing before the train. (In other words, a car will not proceed if there is not enough room). However, it is possible for a car to become disabled as it is crossing the tracks, which will result in an obstruction for the train. In this event, the train operator must bring the train to a stop before the intersection. If the train is not stopped in time, a collision will occur, which will be indicated by a cracked windscreen. It is important that the train operator be able to quickly assess the crossing traffic and determine whether the train must be stopped. On one hand, a collision is a major event, and will result in a significant delay in operation. On the other hand, stopping the train unnecessarily will also cause delays in service. It is up to the vehicle operator to evaluate the situation and determine the best course of action under these constraints. If a collision occurs, the vehicle operator must stop the vehicle and immediately contact the CTC operator to report the collision. In the case of a collision, the windscreen will appear "cracked." This crack will remain for the remainder of the shift. Bonus/penalties are assessed according to the following table:

Obstruction outcome	Bonus/Penalty
Collision avoidance:	+1000
Collision Speed:>= 90km/h:	-1000
Collision 70 km/h<= Speed <= 90km/h:	-800
Collision 40 km/h<= Speed <= 70km/h:	-600
Collision 10 km/h<= Speed <= 40km/h:	-450
Collision Speed <= 10km/h:	-250

Table A-3: Obstruction hazards bonus/penalty scheme

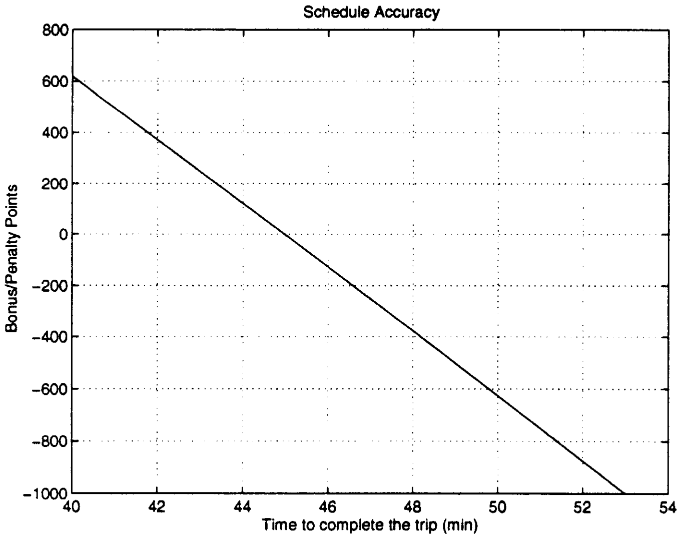


Figure A-5: Schedule accuracy bonus/penalty points

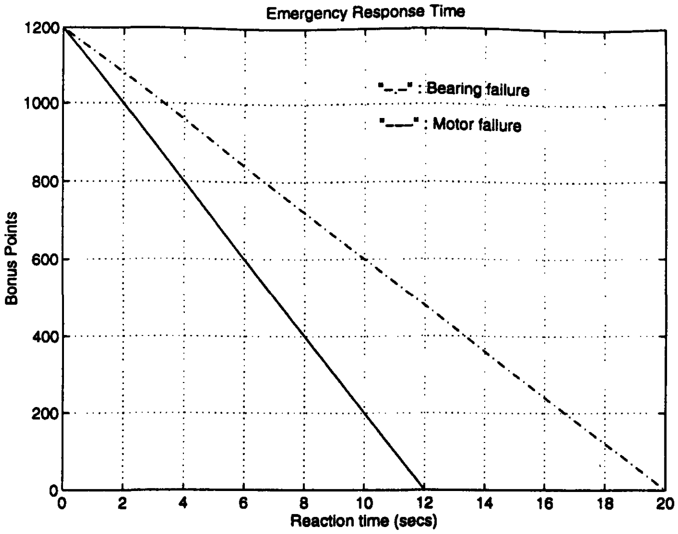


Figure A-6: Emergency response time bonus points

From the graphs one can infer that a bonus of 100 points per second is given for the motor failures as opposed to 60 bonus points per second, whereas the bonus/penalty scheme for schedule accuracy is approximate 12 bonus points per second.

During each shift (from East station to West station and back) you will experience on **average 2 emergency scenarios** from either category 1, 2, or 3 and in addition to these you might get or not obstructions. If you think that an emergency scenario is present you should respond appropriately as described above. However if you took such an action while not necessary you will be penalized -300 points. For instance if you pressed the brake pressure reset key (F10) while there was no reason you lose 300 points. If an emergency was present and you took no action then the penalty is -300 points.

Situation Awareness

To investigate the relationship between control automation and situation awareness the temporary freeze technique. The simulation is suspended at random locations during vehicle operation the screens are blanked and questions relating to the overall state of the vehicle are asked such as the current block number and position (kilometer post), train speed, speed limit of the next block etc. Each correct answer to the questions asked gets 1000 bonus points while there is no penalty for wrong answers (guessing is encouraged).

Appendix B

Review Quiz

Select only one answer form the ones presented below based on what is stated in the tutorial document.

1. What is the length of a block used throughout the simulation?

   a. 1km

   b. 2 km

   c. 1 mile

   d. None of the above.

2. How many trains are allowed to occupy a single block at any given time?

   a. one

   b. two

   c. three

   d. none of the above

3. Assume that the train in the simulation is traveling Eastbound and enters block number 10. What will the indication of the kilometer post be when entering the same block but traveling Westbound?

   a. 19

   b. 20

   c. 21

   d. 22

4. What is the maximum allowable speed in an upcoming block with a Green/Yellow signal?
    a. 80 km/hr
    b. 150 km/hr
    c. 230 km/hr
    d. 300 km/hr

5. In the simulation which blocks have civil speed limits?
    a. 0, 1 and 11
    b. 0, 1, 10, 11, 12
    c. 0, 1, 11, 13, 14
    d. 0, 1, 11, 12, 13

6. What is the approximate braking distance from 300km/h to 100km/hr under full service braking?
    a. 1.5 km
    b. 2.5 km
    c. 3.5 km
    d. 4.5 km

7. What is the approximate braking distance from 100km/hr to a complete stop under emergency braking?
    a. 1km
    b. 500m
    c. 200m
    d. 100m

8. In the event of a motor failure how long will the failure remain active before it is reset automatically?
    a. 20 secs
    b. 12 secs
    c. 30 secs
    d. 50 secs

9. In the event of a bearing failure how long will the failure remain active before it is reset automatically?

- a. 20 secs
- b. 12 secs
- c. 30 secs
- d. 50 secs

10. Assume that you are entering a bock with Red/Yellow signal.For how long are you allowed maintain a speed of 90 km/h while within that block?

- a. Not for a single second. ATP will apply emergency brakes immediately
- b. For 10 secs
- c. For 15 secs
- d. For 20 secs

11. Assume you are run as a test subject what will happen when the simulation is suspended?

- a. The experiment will be over
- b. Screens will be blanked and questions relating to the state of the train are asked
- c. None of the above can happen
- d. Both a and b can happen

12. Assume that while you are driving the train you are entering a block which has a Red/Yellow signal while the incab signal shows Yellow for that block. What the maximum allowable speed at that block?

- a. 80 km/h
- b. 150 km/h
- c. 0 km/h. Stop the train and immediately contact the CTC requesting directions
- d. 50 km/h

13. Assume that you are driving the train using the automatic control mode and the next block has a civil speed limit. What will you do?

a. Nothing. Leave the automation handle the situation.

b. Observe down the track to see whether there is an obstruction. If so you assume manual control for regulating the speed of the train or apply emergency brakes.

c. Assume manual control immediately

d. Slow down the speed of the train and notify the CTC requesting further instructions

14. How far will the vehicle have traveled before a brake failure is automatically reset?

a. 800 meters

b. 900 metes

c. 1000 meters

d. None of the above

15. How many brake pressure tanks are available in the simulation system?

a. one

b. two

c. None. The system uses regenerative braking.

d. None of the above.

16. How far can the operator see down the track in the simulation?

a. More than half a kilometer but less that a kilometer

b. More than a kilometer but less that a 2 kilometers

c. Depends on the position of the vehicle

d. Depends on the weather conditions (fog, snow)

17. From within which mode you can slow down the train (neglect programmed stop mode)

a. Manual

b. Cruise control

c. Autopilot

d. All of the above

Appendix C

Subject Consent Form

The U. S. Department of Health and Human Services requires that all persons used as subjects in experiments sign a consent agreement.

The procedures to be followed in our experiments involve making observations from a computer or related displays, making decisions and communicating these by mechanical or verbal means to be provided and explained to you in detail. These experiments do not, in our judgment, pose any risks or hazards to your health or well-being. You are free to ask any questions and have them answered to your satisfaction, and are free to withdraw consent and discontinue participating at any time without prejudice.

I understand that I may contact the Chairman of the Committee on the Use of Humans as Experimental Subjects, Dr. H. Walter Hones, E23-425, MIT (tel. 253-1772), if I feel I have been treated unfairly as a subject, and that further information may be obtained by calling the MIT Insurance and Legal Affairs Office, 4-104 (tel. 253-2822).

I consent to be a subject in the MIT Human-Machine Systems Laboratory under the above stated conditions.

(name)	(signature)	(date)

Appendix D

Experiment Design

Below are presented the designated positions along the track at which emergencies were set to occur. To make the notation more compact we use the codes: m for motor failures, b for bearing failures and o for obstructions.

Motor failure codes	Distance from last station (km)	Direction of travel	Time block
1	13.25	Eastbound	1st
2	40.58	Eastbound	2nd
3	30.52	Westbound	3rd

Bearing failure codes	Distance from last station (km)	Direction of travel	Time block
1	33.28	Eastbound	1st
2	3.06	Westbound	2nd
3	26.05	Westbound	3rd

grading crossing codes	Distance from last station (km)	Direction of travel
1	1.523	Eastbound
2	21.451	Eastbound
3	22.726	Eastbound
4	22.607	Westbound
5	26.549	Westbound
6	48.181	Westbound

The next appendix goes in detail to present which scenario each subject went through and what its reaction was.

Appendix E

Subject Emergency Scenarios and Data

Subject 1

Control mode	Emergency	Response	Reaction
Automatic	m1	hit	8.205 secs
	b3	miss	-
	o6	No collision	Stop 91.22 m
Cruise	m2	hit	2.28 secs
	b3	hit	5.595 secs
	o5	No collision	Stop 80.14 m
Manual	b1	hit	8.22 secs
	b2	hit	15 secs
	m3	hit	1.981 secs
	o4	No collision	Stop 185.07 m

Subject 2

Control mode	Emergency	Response	Reaction
Automatic	b1	hit	7.206 secs
	b2	hit	13.194 secs
	m3	hit	3
	o1	No collision	Stop 66.37 m
Cruise	b2	miss	-
	m3	hit	3.1
	o3	No collision	Stop 89.21 m
Manual	m1	hit	9.48
	b3	miss	-
	o6	No collision	Stop 79.6 m

Subject 3

Control mode	Emergency	Response	Reaction
Automatic	b2	hit	9.86 secs
	m3	hit	8.7 secs
	o3	No collision	Stop 140 m
Cruise	m1	hit	5.45 secs
	b3	hit	13.1 secs
	o6	No collision	Stop 26.45 m
Manual	b1	hit	12.85 secs
	m2	miss	-
	b3	hit	15.7 secs
	o1	No collision	196.77 m

Subject 4

Control mode	Emergency	Response	Reaction
Automatic	b2	miss	-
	m3	hit	7.38 secs
	o3	No collision	Stop 111.5 m
Cruise	m1	hit	2.945 secs
	m3	miss	-
	o6	No collision	Stop 76.17 m
Manual	b1	hit	6.364 secs
	m2	miss	-
	b3	miss	-

Subject 5

Control mode	Emergency	Response	Reaction
Automatic	b1	hit	12.98 secs
	m2	hit	3.54 secs
	o5	No collision	Stop 4.4 m
Cruise	m1	miss	-
	m3	hit	3.06 secs
	o6	No collision	Stop 3.71 m
Manual	b1	hit	3.46 secs
	m2	hit	3.66 secs
	b3	hit	17.6 secs
	04	No collision	Stop 94.78 m

Subject 6

Control mode	Emergency	Response	Reaction
Automatic	b1	hit	15.59 secs
	m2	miss	-
	b3	hit	15.74
	o5	No collision	Stop 38.67 m
Cruise	m2	hit	7.8 secs
	b1	miss	-
	o6	No collision	Stop 17 m
Manual	b1	hit	8.154 secs
	m2	hit	6.24 secs
	o1	No collision	Stop 70.09 m

Subject 7

Control mode	Emergency	Response	Reaction
Automatic	b1	hit	16.06 secs
	m2	hit	3.13 secs
	o5	No collision	Stop 27.18 m
Cruise	m1	hit	2.465 secs
	b2	hit	17.132 secs
	o6	No collision	87.19 m
Manual	b2	hit	16.87 secs
	b3	hit	19.68 secs
	o1	No collision	Stop 86.4 m

Subject 8

Control mode	Emergency	Response	Reaction
Automatic	m1	miss	-
	b2	hit	15.6 secs
	o5	No collision	Stop 56.48 m
Cruise	m1	miss	-
	b2	miss	-
	m3	hit	3.48 secs
	o6	No collision	171.1 m
Manual	m1	hit	10.8 secs
	m3	hit	7.2 secs
	o1	No collision	Stop 68.83 m

Subject 9

Control mode	Emergency	Response	Reaction
Automatic	m1	hit	3.545 secs
	b3	hit	17.405 secs
	o6	collision	-
Cruise	m2	hit	6.481 secs
	b3	miss	-
	o5	No collision	Stop 158 m
Manual	m1	miss	-
	b2	miss	-
	m3	hit	4.174 secs
	o1	No collision	Stop 132.66 m

Subject 10

Control mode	Emergency	Response	Reaction
Automatic	b2	hit	16.2 secs
	m3	hit	3 secs
Cruise	b1	hit	7.32 secs
	m2	miss	-
	b3	miss	-
	o5	collision	-
Manual	m1	hit	3.125 secs
	b2	hit	13.8 secs
	o6	No collision	Stop 163.66 m

Subject 11

Control mode	Emergency	Response	Reaction
Automatic	m1	hit	3.125 secs
	m3	miss	-
	o5	No collision	Stop 95.75 m
Cruise	b1	hit	8.288 secs
	m2	hit	4.51 secs
	b3	hit	11.04 secs
	o2	collision	Stop 62.79 m
Manual	m2	hit	3.361 secs
	m3	miss	-
	o4	No collision	Stop 125.44 m

Subject 12

Control mode	Emergency	Response	Reaction
Automatic	b1	hit	12.91 secs
	m2	hit	4.4 secs
	b3	hit	15.34 secs
	o4	No collision	Stop 134.88 m
Cruise	m1	hit	2.9 secs
	b2	hit	7.261 secs
	m3	miss	-
	o5	No collision	Stop at 142.92 m
Manual	b2	hit	8.4 secs
	o3	No collision	Stop 199.39 m