Investigating Information Use and Learning with ATIS: Simulation Description and Experimental Design

By

Kenneth Vaughn
Prasuna Reddy
Mohamed Abdel-Aty
Ryuichi Kitamura
Paul Jovanis

UCD-ITS-RR-94-7

April 1994

Funding for this study from
Partners for Advanced Transit and Highway (PATH)
and
California Department of Transportation (Caltrans)

Institute of Transportation Studies
University of California, Davis
Davis, California 95616
(916) 752-6548 Fax: (916) 752-6572
http://www.engr.ucdavis.edu/~its

TABLE OF CONTENTS

Investigating Information Use and Learning with ATIS:

Simulation Description and Experimental Design

ABSTRACT

The application of advanced technology and its potential to alleviate a host of traffic related problems has spawned numerous research programs into the investigation of how advances in information technologies may be utilized to significantly alter traveler behavior. If real-time, accurate information on the characteristics of the travel environment can be provided to travelers prior to departure and while en-route, will behavior be altered in such a way as to improve the overall characteristics of the travel environment? In order to accurately model the macro-level effects of Advanced Traveler Information Systems (ATIS), the micro-level effects of these systems on individual driver behavior must be analyzed and understood. The use of computer simulation and modeling is being utilized as a data collection tool to study advanced traveler devices to supplement real operational systems and as a cost effective alternative to field studies (Bonsall, 1990, Adler, 1993, McNally, 1993). Researchers at the University of California at Davis are utilizing PC based computer simulation to study the effects of information on individuals' route choice behavior and learning. Previous research utilized a simplistic binary route choice simulation to collect sequential data on subjects decisions (Vaughn et al. 1993, Yang et al. 1993). Building on the efforts from this previous simulation, a new set of experiments utilizes an expanded traffic network and provides various levels of information content to the subjects. This framework allows for investigation of both pre-trip and en-route decision behavior and captures the effect of different levels of information on route choice and diversion. In this series of experiments further investigation of drivers learning and adaptive processes are also explored.

1.0 OVERVIEW

The simulation is an interactive PC program running on a DOS based platform. The screen display is composed of three main windows: a network window, an information window, and an instruction window. The network window displays a hypothetical traffic network, composed of three primary routes from an origin to a destination. The primary routes are composed of a freeway route and two arterial routes. These primary routes are cross connected with a series of surface streets creating a network of approximately 34 roadway links and 23 intersections (or potential decision points). The travel environment is generated by a stochastic assignment of travel speeds and stop delays to the network links and nodes. A random incident generator is used to assign an incident of random severity to the network for each travel day.

The information and instruction windows are used to simulate an in-vehicle information system. A structured in-laboratory simulation experiment has been designed utilizing different levels of information as experimental treatments. The information treatments under analysis include incident information, en-route guidance, pre-trip guidance and congestion level information. The simulation experiments described here were just completed at the time of this writing and took place in March of 1994 utilizing 100 commuters from the Sacramento California region as test subjects. This paper describes the simulator and the experimental design in detail as well as summarizes the testable hypothesis of the experiment. A future report will provide an analysis of the findings of this study.

A focus of the previous simulation experiments was to produce a data set which would allow the exploration of information accuracy requirements. It was found that subjects initially had a high compliance rate in the absence of any knowledge about the accuracy level of the information being provided, but that they could quickly discern the accuracy from experiences and in general tended to accept advice at a rate equivalent to the accuracy of the system (see Vaughn et al. 1993). In the binary route choice framework which was used, a particular piece of advice for a given travel day can only be correct or incorrect. This creates an overly simplistic world in which incorrect advice is easily identifiable. In the realistic route choice problem when multiple alternatives are available and en-route deviations allow for almost an unrestricted number of possible routes the notion of information or advice being "correct" or "incorrect" becomes very grey. In this infinitely complex environment certainly no information system will ever be 100 percent accurate. The question then becomes how accurate must a system of this type be in order for users to perceive it as being accurate, and how does their perception of accuracy relate to their perceived benefit from such information. In order to begin to understand the effects of the accuracy of information or the perception of accuracy, this set of experiments creates a framework in which subjects are provided with information which is not always 100 percent accurate. An hypothesis of these experiments is that subjecting drivers to a more complex travel environment will significantly reduce their ability to access the accuracy of information. If drivers cannot accurately perceive the quality of information which they receive, then the potential exists for the accuracy level to be either over or under estimated. The prior being beneficial to ATIS implementation and the later being detrimental. The accuracy level used in these experiments is 75% which was one of the accuracy levels investigated in the previous binary route choice simulations. This level of accuracy was purposely chosen to allow for comparisons between the decisions and accuracy perceptions between the previous simple travel choice environment and the more complex environment provided in this new set of experiments. An additional dimension in which the experimental design could have been expanded was to treat the accuracy level as one of the experimental treatments and to vary accuracy across subjects as was done in the previous experiments. The use of another variable as an experimental treatment would have significantly increased the sample size required to obtain the desired testable hypotheses. The definition of accuracy in the context of these simulation experiments and the experimental design is discussed in more detail in subsequent sections of this paper.

In this series of experiments further investigation of drivers learning and adaptive processes will be explored. Building on the efforts from the TRAFFIC.COM simulation (Vaughn et al. 1993), this new set of experiments will utilize an expanded traffic network and provide various levels of information content to the subjects. This framework will allow for investigation of both pre-trip and en-route decision behavior and will attempt to capture what effects different levels of information have on route choice and diversion.

The objectives of this study are to investigate the main and interaction effects of the incident information, en-route guidance information, pre-trip guidance information and congestion level information on drivers' route choice decisions, their perceptions of the usefulness of these information types, and how the different information types affect drivers' perceptions of accuracy. All of these effects take place in a dynamic learning environment in which drivers' perceptions and knowledge are updated over the sequential series of their experiences. A final goal of this study is to investigate and begin to understand these learning process.

2.0 SIMULATION

The simulation is an interactive PC program running on a DOS based platform. The screen display is composed of three main windows: a network window, an information window, and an instruction window. The program is designed to be as self explanatory as possible with built in self paced instructions. The program also has an experimentation phase where subjects are allowed to make preliminary trials on the system and request help until they reach a point of familiarity with the system and then can proceed to the actual simulation. All data described in the data collection section later in this paper is also collected in this experimentation phase including the total number of practice trials taken by each subject for comparative results. Also an interface is provided to allow the experimenter to set up the desired experimental conditions which will be in effect for each subject. In the previous experiments this consisted of a numbering scheme which was used to identify all the possible combinations of treatments. A similar scheme has been developed for this simulation and all data will be stored in a specific directory in a separate file for each subject identified by this number.

Subjects were randomly recruited from the Sacramento area using a telephone survey company. One hundred subjects were recruited and participated in the experiments and were selected to meet the following subject requirements:

- All subjects will be recruited at random from the Sacramento California vicinity and must be willing to travel to the University of California Davis to participate in the experiment (approximately 10 miles from Sacramento CBD to UC Davis).

- All subjects will be regular (i.e. full time) commuters who commute during the morning commute hours of 6:00 am to 10:00 am Monday through Friday.

- The subjects will be broken down into the following categories:

  80 solo drivers, 20 carpoolers
  50 males, 50 females
  50 with at least some college education, 50 no college
  50 who drive less than 10,000 miles/year (LO MILES), 50 who driver more than 10,000 miles/yr (HI MILES)
  25 age 16 to 25
  25 age 26 to 40
  25 age 41 to 55
  25 age 56 and over

  Carpoolers are either carpool drivers or passengers and must consider themselves to be regular carpoolers.

- The structure for these categories will be as outlined in Figures 1 and 2 below.

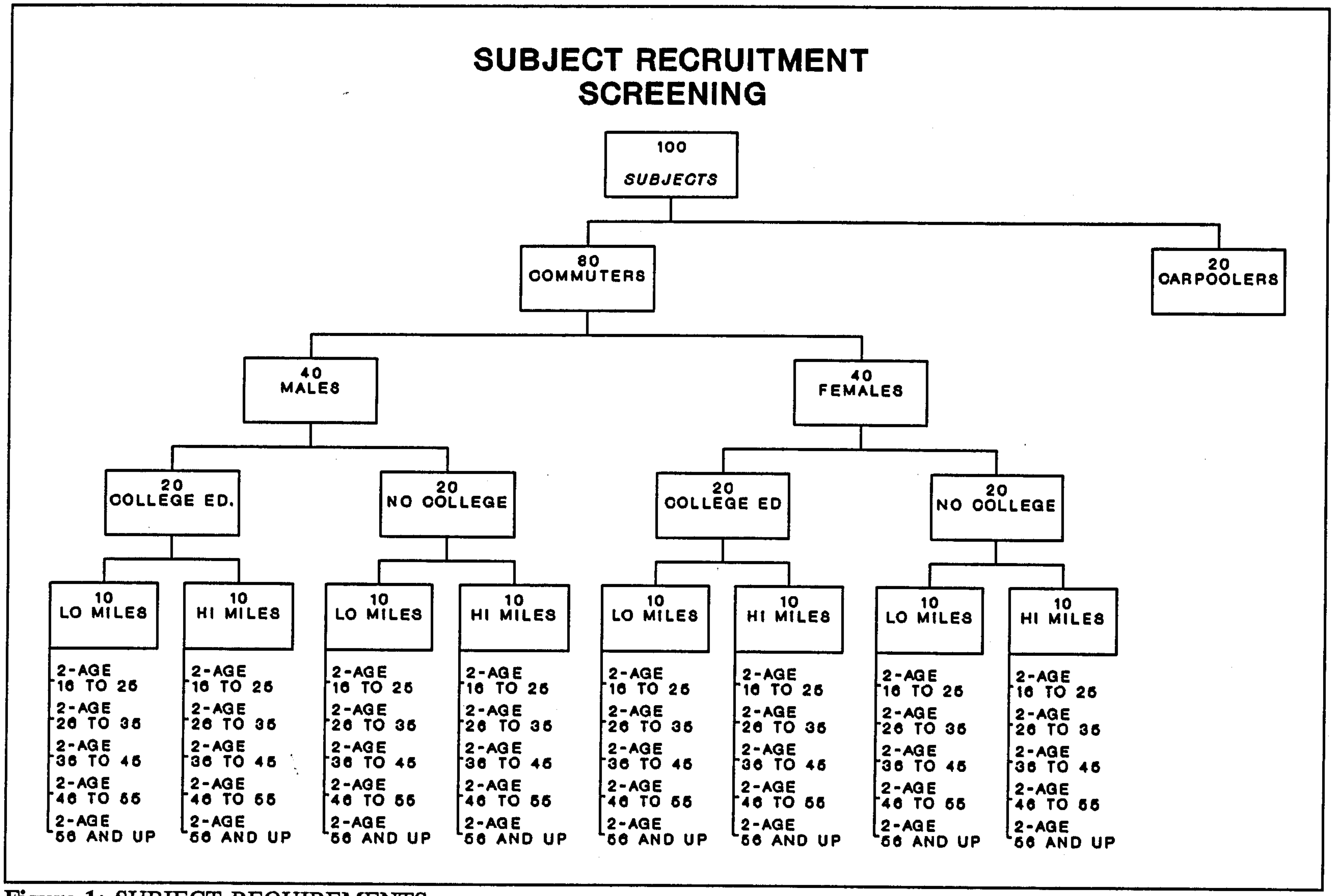

Figure 1: SUBJECT REQUIREMENTS

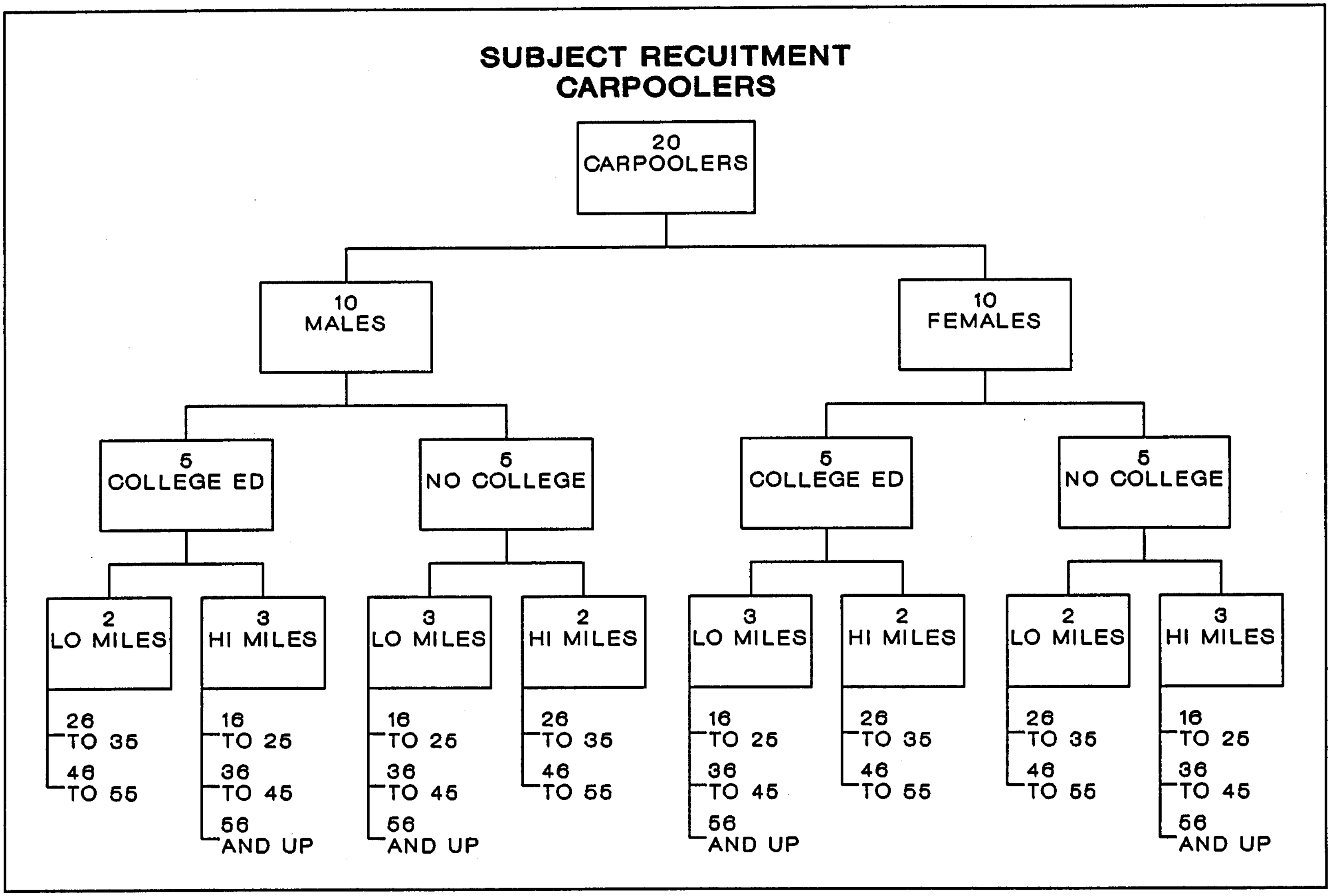

Figure 2: SUBJECT REQUIREMENTS CONTINUED

At the beginning of the simulation, subjects are asked several questions in the form of a computer prompted "mini-survey". The responses are automatically recorded in the subjects data file. The following is a list of the questions as asked in the simulation.

1. Approximately how many miles is it from your home to your usual work place?
   ___ miles.

2. What time do you normally leave home for work? ___

3. What time do you normally arrive at work? ___

4. What is your normal work start time? ___

5. Is your work start time fixed or is it flexible?
   1. Fixed.
   2. Flexible.

6. What is your primary method of travel to your usual work place?

   1. Drive alone (including motorcycle).
   2. Drive with one or more workers.
   3. Drive with one or more children.
   4. As a passenger in a car, truck or van.
   5. Other.

7. Do you listen to traffic reports before you leave home for work?

   1. no
   2. yes, if yes then

   7.1 How often do you listen to traffic reports before you leave home?

      1. Only every now and then, or on special occasions
      2. On some days
      3. Most days
      4. Every day, or nearly every day

   7.2 Approximately how many times per month do you change your route to work based on these reports?

   7.3 Approximately how many times per month do you change your departure time for work based on these reports?

7.4 Approximately how many times per month do you change your transportation mode for work based on these reports?

8. Do you listen to traffic reports while driving to work?

1. no
2. yes, if yes then

8.1 How often do you listen to traffic reports while driving?

1. Only every now and then, or on special occasions
2. On some days
3. Most days
4. Every day, or nearly every day

8.2 Approximately how many times per month do you change your route to work based on these reports?

9. How accurate is the information you receive from traffic reports?

1. Extremely accurate
2. Very accurate
3. Somewhat accurate
4. Not very accurate
5. Not at all accurate

10. Approximately how many miles do you drive in a year?

1. Less than 5,000 miles
2. 5001 to 10,000 miles
3. 10,001 to 15,000 miles
4. 15,001 to 20,000 miles
5. More than 20,000 miles

11. What is your age?

12. What is your gender?
1. Male
2. Female

13. Do you live in a...
1. Single family house
2. Apartment

3. Condominium or Townhouse
4. Duplex Unit
5. Mobile home
6. Hotel or Motel
7. Other

14. Do you own or rent?
    1. Own
    2. Rent
    3. Don't Know

15. Including yourself, how many people live in your household? Count all babies, relatives, roommates or others who regularly live with you.

    Total # of persons______.

16. Including yourself, how many of these household members are five years of age or older?

    # of persons 5 or older_____.

17. Including yourself, how many of these household members are sixteen years of age or older?

18. Including yourself, how many of these household members are employed?

19. Which of the following best describes your relationship in your household?
    1. spouse with children in the home
    2. spouse with no children
    3. adult single with children in the home
    4. adult single with no children
    5. co-habitant
    6. child (family member under 18 years old)
    7. other relative
    8. roommate
    9. friend
    10. other.

20. Which of the following best represents your education level?

    1. did not complete high school
    2. high school graduate
    3. completed some college, including 2 year (AA)
    4. college graduate

21. For statistical purposes, please indicate which income category best represents your last years total household income.

1. Less than $15,000
2. $15,000 to $25,000
3. $25,001 to $35,000
4. $35,001 to $45,000
5. $45,001 to $55,000
6. $55,001 to $75,000
7. $75,001 to $100,000
8. $100,001 to $150,000
9. More than $150,000

2.1 Network

The network window displays a hypothetical traffic network. The network is composed of three primary routes from an origin to a destination. The primary routes are composed of a freeway route and two arterial routes. These primary routes are cross connected with a series of surface streets creating a network of 34 roadway links and 23 intersections (or potential decision points) (see figure 3 for current configuration). The links running from node 2 to node 22 make up the freeway route, and the links running from nodes 3 to 23 and from 4 to 24 make up the two arterial routes. While this network is still simplistic in comparison to the real traffic environment it is a considerable step up in complexity from the two-link, two-node traffic network utilized in the first set of learning experiments and may adequately represent routing alternatives which are available to many commuters.

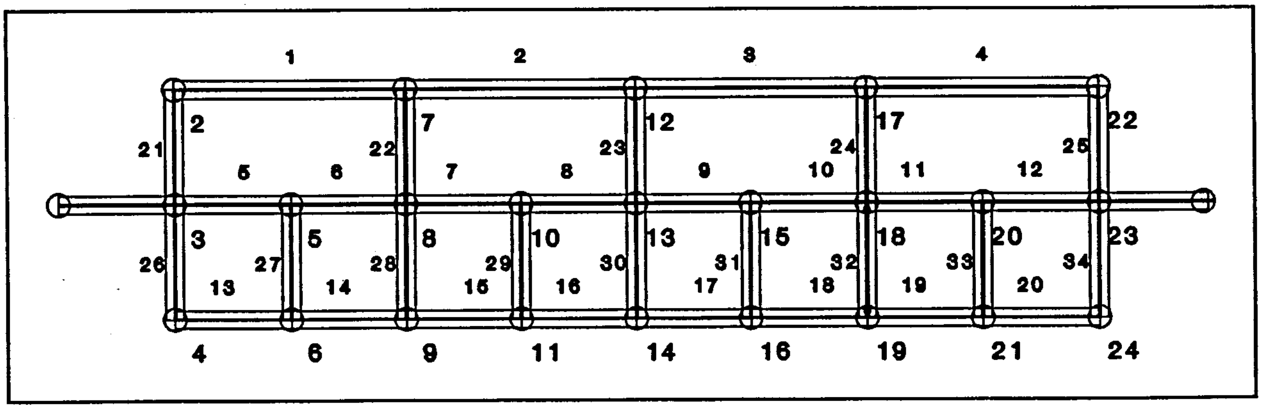

Figure 3: Simulation Network

The network window is animated in that the simulated vehicle movements will take place within this window. A simulated vehicle (cursor) will move through the network in response to decision inputs by the subject. Control commands will be input via the key board to indicate desired turning movements. The simulation is currently designed with a 1:30 time scale (1 minute real time = 2 seconds simulated time).

Consider one of the freeway links (there are 4 total). The program has been configured such that each freeway link is 10 characters in length. The movement is generated by having the cursor move from character to character along the link. The time the cursor stays at each character location is determined by the delay setting of the link. This delay setting is currently in 1/1000 second increments. Thus if the delay assigned on a freeway link is 1000 units, then the cursor will move from character to character at a rate of once per second. If the delay assigned is 2000 units, then the cursor will move from character to character at a rate of once every two seconds or at half the speed. Thus the simulated speed of the cursor is inversely proportional to the delay assigned to the link. Assuming one freeway link represents 5 miles this represents a 20 mile freeway segment and each character position represents .5 miles. In the real world, assuming a travel speed of 30 mph, it would take 1 min to traverse .5 miles thus with a 1:30 time scale, it takes 2 seconds to traverse the .5 mile character position. Also, if the overall average freeway route speed were 30 mph, it would take 1 minute and 20 seconds to traverse the 4 freeway links.

A simple formula converts the link speeds to segment delay values:

link delay =character length (miles)*time scale*60s/min*60min/hr * 1000 / speed(in mph).
=60000/speed (mph).

For example if a freeway link speed is 50 mph, then the delay assignment is 1200 or 1.2 seconds per character movement and the 5 mile link would be traversed in 12 seconds.

The intent within this simulation is to present the subjects with a sequence of travel days. To accomplish this a series of network characteristics which will be the basis for the daily travel experiences encountered by the subjects have been generated. Additionally, because the focus is on the effects of the information, we also want to filter out as much variability due to the network as possible. To accomplish this, the sequentially generated travel days are dynamic across the sequence, but static from subject to subject. Thus, the travel conditions experienced by subject #1 on travel day 6 (for example) will be the same as the travel conditions experienced by any other subject on travel day 6. Some differences will be experienced which are choice-based, in that with this more complex network for a given travel day, subjects may not traverse (and therefore experience) the exact same segments of the network, or may traverse them in a different sequence, and therefore may have a fundamentally different experience.

2.2 Link Delay

The simulated network characteristics are pre-generated and stored in a network data file. This data file will contain all of the network characteristics identified by travel day and link or node number. The simulation program will then simply read this data file to create the travel environment for each day. A subroutine has been developed to create and write this file. An example of the file content is shown in Table 1 below. The primary network characteristic will be delay. The delay will be of two forms, congestion delay experienced on a link, and stop delay experienced at nodes or intersections. Additionally, the congestion delay will be of two types, pure congestion, and incident caused congestion.

Table 1

Variables	Description
Link1 to Link34	Delay assignments for links 1 to 34
Node2 to Node24	Stop delay assignments for nodes 2 to 24
Incident Link	Link number on which random incident has occurred
Severity	Severity of incident (either moderate or severe)
Incident type	Random assignment of description of incident

On roadway links, the delay assigned will be inversely proportional to the travel speed on that link for that day. When an incident has been assigned to occur on a particular link, the delay assigned to that link will be elevated due to the incident.

The use of the normal distribution is suggested (A.D. May, 1990) to represent speed distributions on roadway links. For this simulation the network characteristic subroutine creates normally distributed link speeds for each link in the network. These link speed assignments are established independent of each other in the current configuration of the simulation program. It is assumed that the freeway is congested 40 percent of the time and in the simulation, the freeway is randomly assigned to be in a congested state at this level. The freeway links are N(μ=35mph, σ=7mph) on "congested" days and N(μ=45mph, σ=5mph) on "uncongested" days, 1st street is N(μ=35mph, σ=5mph) and 2nd street is N(μ=40mph, σ=2mph) and side streets are N(μ=25mph, σ=1mph). This creates a scenario within the network where the three primary routes through the network have the similar mean speeds on over the sequence of trials, but the variance in travel speed is much greater on the freeway than on the arterials and the surface streets have lower mean speeds, but experience little variation. For the normal distribution, approximately 95% of the observations will be between $\pm 2\sigma$. This means that typical speeds will range from 15 mph to 55 mph on freeway links, 23 mph to 47 mph on arterials and from 23 mph to 27 mph on surface streets. The speed distributions have been right truncated at the free flow or design speed of the links such that the maximum speed is 55 mph on freeway links, 45 mph on arterials and 30 mph on surface streets. The subroutine converts the speed assignments to units of delay which the program uses to simulate movement in the network.

2.3 Incident Delay

For this simulation it will be assumed that it is not unreasonable to expect at least one incident to occur within the network on each simulated day. It is also reasonable to assume that incidents are more likely to occur on the freeway and arterial links than on the surface streets. For this simulation, we will assume that the probability of an accident occurring on a given travel day is 1.0 and that the probability of the accident being on the freeway or arterial links is 4/5 and the probability of being on the surface streets is 1/5. This creates a situation in which there is always an incident of varying intensity some where on the network for any given day. Then,

P(incident on freeway route | it is on freeway or arterial) = 2/3
P(incident on arterial route 1 or 2 | it is on freeway or arterial) = 1/3

i.e. twice as likely to occur on the freeway as opposed to the arterials.

then,

P(incident is on a particular freeway link | incident on freeway route) = 1/(number of freeway links) = 1/4

P(incident is on a particular arterial link | incident on arterial route 1 or 2) = 1/(number of arterial links) = 1/16

P(incident is on a particular surface link | incident is on surface streets) = 1/(number of surface links) = 1/14

Then,

Total probability of the incident occurring on a specific freeway link = 1/4 * 2/3 * 4/5 = 2/15.

Total probability of the incident occurring on a specific arterial link = 1/16 * 1/3 * 4/5 = 1/60.

Total probability of the incident occurring on a specific surface link = 1/14 * 1/5 = 1/70.

Incidents occurring on the network are randomly assigned as being either moderate or severe in nature. Links with incidents assigned have their link speed assignments reset to 15 mph for severe incidents and 20 mph for moderate incidents. For each incident one of three incident types is also randomly assigned. Moderate incidents are assigned as either Accident, Stalled Vehicle, or Maintenance while severe incidents are assigned as Injury Accident, Truck Accident, or Maintenance Lane Closed. The structure of the program also allows for assignment of incidents to intersection nodes which would then effect delay assignments on multiple links. This option is not utilized in this series of experiments and all incidents and their effects are confined to a specific link.

2.4 Stop Delay

From the first set of learning experiments, the effects of stop delay had significant effects on driver behavior (Vaughn et al. 1993). Stop delay is the amount of time spent stopped at an intersection due to traffic signals or stop signs. In the real driving environment, this delay also includes the additional time loss due the deceleration and acceleration required to make the stop. It is hypothesized that drivers perceive the stop delay as greater than it actually is or that they assign some other level of dis-utility to the process of stopping and that given two alternative ceteris paribus routes drivers would prefer a route with no stops to one with stops. Evidence in support of this hypothesis was found in the previous experiments where the compliance with advice was significantly lower in conditions where subjects experienced stop delay primarily due to subjects rejection of advice to take the route with stops. These results suggest significant behavioral differences in response to stop delay versus congestion (or moving delay). Within this simulation we will again incorporate the effects of stop delay in a more realistic manner. Additionally, because this is a simulated network, stops become the only route (or link) specific attribute (besides speed or delay).

Stop delay will occur as the result of either stop signs or signalized intersections. In figure 3, nodes 8, 9, 13, 14, 18, and 19 are signalized intersections, nodes 5, 6, 10, 11, 15, 16, 20 and 21 have stop signs only on the surface street approaches. Nodes 2, 7, 12, 17 and 22 represent freeway on/off ramps and are not assigned stop delay in the simulation. At stop sign locations the vehicle tracking cursor will stop for an appropriate amount of time, but at signals, stops are only required when the light is red. Stop signs have been assigned a delay value of 2 seconds for right turn movements and 3 seconds for left turn movements giving a greater penalty for the left turns. Within the simulation a 50% probability of the light being encountered as red is used for every signal encountered . In this scenario, stop delay is mandatory at all stop signs, and stops at signals will be encountered approximately 50% of the time when traversing the network.

As currently envisioned, the network characteristic data file would be generated once for each trial day reflecting a static information situation. To move to a dynamic information system, several files (say 4) could be generated for each trial day then the initial information provided would come from the first file and at some fixed time into the trial, the information could be refreshed or updated from the second table, likewise for a third table and so on for what ever seems appropriate depending on the total travel time in the network and a reasonable refresh rate. Within a scenario of this type we may want to update everything except incident information and assume that the incident durations will be longer than the total travel time in the network (this doesn't seem unreasonable as the network is not that large but it may also be interesting to see the effects of clearing an incident).

3.0 OBJECT ORIENTED PROGRAMMING DESIGN

The simulator design includes the following steps: requirement analysis, database design (data input and data collection), specifications of user-computer interface, design of shortest path module, software design, testing the prototype, and prototype refinement based on user feedback. The simulator was developed using C++ programming language which includes all object oriented features. This includes inheritance, class hierarchy (objects), message passing and concurrence. Each node in the network is a self-contained 'objects'. Whenever the subject reaches the node, a node class is 'instantiated' from a uniform and basic class structure.

In the experiment, we simulated driver's route choice in a network using object orient design. An introduction to object orient design can be found in most books related to programming languages. The continuous event simulation will provide data to analyze the drivers behavior using level of information content.

The information level in the simulation is pre-defined for each subject. The basic unit of time in the simulation is a simulated minute. An object oriented solution to a problem requires that we map the problem space into a set of objects and operations in the solution space. The major components of the problem space are explained below. The problem space includes the following entities.

The network entities are links, nodes, clock time, signals and stops, and cursor (driver's vehicle). The other entities are information window, instruction window, signal display and navigating arrow. These entities are defined as friends class to most of the classes in the simulation. Each entity will be mapped to a class of abstract objects in the solution space. The class construct in C++ supports the encapsulation of data and the decomposition of problems into objects and operations. We discuss each of the classes and communication with other objects that make up the object-oriented solution space for this simulation.

Class 'Main' has four subclasses which are; demonstration, trial runs, network and best info. The 'Demonstration' class has a single method of demonstrating how to use the simulation. The 'Trial runs' will allow user to make as many trial runs as he want. This trial run is a private method. It has a private data number trial each subject made. The 'Network' class is a parent class for all simulation classes. Class 'Network' has private data consisting of number of nodes, links, decision points, origin and destination. It also has private methods title screen, survey details, feed back and end questions. The public methods that are defined for this class are, a constructor that initialize all the classes in the simulation, a data reader from the file and output a data file. The class 'Best info' has a private method for evaluating different information systems and private data how subjects rank the information.

The cursor (driver or subject) class contains the important feature associated with each subject his or her arrival time in the system and at each node arrival and departure time and decision time. These time datums are part of the private section of the class. This private section also has x, y coordinates of present location. In the public section, this class stores the keyboard input which is accessed by many classes. We designated the shortest path, subject-path, event, navigation classes to be a friend of class cursor so that objects of these classes can access private data class cursor. This is the only class which has no methods.

The class 'Shortest path' contains the following private data. A static array of links, a static array of nodes, subjects new location (node), previous node, previous link, total number of links, total number of nodes, least travel time, and cumulative delay at nodes. The private methods are finding next node, shortest path, and total distance. The class 'Time' contains the private datum current time (computer clock time), methods get-time, set-time, and time conversion. The class 'Signal' contains an integer, available, that has the value 1 if the signal is red, otherwise the value '0'. A real value delay due to signal represents that signal is red. It has total number of red signals and green signals as integer values. This class has only method, it delays if the signal is red. Class 'Stop sign' has all the data and methods as signal class expect on or off data.

The class 'Subject path' has private data total number of links (at the beginning), current number of links, total travel time, cumulative travel time, cumulative delay at nodes, cumulative delay on links, reaction time at each node, number of decision points, number of acceptance and not accepted points. This class has methods for counting number of links, decision points and total reaction time. The 'Congestion display' class has private data array of links, link delays and colors. It also has a method of painting the network in different colors.

The class 'Navigation' has two subclasses Navigation by arrow and Navigation by text. These two have the same data and methods. But navigation by arrow method will display and animate arrow (graphics mode) and navigation by text display and animate text (text mode). These classes have private data of the next node and direction and have methods for finding next node, direction and converting direction to display format. The 'Pre-trip' class has private data shortest path inherited from shortest path class stored in an array. This also has private method display shortest path and delete display of shortest path.

The 'Accident' class has the private data location co-ordinates, delay and link congestion level and private methods delay, accident display and accident identifier. The accident display will indicate accident when a subject makes his first decision, if he is assigned to that information configuration. If there is no accident information in his configuration file then identifier function will display accident when he reaches a specific range from the accident. This phenomenon is included to simulate realistic conditions. If a person does not have any information about an accident, but he reaches the accident location, he will observe the accident and understand the reason for the delay in his travel time.

This ability to sense the type of the object is crucial and enables object-oriented programming language to reduce the amount of programming necessary in large software applications.

4.0 EXPERIMENTAL DESIGN

Within this controlled simulated travel environment we wish to apply experimental treatments consisting of various types or levels of information. In addition, we may also wish to consider several blocking factors such as gender, age, driving experience, and education. As an example suppose we wish to consider 7 treatments each with two treatment levels. If enough subjects are available then a complete factorial experiment could be performed. This design is known as a 2^7 factorial design and would require 128 subjects for single replication. In cases such as ours where a large number of treatment combinations are being considered, fractional factorial designs are often used to limit the data requirements to a reasonable level. The disadvantage of the fractional factorial design is that not all factor level and interaction effects are estimable. By fractioning the design, certain factor effects and interactions will be aliased or confounded with each other resulting in the inseparability of these effects. It is however common that in experiments with high order interaction that the effects of these higher order interactions are not significant (Montgomery, D.C. 1991). Thus the goal in a fractional factorial design is to try and alias the effects of interesting interactions (typically the main effects and two-way interactions) with higher order interaction terms which can reasonably be assumed to be insignificant.

Back to the example, if we perform a 1/8 replicate of the 2^7 design we require only 16 subjects and this is considered a 2^{7-3} fractional factorial. With this design, all main factor effects are aliased with three-way or higher interactions and each two-way interaction is aliased with two two-way interactions. With this design if it is reasonable to assume that all three-way and higher interactions are insignificant, then all of the main effects are estimable and 7 of the 21 two-way interactions are estimable if effects of the 14 aliased two-way interactions can also be assumed to be negligible.

Factorial experiments are typically used in production process environments where the variability between the units of analysis are generally small. In behavioral experiments where the experimental units are people, individual difference can be great due to differences in background, education, life experiences, etc. In this case, between subject variability can become large resulting in difficulty in detecting real differences between treatments. To deal with this problem, often repeated observations are required for each treatment combination. Assuming we can reasonable recruit and test at least 100 subjects, then the 2^{7-3} fractional factorial design would allow for 6 subjects within each treatment combination.

4.1 Experimental Treatments

The experimental design selected for this simulation study is a 1/4 fraction of a 2^7 factorial design. With two levels for each factor, a full factorial design would require a minimum of $2^7 = 128$ runs. A one-quarter fraction of this design can be used requiring a minimum of only 32 runs and if 3-way and higher interactions are assumed to be negligible, then all of the main effects are estimable and 15 of the 21 2-way interactions are also estimable. With three subjects per design block, 96 subjects are required.

The design matrix for this one-quarter fractional factorial design is shown in table 1 below. The term Run refers to one completed experimental trial. The letters A through G refer to the experimental treatments. Since all treatments have two levels, a "-" in the table refers to the low or off value of the treatment and a "+" in the table refers to the high or on level of the treatment. For example, for run 1 in the design table, treatments A through E are in the low level or are off while F and G treatments are in the high level or are turned on, this gives treatment fg in the Treatment column. this experimental design will provide estimates of the following effects:

Estimable effects:

Main effects: A,B,C,D,E,F,G

Two-way interactions: AB,AC,AD,AE,AF,AG,BC,BD,BE,BF,BG,CD,DE,DF,DG

Aliased two-way interactions:

CE=FG, CF=EG, CG=EF

The simulation will apply four information treatments and will use 3 blocking factors to make up the seven experimental treatments. The information treatments are labeled A through D and the blocking factors are E through G. All treatments have two levels and are described below:

A. Incident with description. Red Icon displayed at the location of a severe incident, yellow Icon displayed at the location of a moderate incident. Also in the information window, display textually the location and classification of the incident for example: "Sever injury accident on First Street between F st. and G st." (+=on, -=off)

B. En-route Guidance. Graphical arrows indicating advised turning movements and textual description of advice. (+=on, -=off) At every node, the information system provides turning movement recommendations for the next node in the form of a blinking arrow for turn left, turn right or continue straight ahead. Also, text information is provided such as "go north on B street". The turning movement advice is based on the computed minimum path from the current cursor position to the destination.

C. Pre-trip Guidance. Minimum path displayed at beginning of trip. (+=on, -=off) At the start of the trail the initial calculated minimum path from the origin to the destination is outlined on the network and remains until the subject make their first decision and then is turned off as they under way.

D. Congestion Information. Color coded links for moderate and severely congested links with green indicating normal congestion, yellow indicating moderate congestion and red indicating severe congestion. (+=on, -=off)

Three blocking factors:

E. Gender, +=male/-=female.

F. Age, +=young/-=old.

G. Education, +=high/-=low.

A 2^{7-2} fractional factorial design

Run	A	B	C	D	E	F	G	Treatment
1	−	−	−	−	−	+	+	fg
2	+	−	−	−	−	−	−	a
3	−	+	−	−	−	−	−	b
4	+	+	−	−	−	+	+	abfg
5	−	−	+	−	−	−	+	cg
6	+	−	+	−	−	+	−	acf
7	−	+	+	−	−	+	−	bcf
8	+	+	+	−	−	−	+	abcg
9	−	−	−	+	−	−	−	d
10	+	−	−	+	−	+	+	adfg
11	−	+	−	+	−	+	+	bdfg
12	+	+	−	+	−	−	−	abd
13	−	−	+	+	−	+	−	cdf
14	+	−	+	+	−	−	+	acdg
15	−	+	+	+	−	−	+	bcdg
16	+	+	+	+	−	+	−	abcdf
17	−	−	−	−	+	+	−	ef
18	+	−	−	−	+	−	+	aeg
19	−	+	−	−	+	−	+	beg
20	+	+	−	−	+	+	−	abef
21	−	−	+	−	+	−	−	ce
22	+	−	+	−	+	+	+	acefg
23	−	+	+	−	+	+	+	bcefg
24	+	+	+	−	+	−	−	abce
25	−	−	−	+	+	−	+	deg
26	+	−	−	+	+	+	−	adef
27	−	+	−	+	+	+	−	bdef
28	+	+	−	+	+	−	+	abdeg
29	−	−	+	+	+	+	+	cdefg
30	+	−	+	+	+	−	−	acde
31	−	+	+	+	+	−	−	bcde
32	+	+	+	+	+	+	+	abcdefg

5.0 INFORMATION ACCURACY

In order to investigate the effect of accuracy on the decision and learning process, the information provided within the simulation will not always be 100 per cent accurate. This will allow us to investigate how subjects respond to receiving inaccurate information and what effect this may have on future decisions. The information content being provided to subjects is basically of three types, incident, guidance/advice, and congestion levels. It is a reasonable assumption that providing very accurate guidance/advice and congestion levels is much more difficult than providing incident location information. Guidance requires accurate estimates of link travel conditions and errors are compounded across the network and congestion can build up rapidly or dissipate between information updates. Incidents are static in nature and are more easily identified and typically remain in the network for a significant length of time.

While the effects of a particular incident may be difficult to predict, its location should be accurate. Within the simulation, the locational information of incidents will be provided at 100% accuracy (i.e. always correct). Route guidance/advice and congestion information will be provided at a level of 75% accuracy. This definition means on 75% of the trial days, the guidance/advice or the congestion information provided to the subject will be accurate, but on 25% of the trial days, it will be inaccurate. Out of every four trial days, 1 day was randomly selected to be the day in which information would be inaccurate creating a balanced level of accuracy across the 20 commute days.

6.0 DATA COLLECTION

Data will automatically be recorded as one record for each trial day with 20 records per subject.

Initial data items:

Subject number
All data items from pre-survey
All link and node attributes (these already exist, just re-write to this file)
Treatment combination

At each node:

Record the decision made at each node:
- 0=no decision, node not traversed
- 1=straight ahead
- 2=right turn
- 3=left turn

Record the optimal decision for each node:
- 1=straight ahead
- 2=right turn
- 3=left turn

System advice
- 0=freeway
- 1=other roads

Drivers' decision
- 0=freeway
- 1=other roads

Decision time at each node (seconds)
Cumulative freeway/arterial/surface street distances at each node
Cumulative travel time
Amount of cumulative stop time

Each Day (trial)

Trial Number
Total travel time
Best possible travel time
Performance measure (travel time - best possible)
Number of stops on route
Number of stops on optimal route
Amount of stop delay on route

Amount of stop delay on optimal route
Total freeway/arterial/surface street distances used
Total freeway/arterial/surface street distances on optimal route
Accuracy of information (1/0)

Summary Questions: The following questions were asked at end of each trial day.

1. Do you feel the information you received today was:

   1. Correct
   2. Probably Correct
   3. Don't Know
   4. Probably Incorrect
   5. Incorrect

2. How important was the information you received in determining the route which you followed today?

   1. Very important, basically determined my route.
   2. Of some importance, strongly influenced my route selection.
   3. Moderate importance, had some effect on my selection.
   4. Of little importance, weakly influenced my route selection.
   5. Had no effect on my route selection at all.

3. Do you think the route which you followed today was:

   1. Faster than the other routes which were available today.
   2. About the same as other routes.
   3. Slower than other routes which were available today.

4. Do you think the route which you followed today was:

   1. Faster than the route which you followed yesterday.

2. About the same as the route you followed yesterday.
3. Slower than the route which you followed yesterday.

Each Subject

The following questions were asked after completion of the simulation.

1. Do you feel that the information you received during this simulation was:

   1. Extremely Accurate
   2. Frequently Accurate
   3. Moderately Accurate
   4. Frequently Inaccurate
   5. Completely Inaccurate

2. In general, how important was the information you received in determining which route you followed during the simulation?

   1. Very important.
   2. Of some importance.
   3. Moderate importance.
   4. Of little importance.
   5. Had no effect on my route selection at all.

3. If a traffic information system, similar to what you experienced in this simulation, were available on the market today how likely would you be to purchase such a system?

   1. Very likely
   2. Likely
   3. Undecided
   4. Unlikely
   5. Very unlikely

4. If you were to buy such a system, how much would you be willing to pay?

   1. less than $200
   2. $200 to $400
   3. $400 to $600
   4. $600 to $800
   5. $800 to $1000
   6. $1000 to $1200
   7. more than $1200

5. If such a system were only available as a monthly service, how much would you be willing to pay to subscribe to such a service?

   1. less than $10 per month
   2. $10 to $20 per month
   3. $20 to $30 per month
   4. $30 to $50 per month
   5. $50 to $75 per month
   6. $75 to $100 per month
   7. more than $100 per month

7.0 SUMMARY

This paper has described the development of a traffic network and travel simulator. This simulator is also capable of simulating various types of information systems for use by simulation participants. We have also described the experimental design used to collect sequential travel data in a recently completed set of simulation experiments. The strength of this simulator is that it is independent of the experimental design, that is to say the simulator was not designed to support a specific experimental design but the converse is true, the experiment was designed around the simulator. Due to this developmental approach, the simulation software is easily adaptable and therefore can support different experimental frameworks. The network can be easily modified simply by expanding the link/node structure. The location of traffic signals and stop signs, turning restrictions and the hierarchy of link types can all be easily manipulated. The location of origin and destination points within the network can also be modified and required stop locations between origin and destination can also be added to allow for trip chaining types of analysis, route choice under constraints and analysis of non-work related trips.

The first set of experiments utilizing this software have been successfully completed the results of which will be forthcoming in a future paper.

ACKNOWLEDGEMENTS:

This research has been funded by the California Department of Transportation (Caltrans) and the Partners for Advanced Transit and Highways (PATH). We would also like to acknowledge the staff and researchers at the Institute who contributed their time during the pre-testing of the simulation. The views expressed in this paper are those of the authors and do not necessarily represent the views of the funding agencies.

REFERENCES

Adler, J.L., McNally, M.G., and Recker, W.W., "Interactive Simulation for Modeling Dynamic Driver Behavior in Response to ATIS", , Proc. Fifth International Conference Computing in Civil and Building Engineering, ASCE 1993.

Bonsall, P. and Parry, T., "A Computer Simulation Game to Determine Drivers' Reactions to Route Guidance Advice", Proc 18th PTRC Summer Annual Meeting, London, 1990.

Chen, P., Mahamassani, H.S., "A Dynamic Interactive Simulator for the study of Commuter Behavior under Real-Time Traffic Information Supply Strategies", Presented at the 72nd annual meeting of the Transportation Research Board, Washington D.C. 1993.

May, A.D., "Traffic Flow Fundamentals", Prentice Hall, 1990.

Montgomery, D.C., "Design and Analysis of Experiments" Third Edition, John Wiley and Sons, 1991.

Vaughn K.M., Abdel-Aty M.A., Kitamura R., and Jovanis P.P., "Experimental Analysis and Modeling of Sequential Route Choice Under ATIS in a Simplistic Traffic Network", in Transportation Research Record 1408, Transportation Research Board, Washington D.C., 1993.

Yang H., Kitamura R., Jovanis P.P., Vaughn K.M., and Abdel-Aty M.A., "Exploration of Route Choice Behavior with Advanced Traveler Information using Neural Network Concepts", Transportation, Kluwer Academic Publishers, 1993.

APPENDIX

Driving Simulation Instructions

Background:

In this simulation, you will be presented with 20 commute days and will be asked to make route choice decisions in order to get from home to work. You will have three main routes from which to choose:

1. Freeway 80. The freeway has the advantages of a faster top speed and no stops or intersections but has the disadvantages of heavier traffic congestion and a greater occurrence of traffic incidents (accidents, maintenance etc.).

2. 1st Street. This is a major street and has the disadvantages of slower top speeds, stop lights and intersections, but is less likely to have traffic incidents and may be faster than the freeway route on days when the freeway is heavily congested or when an incident has occurred on the freeway.

3. 2nd Street. This is another major street which runs parallel to both Freeway 80 and 1st Street. Like 1st Street, 2nd Street has a top speed which is slower than the freeway and also has stop lights and intersections. The 2nd Street does have a faster top speed than 1st Street, but the 2nd Street route is longer (in distance) than the 1st Street route.

In addition, these main routes are interconnected by short surface streets which will allow you to change your route during your trip in order to avoid traffic congestion or traffic incidents. The traffic network used in the simulation is shown in Figure 1 below.

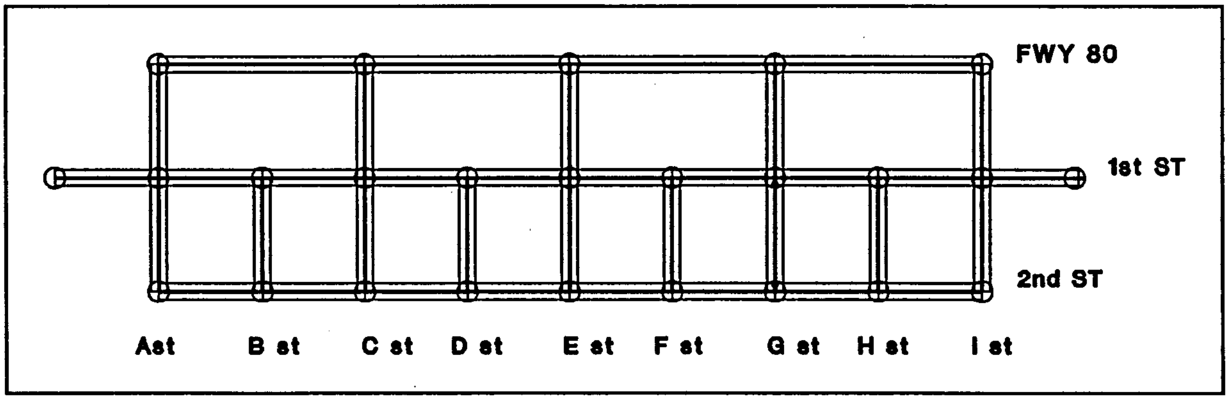

Figure 4: Traffic Network

In the simulation your vehicle will be represented by a moving green curser which you will "drive" through the traffic network by using the arrow keys on the computer keyboard. Your goal is to get from home to work as quickly as possible. Your score in the simulation will be the time it takes you to get from home to work, added over all 20 commute days. The lower this total travel time, the better you have performed.

In order to reduce your travel time, your vehicle has been equipped with a new Traffic Information System. The types of information you will receive include:

1. Traffic incident information with description. A color coded Icon or symbol will be displayed on the traffic network at the location of any incident. Also, in the information window, text information will be provided describing the incident for example: "Severe injury accident on 1st Street between F Street and G Street".

2. En-route Guidance. Graphical arrows indicating advised turning movements at each upcoming intersection will be provided as well as a textual description of the advice.

3. Pre-trip Guidance. An advised minimum path from the origin to the destination will be displayed at the beginning of trip by outlining the route in white.

4. Congestion Information. The street segments in the simulation will be color coded to indicate different levels of traffic congestion with green indicating normal congestion, yellow indicating moderate congestion and red indicating severe congestion.

Note:

The information provided by this system will not always be correct, but it should help to improve your average travel time to work.

Procedures:

Once you have completed reading these instructions, a demonstration program will be executed which will allow you to become familiar with the simulation. The demo will allow you to make some practice trials so that you can become familiar with how the simulation works, how the various types of information are displayed, and how to maneuver your vehicle curser using the keyboard arrow keys. When you have performed enough practice trails so that you feel comfortable using the simulation, you can select to move on to the actual experiment.

If you have any questions during this Demonstration stage, please do not hesitate to ask the experiment coordinator before continuing on and beginning the experiment.

Once the actual experiment begins, the main keyboard keys which you will need to use are the space bar and the four arrow keys. The space bar is used to begin the trial on each commute day. At the start of the trail traffic information will be displayed depending on the types of systems which have been provided for you.

Different subjects will be receiving different combinations of traffic information or if you are in the control group, you will not have any information displayed and will have to make decisions based on your own judgements.

Use the four arrow keys to move your vehicle curser from the origin to the destination. If at any time the vehicle curser changes color from green to red, this indicates that the simulation is waiting for you to make a decision. The time that it takes you to make decisions is being recorded, so try to respond as quickly as possible.

In the simulation you are allowed to make turning decisions in advance and each time you press an arrow key, that turning movement is stored for the next intersection. For example, if you know in advance that you want to go straight through the next two intersections and then turn south at the third intersection, you could at that point press the straight arrow twice and the south arrow once and then just wait until you get to the fourth intersection before making any more decisions. During the demonstration program you should try this feature to get familiar with it.

At the completion of each trail day you will be asked to respond to a few questions regarding your days travel and you will be shown the travel time on the route which you followed as well as the travel time for the best possible route for that travel day. Also, at the end of the simulation, you will be asked to respond to a few short questions.

If at any time you have any questions please ask the experimenter.